



Appendix - UFM Subnet Manager Default Properties

Table of contents

Configuring UFM for SR-IOV

Isolating Switch From Routing

The following table provides a comprehensive list of UFM SM default properties.

Category	Property	Config File Attribute	Default	Mode/Field	Description
Generic	Subnet Prefix	subnet_prefix	0xfe80000000 000000	RW	Subnet prefix used on the subnet 0xfe80000000000000
	LMC	lmc	0	RW	The LMC value used on the subnet: 0-7 Changes to the LMC parameter require a UFM restart.
	SM LID	master_sm_lid	0		Force LID for local SM when in MASTER state Selected LID must match configured LMC 0 disables the feature
Keys	M_Key	m_key	0x0000000000 000000	RW	M_Key value sent to all ports -used to qualify the set(PortInfo)
	M_Key Lease Period	m_key_lease_period	0	RW	The lease period used for the M_Key on the subnet in [sec]
	SM_Key	sm_key	0x0000000000 000001	RO	SM_Key value of the SM used for SM authentication
	SA_Key	sa_key	0x0000000000 000001	RO	SM_Key value to qualify rcv SA queries as 'trusted'
	Partition enforcement	part_enforce	<ul style="list-style-type: none"> • Out • In • Both (default-outbound and 	RO	Partition enforcement type (for switches)

Category	Property	Config File Attribute	Default	Mode/Field	Description
			inbound enforcement enabled)		
	MKEY lookup	m_key_lookup	FALSE	RW	If FALSE, SM will not try to determine the m_key of unknown ports.
	M_Key Per Port	m_key_per_port	FALSE	RW	When m_key_per_port is enabled, OpenSM will generate an M_Key for each port
Limits	Packet Life Time	packet_life_time	0x12	RW	The maximum lifetime of a packet in a switch. The actual time is 4.096usec * $2^{<packet_life_time>}$ The value 0x14 disables the mechanism
	VL Stall Count	vl_stall_count	0x07	RO	The number of sequential packets dropped that cause the port to enter the VL Stalled state. The result of setting the count to zero is undefined.
	Leaf VL Stall Count	leaf_vl_stall_count	0x07	RO	The number of sequential packets dropped that causes the port to enter the leaf VL Stalled state. The count is for switch ports driving a CA or gateway port. The result of setting the count to zero is undefined.

Category	Property	Config File Attribute	Default	Mode/Field	Description
	Head Of Queue Lifetime	head_of_queue_lifetime	0x12	RW	The maximum time a packet can wait at the head of the transmission queue. The actual time is $4.096\text{usec} * 2^{\langle \text{head_of_queue_lifetime} \rangle}$ The value 0x14 disables the mechanism
	Leaf Head Of Queue Lifetime	leaf_head_of_queue_lifetime	0x10	RW	The maximum time a packet can wait at the head of queue on a switch port connected to a CA or gateway port.
	Maximal Operational VL	max_op_vls	2	RW	Limit of the maximum operational VLs
	Force Link Speed	force_link_speed	15 (Do NOT change)	RO	Force PortInfo: LinkSpeedEnabled on switch ports. If 0, do not modify. Values are: 1: 2.5 Gbps 3: 2.5 or 5.0 Gbps 5: 2.5 or 10.0 Gbps 7: 2.5 or 5.0 or 10.0 Gbps 2,4,6,8-14 Reserved 15: set to PortInfo: LinkSpeedSupported
Limits	Subnet Timeout	subnet_timeout	18 (1second)	RW	The subnet_timeout code that will be set for all the ports.

Category	Property	Config File Attribute	Default	Mode/Field	Description
					The actual timeout is 4.096usec * 2^<subnet_timeout>
	Local PHY Error Threshold	local_phy_errors_threshold	0x08	RW	Threshold of local phy errors for sending Trap 129
	Overrun Errors Threshold	overrun_errors_threshold	0x08	RW	Threshold of credit overrun errors for sending Trap 130
Sweep	Sweep Interval	sweep_interval	10	RW	The time in seconds between subnet sweeps (Disabled if 0)
	Reassign Lids	reassign_lids	FALSE (disabled)	RW	If TRUE (enabled), all LIDs are reassigned
	Force Heavy Sweep	force_heavy_sweep_window	-1	RW	Forces heavy sweep after number of light sweeps (-1 disables this option and 0 will cause every sweep to be heavy)
	Sweep On trap	sweep_on_trap	TRUE (enabled)	RW	If TRUE every trap 128 and 144 will cause a heavy sweep
	Alternative Route Calculation	max_alt_dr_path_retries	4	RW	Maximum number of attempts to find an alternative direct route towards unresponsive ports

Category	Property	Config File Attribute	Default	Mode/Field	Description
	Fabric Rediscovery	max_seq_redisc	2	RW	Max Failed Sequential Discovery Loops
	Offsweep Rebalancing Enable	offsweep_balancing_enabled	FALSE	RW	Enable/Disable idle time routing rebalancing
	Offsweep Rebalancing Window	offsweep_balancing_window	180	RW	Set the time window in seconds after sweep to start rebalancing
Handover	SM Priority	sm_priority	15	RO	SM (enabled). The priority used for deciding which is the master. Range is 0 (lowest priority) to 15 (highest)
	Ignore Other SMs	ignore_other_sm	FALSE (disabled)	RO	If TRUE other SMs on the subnet should be ignored
	Polling Timeout	sminfo_polling_timeout	10	RO	Timeout in seconds between two active master SM polls
	Polling Retries	polling_retry_number	4	RO	Number of failing remote SM polls that declares it non-operational
	Honor GUID-to-LID File	honor_guid2lid_file	FALSE (disabled)	RO	If TRUE, honor the guid2lid file when coming out of standby state, if the

Category	Property	Config File Attribute	Default	Mode/Field	Description
					guid2lid file exists and is valid
	Allowed SM GUID list	allowed_sm_guids	(null) (disabled)		List of Host GUIDs where SM is allowed to run when specified. OpenSM ignores SM running on port that is not in this list. If 0, does not allow any other SM. If null, the feature is disabled.
Threading	Max Wire SMPs	max_wire_smps	8	RW	Maximum number of SMPs sent in parallel
	Transaction Timeout	transaction_timeout	200	RO	The maximum time in [msec] allowed for a transaction to complete
	Max Message FIFO Timeout	max_msg_fifo_timeout	10000	RO	Maximum time in [msec] a message can stay in the incoming message queue
	Routing Threads	routing_threads_num	0	RW	Number of threads to be used for parallel minhop/updn calculations. If 0, number of threads will be equal to number of processors.
	Routing Thread	max_threads_per_core	0	RW	Max number of threads that are allowed to run on

Category	Property	Config File Attribute	Default	Mode/Field	Description
	s Per Core				the same processor during parallel computing. If 0, threads assignment per processor is up to operating system initial assignment.
Logging	Log File	log_file	/opt/ufm/files/log/opensm.log	RO	Path of Log file to be used
	Log Flags	log_flags	Error and Info 0x03	RW	The log flags, or debug level being used.
	Force Log Flush	force_log_flush	FALSE (disabled)	RO	Force flush of the log file after each log message
	Log Max Size	log_max_size	4096	RW	Limit the size of the log file in MB. If overrun, log is restarted
	Accumulate Log File	accum_log_file	TRUE (enabled)	RO	If TRUE, will accumulate the log over multiple OpenSM sessions
	Dump Files Directory	dump_files_dir	/opt/ufm/files/log	RO	The directory to hold the file SM dumps (for multicast forwarding tables for example). The file is used collects information.
	Syslog log	syslog_log	0x0	RW	Sets a verbosity of messages to be printed in syslog
Misc	Node Names	node_name_map_name	Null	RW	Node name map for mapping node's to more

Category	Property	Config File Attribute	Default	Mode/Field	Description
	Map File				descriptive node descriptions
	SA database File	sa_db_file	Null	RO	SA database file name
	No Clients Reregistration	no_clients_rereg	FALSE (disabled)	RO	If TRUE, disables client reregistration
	Exit On Fatal Event	exit_on_fatal	TRUE (enabled)	RO	If TRUE (enabled), the SM exits for fatal initialization issues
	Switch Isolation From Routing	held_back_sw_file	Null	RW	File that contains GUIDs of switches isolated from routing
	Enable NVIDIA SHARP support	sharp_enabled	Enabled	RW	Defines whether to enable/disable NVIDIA SHARP on supporting ports.
Multicast	Disable Multicast	disable_multicast	FALSE (disabled)	RO	If TRUE, OpenSM should disable multicast support and no multicast routing is performed
	Multicast Group Parameters	default_mcg_mtu	0	RW	Default MC group MTU for dynamic group creation. 0 disables this feature, otherwise, the value is a valid IB encoded MTU
Multicast	Multicast	default_mcg_rate	0	RW	Default MC group rate for dynamic group creation. 0

Category	Property	Config File Attribute	Default	Mode/Field	Description
	Group Parameters				disables this feature, otherwise, the value is a valid IB encoded rate
Multicast	Enable incremental multicast routing	enable_inc_mc_routing	FALSE	RW	Enable incremental multicast routing
Multicast	MC root file	mc_roots_file	null	RW	Specify predefined MC groups root guides
QoS	Settings	qos	FALSE (disabled) *From UFM v3.7 and on	RW	If FALSE (disabled), SM will not apply QoS settings
Unhealthy Ports	Enabling Unhealthy Ports	hm_unhealthy_ports_checks	TRUE	RW	Enables Unhealthy Ports configuration
	Configuration file	hm_ports_health_policy_file	null	RW	Specifies configuration file for health policy
	Unhealthy actions	hm_sw_manual_action	no_discover	RW	Specifies what to do with switch ports which were manually added to health policy file
	MADs validation	validate_smp	TRUE	RW	If set to TRUE, opensm will ignore nodes sending non-spec compliant MADs. When set to FALSE, opensm will log the warning in the opensm

Category	Property	Config File Attribute	Default	Mode/Field	Description
					log file about non-compliant node
Routing	Unicast Routing engine	routing_engine	(null)	RW	By default, ar_updn routing engine is used by the SM. Supported routing engines are minhop, updn, dnuip, ftree, dor, torus-2QoS, kdor-hc, kdor-ghc, dfp, dfp2, ar_updn, ar_ftree and ar_dor.
	Randomization	scatter_ports	8	RW	Assigns ports in a random order instead of round-robin. If 0, the feature is disabled, otherwise use the value as a random seed. Applicable to the MINHOP/UPDN routing algorithms
	Randomization	guid_routing_order_no_scatter	TRUE	RO	Do not use scatter for ports defined in guid_routing_order file
	Unicast Routing Caching	use_ucast_cache	TRUE	RW	Use unicast routing cache for routing computation time improvement
	GUID Ordering During Routing	guid_routing_order_file	NULL	RW	The file holding guid routing order of particular guides (for MinHop, Up/Down)

Category	Property	Config File Attribute	Default	Mode/Field	Description
	Torus Routing	torus_config	/opt/ufm/files/conf/opensm/torus-2QoS.con	RW	Torus-2QoS configuration file name
	Routing Chains	pgrp_policy_file	NULL	RW	The file holding the port groups policy
		topo_policy_file	NULL	RW	The file holding the topology policy
		rch_policy_file	NULL	RW	The file holding the routing chains policy
		max_topologies_per_sw	1	RO	Defines maximal number of topologies to which a single switch may be assigned during routing engine chain configuration.
	Incremental Multicast Routing (IMR)	enable_inc_mc_routing	TRUE	RW	If TRUE, MC nodes will be added to the MC tree incrementally. When set to FALSE, the tree will be recalculated per each change.
	MC Global root	mc_primary_root_guid/mc_secondary_root_guid	0x0000000000000000 (for both)	RW	Primary and Secondary global mc root guid
	Scatter ports	use_scatter_for_switch_lid	FALSE	RW	Use scatter when routing to the switch's LIDs
	updn lid tracking mode	updn_lid_tracking_mode	FALSE	RW	Controls whether SM will use LID tracking or not when updn or ar_updn routing engine is used

Category	Property	Config File Attribute	Default	Mode/Field	Description
Events	Event Subscription Handling	drop_subscr_on_report_fail	FALSE	RW	Drop subscription on report failure (o13-17.2.1)
	Event Subscription Handling	drop_event_subscriptions	TRUE	RW	Drop event subscriptions (InformInfo and ServiceRecords) on port removal and SM coming out of STANDBY
Virtualization	Virtualization enabled	virt_enabled	Enabled	RW	Enables/disables virtualization support
	Maximum ports in virtualization process	virt_max_ports_in_process	64	RW	Sets a number of ports to be handled on each virtualization process cycle
Router	Router aguid enable	rtr_aguid_enable	0 (Disabled)	RW	Defines whether the SM should create alias GUIDs required for router support for each HCA port
	Router path record flow label	rtr_pr_flow_label	0	RW	Defines flow label value to use in multi-subnet path query responses
	Router path record tclass	rtr_pr_tclass	0	RW	Defines tclass value to use in multi-subnet path query responses.

Category	Property	Config File Attribute	Default	Mode/Field	Description
	Router path record sl	rtr_pr_sl	0	RW	Defines sl value to use in multi-subnet path query responses
	Router path record MTU	rtr_pr_mtu	4 (IB_MTU_LEN_2048)	RW	Define MTU value to use in multi-subnet path query responses
	Router path record rate	rtr_pr_rate	16 (IB_PATH_RECORD_RATE_100_GBS)	RW	Defines rate value to use in multi-subnet path query responses
SA Security	SA Enhanced Trust Model (SAETM)	sa_enhanced_trust_model	FALSE	RW	Controls whether SAETM is enabled.
	Untrusted Guidinfo records	sa_etm_allow_untrusted_guidinfo_rec	FALSE	RW	Controls whether to allow Untrusted Guidinfo record requests in SAETM.
	Guidinfo record requests by VF	sa_etm_allow_guidinfo_rec_by_vf	FALSE	RW	Controls whether to allow Guidinfo record requests by vf in SAETM.
	Untrusted proxy	sa_etm_allow_untrusted_proxy_requests	FALSE	RW	Controls whether to allow Untrusted proxy requests in SAETM.

Category	Property	Config File Attribute	Default	Mode/Field	Description
	requests				
	Max number of multicast groups	sa_etm_max_num_mcgs	128	RW	Max number of multicast groups per port/vport that can be registered.
	Max number of service records	sa_etm_max_num_srvcs	32	RW	Max number of service records per port/vport that can be registered.
	Max number of event subscriptions	sa_etm_max_num_event_subs	32	RW	Max number of event subscriptions (InformInfo) per port/vport that can be registered.
	SGID spoofing	sa_check_sgid_spoofing	TRUE	RW	If enabled, the SA checks for SGID spoofing in every request with GRH included, unless the SLID is from a router port at that request.

Configuring UFM for SR-IOV

Single-root I/O virtualization (SR-IOV) enables a PCI Express (PCIe) device to appear to be multiple separate physical PCIe devices.

UFM is ready to work with SR-IOV devices by default. You can fine-tune the configuration using the SM configuration.

The following arguments are available for ConnectX-5 and later devices:

Argument	Value	Description
virt_enabled	<ul style="list-style-type: none"> • 0 – no virtualization support • 1 – disable virtualization on all virtualization supporting ports • 2 – enable virtualization on all virtualization supporting ports (default) 	Virtualization support
virt_max_ports_in_process	Possible values: 0-65535; where 0 processes all pending ports Default: 64	Maximum number of ports to be processed simultaneously by the virtualization manager
virt_default_hop_limit	Possible values: 0-255 Default: 2	Default value for hop limit to be returned in path records where either the source or destination are virtual ports

Isolating Switch From Routing

UFM can isolate particular switches from routing in order to perform maintenance of the switches with minimal interruption to the existing traffic in the fabric.

Isolating a switch from routing is done via UFM Subnet Manager as follows:

1. Create a file that includes either the node GUIDs or system GUID of the switches under maintenance. For example:

```
0x1234566
0x1234567
```

2. Set the filename of the parameter held_back_sw_file in the /conf/opensm.conf file (the same as the file created in Step 1).
3. Run:

```
kill -s HUP 'pidof opensm'
```

Once SM completes rerouting, the traffic does not go through the ports of isolated switches.

To attach the switch to the routing:

1. Remove the GUID of the switch from the list of isolated switches defined in Step 1 of the isolation process.

2. Run:

```
kill -s HUP 'pidof opensm'
```

Once SM completes rerouting, traffic will go through the switch.

© Copyright 2024, NVIDIA. PDF Generated on 06/06/2024