NAME

cxgbe - Chelsio T4-, T5-, and T6-based 100Gb, 40Gb, 25Gb, 10Gb, and 1Gb Ethernet adapter driver

SYNOPSIS

To compile this driver into the kernel, place the following lines in your kernel configuration file:

```
device cxgbe
```

To load the driver as a module at boot time, place the following lines in loader.conf(5):

```
t4fw_cfg_load="YES"
t5fw_cfg_load="YES"
t6fw_cfg_load="YES"
if_cxgbe_load="YES"
```

DESCRIPTION

The **cxgbe** driver provides support for PCI Express Ethernet adapters based on the Chelsio Terminator 4, Terminator 5, and Terminator 6 ASICs (T4, T5, and T6). The driver supports Jumbo Frames, Transmit/Receive checksum offload, TCP segmentation offload (TSO), Large Receive Offload (LRO), VLAN tag insertion/extraction, VLAN checksum offload, VLAN TSO, VXLAN checksum offload, VXLAN TSO, and Receive Side Steering (RSS). For further hardware information and questions related to hardware requirements, see *http://www.chelsio.com/*.

The **cxgbe** driver uses different names for devices based on the associated ASIC:

ASIC	Port Name	Parent Device	Virtual Interface
T4	cxgbe	t4nex	vcxgbe
T5	cxl	t5nex	vcxl
T6	cc	t6nex	vcc

Loader tunables with the hw.cxgbe prefix apply to all cards. The driver provides sysctl MIBs for both ports and parent devices using the names above. For example, a T5 adapter provides port MIBs under dev.cxl and adapter-wide MIBs under dev.t5nex. References to sysctl MIBs in the remainder of this page use dev.<port> for port MIBs and dev.<nexus> for adapter-wide MIBs.

For more information on configuring this device, see ifconfig(8).

HARDWARE

The **cxgbe** driver supports 100Gb and 25Gb Ethernet adapters based on the T6 ASIC:

- Chelsio T6225-CR
- Chelsio T6225-SO-CR
- Chelsio T62100-LP-CR
- Chelsio T62100-SO-CR
- Chelsio T62100-CR

The **cxgbe** driver supports 40Gb, 10Gb and 1Gb Ethernet adapters based on the T5 ASIC:

- Chelsio T580-CR
- Chelsio T580-LP-CR
- Chelsio T580-LP-SO-CR
- Chelsio T560-CR
- Chelsio T540-CR
- Chelsio T540-LP-CR
- Chelsio T522-CR
- Chelsio T520-LL-CR
- Chelsio T520-CR
- Chelsio T520-SO
- Chelsio T520-BT
- Chelsio T504-BT

The **cxgbe** driver supports 10Gb and 1Gb Ethernet adapters based on the T4 ASIC:

- Chelsio T420-CR
- Chelsio T422-CR
- Chelsio T440-CR
- Chelsio T420-BCH
- Chelsio T440-BCH
- Chelsio T440-CH
- Chelsio T420-SO
- Chelsio T420-CX
- Chelsio T420-BT
- Chelsio T404-BT

LOADER TUNABLES

Tunables can be set at the loader(8) prompt before booting the kernel or stored in loader.conf(5). There are multiple tunables that control the number of queues of various types. A negative value for such a tunable instructs the driver to create up to that many queues if there are enough CPU cores available.

hw.cxgbe.ntxq

Number of NIC tx queues used for a port. The default is 16 or the number of CPU cores in the system, whichever is less.

hw.cxgbe.nrxq
Number of NIC rx queues used for a port. The default is 8 or the number of CPU cores in the system, whichever is less.

hw.cxgbe.nofldtxq
Number of TOE tx queues used for a port. The default is 8 or the number of CPU cores in the system, whichever is less.

hw.cxgbe.nofldrxq
Number of TOE rx queues used for a port. The default is 2 or the number of CPU cores in the system, whichever is less.

hw.cxgbe.num_vis
Number of virtual interfaces (VIs) created for each port. Each virtual interface creates a separate network interface. The first virtual interface on each port is required and represents the primary network interface on the port. Additional virtual interfaces on a port are named using the Virtual Interface name from the table above. Additional virtual interfaces use a single pair of queues for rx and tx as well an additional pair of queues for TOE rx and tx. The default is 1.

hw.cxgbe.holdoff_timer_idx

hw.cxgbe.holdoff_timer_idx_ofld
Timer index value used to delay interrupts. The holdoff timer list has the values 1, 5, 10, 50, 100, and 200 by default (all values are in microseconds) and the index selects a value from this list. holdoff_timer_idx_ofld applies to queues used for TOE rx. The default value is 1 which means the timer value is 5us. Different interfaces can be assigned different values at any time via the dev.<port>.X.holdoff_tmr_idx and dev.<port>.X.holdoff_tmr_idx_ofld sysctls.

hw.cxgbe.holdoff_pktc_idx

hw.cxgbe.holdoff_pktc_idx_ofld
Packet-count index value used to delay interrupts. The packet-count list has the values 1, 8, 16, and 32 by default, and the index selects a value from this list. holdoff_pktc_idx_ofld applies to queues used for TOE rx. The default value is -1 which means packet counting is disabled and interrupts are generated based solely on the holdoff timer value. Different interfaces can be assigned different values via the dev.<port>.X.holdoff_pktc_idx and dev.<port>.X.holdoff_pktc_idx_ofld sysctls. These sysctls work only when the interface has

never been marked up (as done by ifconfig up).

hw.cxgbe.qsize_txq

Number of entries in a transmit queue's descriptor ring. A buf_ring of the same size is also allocated for additional software queuing. See ifnet(9). The default value is 1024. Different interfaces can be assigned different values via the dev.<port>.X.qsize_txq sysctl. This sysctl works only when the interface has never been marked up (as done by ifconfig up).

hw.cxgbe.qsize_rxq

Number of entries in a receive queue's descriptor ring. The default value is 1024. Different interfaces can be assigned different values via the dev.<port>.X.qsize_rxq sysctl. This sysctl works only when the interface has never been marked up (as done by ifconfig up).

hw.cxgbe.interrupt_types

Permitted interrupt types. Bit 0 represents INTx (line interrupts), bit 1 MSI, and bit 2 MSI-X. The default is 7 (all allowed). The driver selects the best possible type out of the allowed types.

hw.cxgbe.pcie_relaxed_ordering

PCIe Relaxed Ordering. -1 indicates the driver should determine whether to enable or disable PCIe RO. 0 disables PCIe RO. 1 enables PCIe RO. 2 indicates the driver should not modify the PCIe RO setting. The default is -1.

hw.cxgbe.fw_install

0 prohibits the driver from installing a firmware on the card. 1 allows the driver to install a new firmware if internal driver heuristics indicate that the new firmware is preferable to the one already on the card. 2 instructs the driver to always install the new firmware on the card as long as it is compatible with the driver and is a different version than the one already on the card. The default is 1.

hw.cxgbe.fl_pktshift

Number of padding bytes inserted before the beginning of an Ethernet frame in the receive buffer. The default value is 0. A value of of 2 would ensure that the Ethernet payload (usually the IP header) is at a 4 byte aligned address. 0-7 are all valid values.

hw.cxgbe.fl_pad

A non-zero value ensures that writes from the hardware to a receive buffer are padded up to the specified boundary. The default is -1 which lets the driver pick a pad boundary. 0 disables trailer padding completely.

hw.cxgbe.cong_drop

Controls the hardware response to congestion. -1 disables congestion feedback and is not recommended. 0 instructs the hardware to backpressure its pipeline on congestion. This usually results in the port emitting PAUSE frames. 1 instructs the hardware to drop frames destined for congested queues. 2 instructs the hardware to both backpressure the pipeline and drop frames.

hw.cxgbe.pause_settings

PAUSE frame settings. Bit 0 is rx_pause, bit 1 is tx_pause, bit 2 is pause_autoneg. rx_pause = 1 instructs the hardware to heed incoming PAUSE frames, 0 instructs it to ignore them. tx_pause = 1 allows the hardware to emit PAUSE frames when its receive FIFO reaches a high threshold, 0 prohibits the hardware from emitting PAUSE frames. pause_autoneg = 1 overrides the rx_pause and tx_pause bits and instructs the hardware to negotiate PAUSE settings with the link peer. The default is 7 (all three = 1). This tunable establishes the default PAUSE settings for all ports. Settings can be displayed and controlled on a per-port basis via the dev.<port>.X.pause_settings sysctl.

hw.cxgbe.fec

Forward Error Correction settings. -1 (default) means driver should automatically pick a value. 0 disables FEC. Finer grained control can be achieved by setting individual bits. Bit 0 enables RS FEC, bit 1 enables BASE-R FEC (aka Firecode FEC), bit 2 enables NO FEC, and bit 6 enables the FEC that is recommended by the transceiver/cable that is plugged in. These bits can be set together in any combination. This tunable establishes the default FEC settings for all ports. Settings can be controlled on a per-port basis via the dev.<port>.X.requested_fec sysctl. The FEC in use on the link is available in dev.<port>.X.link_fec when the link is up.

hw.cxgbe.autoneg

Link autonegotiation settings. This tunable establishes the default autonegotiation settings for all ports. Settings can be displayed and controlled on a per-port basis via the dev.<port>.X.autoneg sysctl. 0 disables autonegotiation. 1 enables autonegotiation. The default is -1 which lets the driver pick a value. dev.<port>.X.autoneg is -1 for port and module combinations that do not support autonegotiation.

hw.cxgbe.buffer_packing

Allow the hardware to deliver multiple frames in the same receive buffer opportunistically. The default is -1 which lets the driver decide. 0 or 1 explicitly disable or enable this feature.

hw.cxgbe.largest_rx_cluster

hw.cxgbe.safest_rx_cluster

Sizes of rx clusters. Each of these must be set to one of the sizes available (usually 2048, 4096, 9216, and 16384) and largest_rx_cluster must be greater than or equal to safest_rx_cluster. The

defaults are 16384 and 4096 respectively. The driver never attempts to allocate a receive buffer larger than largest_rx_cluster and falls back to allocating buffers of safest_rx_cluster size if an allocation larger than safest_rx_cluster fails. Note that largest_rx_cluster merely establishes a ceiling -- the driver is allowed to allocate buffers of smaller sizes.

hw.cxgbe.config_file

Select a pre-packaged device configuration file. A configuration file contains a recipe for partitioning and configuring the hardware resources on the card. This tunable is for specialized applications only and should not be used in normal operation. The configuration profile currently in use is available in the dev.<nexus>.X.cf and dev.<nexus>.X.cfcsum sysctls.

hw.cxgbe.linkcaps_allowed

hw.cxgbe.niccaps_allowed

hw.cxgbe.toecaps_allowed

hw.cxgbe.rdmacaps_allowed

hw.cxgbe.iscsicaps_allowed

hw.cxgbe.fcoecaps_allowed

Disallowing capabilities provides a hint to the driver and firmware to not reserve hardware resources for that feature. Each of these is a bit field with a bit for each sub-capability within the capability. This tunable is for specialized applications only and should not be used in normal operation. The capabilities for which hardware resources have been reserved are listed in dev.<nexus>.X.*caps sysctls.

hw.cxgbe.tx_vm_wr

Setting this to 1 instructs the driver to use VM work requests to transmit data. This lets PF interfaces transmit frames to VF interfaces over the internal switch in the ASIC. Note that the cxgbev(4) VF driver always uses VM work requests and is not affected by this tunable. The default value is 0 and should be changed only if PF and VF interfaces need to communicate with each other. Different interfaces can be assigned different values using the dev.<port>.X.tx_vm_wr sysctl when the interface is administratively down.

hw.cxgbe.attack_filter

Set to 1 to enable the "attack filter". Default is 0. The attack filter will drop an incoming frame if any of these conditions is true: src ip/ip6 == dst ip/ip6; tcp and src/dst ip is not unicast; src/dst ip is loopback (127.x.y.z); src ip6 is not unicast; src/dst ip6 is loopback (::1/128) or unspecified

(::/128); tcp and src/dst ip6 is mcast (ff00::/8). This facility is available on T4 and T5 based cards only.

hw.cxgbe.drop_ip_fragments
Set to 1 to drop all incoming IP fragments. Default is 0. Note that this drops valid frames.

hw.cxgbe.drop_pkts_with_l2_errors
Set to 1 to drop incoming frames with Layer 2 length or checksum errors. Default is 1.

hw.cxgbe.drop_pkts_with_l3_errors
Set to 1 to drop incoming frames with IP version, length, or checksum errors. The IP checksum is validated for TCP or UDP packets only. Default is 0.

hw.cxgbe.drop_pkts_with_l4_errors
Set to 1 to drop incoming frames with Layer 4 (TCP or UDP) length, checksum, or other errors. Default is 0.

SUPPORT

For general information and support, go to the Chelsio support website at: *http://www.chelsio.com/*.

If an issue is identified with this driver with a supported adapter, email all the specific information related to the issue to <*support@chelsio.com*>.

SEE ALSO

arp(4), ccr(4), cxgb(4), cxgbev(4), netintro(4), ng_ether(4), ifconfig(8)

HISTORY

The **cxgbe** device driver first appeared in FreeBSD 9.0. Support for T5 cards first appeared in FreeBSD 9.2 and FreeBSD 10.0. Support for T6 cards first appeared in FreeBSD 11.1 and FreeBSD 12.0.

AUTHORS

The **cxgbe** driver was written by Navdeep Parhar <*np@FreeBSD.org*>.