Autonomous Link Maintenance (ALM) Plugin

Table of contents

List of Figures

Overview

The primary objective of the Autonomous Link Maintenance (ALM) plugin is to enhance cluster availability and improve the rate of job completion. This objective is accomplished by utilizing machine learning (ML) models to predict potential link failures. The plugin then isolates the expected failing links, implements maintenance procedures on them, and subsequently restores the fixed links to their original state by removing the isolation.

The ALM plugin performs the following tasks:

1. Collects telemetry data from UFM and employs ML jobs to predict which ports need to be isolated/de-isolated

2. Identifies potential link failures and isolates them to avert any interruption to traffic flow

3. Maintains a record of maintenance procedures that can be executed to restore an isolated link

4. After performing the required maintenance, the system verifies if the links can be de-isolated and restored to operational status (brought back online)

The ALM plugin operates in the following two distinct modes:

1. Shadow mode

    ○ Collects telemetry data, runs ML prediction jobs, and saves the predictions to files.

2. Active mode

    ○ Collects telemetry data, runs ML prediction jobs, and saves the predictions to files.

    ○ Automatically isolates and de-isolates based on predictions.

    ○ It is essential to note that a subset of the links must be specified in the allow list to enable this functionality.

Schematic Flow
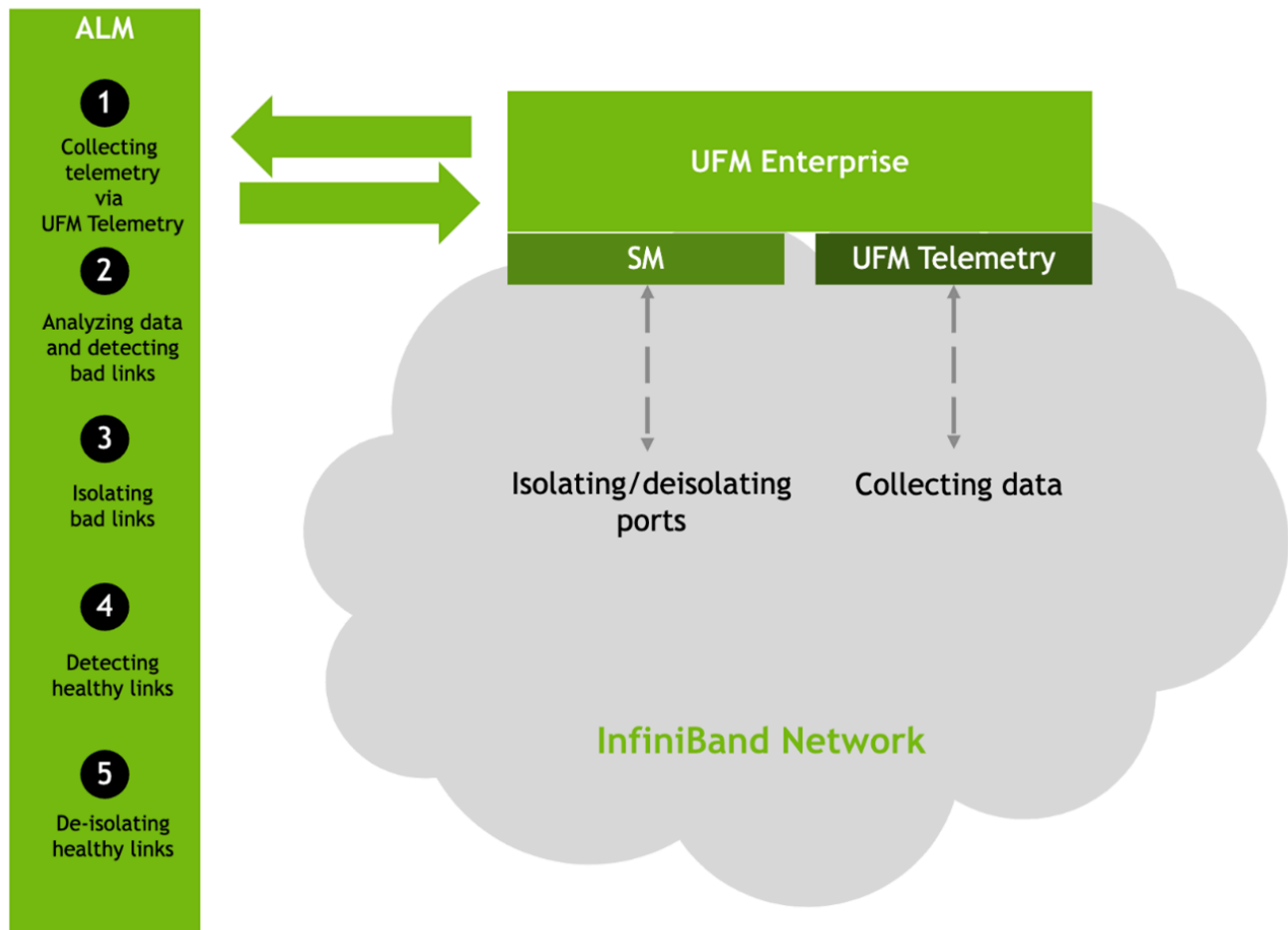
Deployment

The Autonomous Link Maintenance (ALM) plugin can be deployed using the following methods:

1. On the UFM Appliance

2. On the UFM Software

To deploy the plugin, follow these steps:
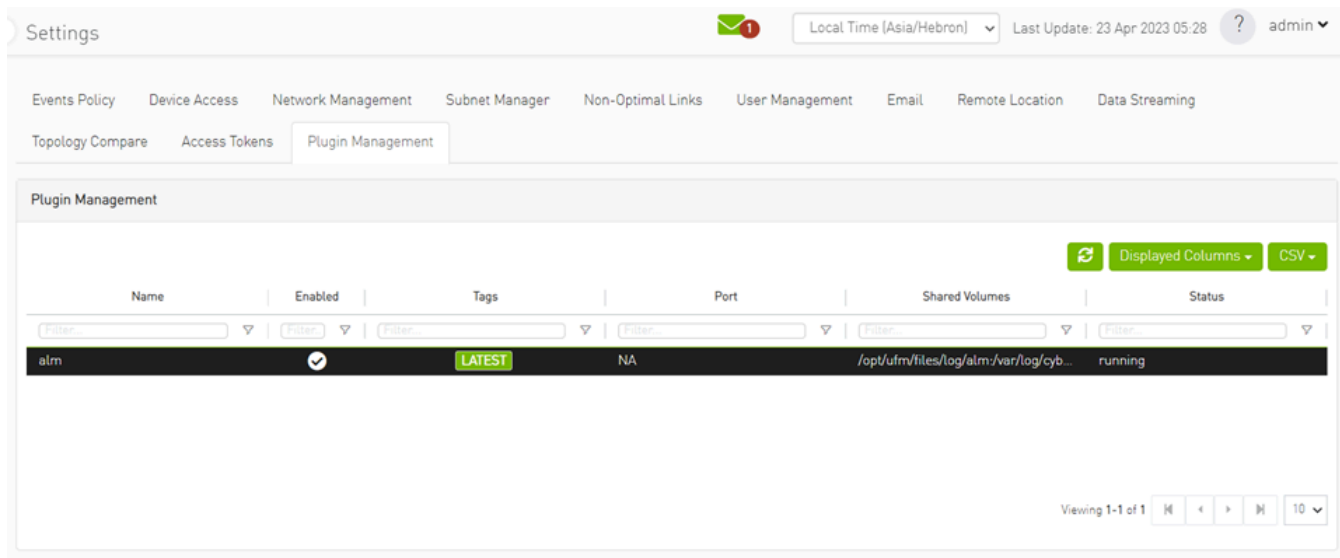
1. Download the ufm-plugin-alm-image from the NVIDIA License Portal (NLP).

2. Load the downloaded image onto the UFM server. This can be done either by using the UFM GUI by navigating to the Settings -> Plugins Management tab or by loading the image via the following instructions:

3. Log in to the UFM server terminal.

4. Run:

```
docker load -I <path_to_image>
```

5. After successfully loading the plugin image, the plugin should become visible within the plugins management table within the UFM GUI. To initiate the plugin's execution, simply right-click on the respective in the table.

> ⓘ **Note**
>
> **The supported InfiniBand hardware technologies are HDR, Beta on NDR.**

Data Collection

The ALM plugin collects data from the UFM Enterprise appliance in the following two methods:

1. Low-frequency collection: This process occurs every 7 minutes and gathers data for the following counter: hist0, hist1, hist2, hist3, hist4, phy_effective_errors, phy_symbol_errors

2. High-frequency collection : This process occurs every 10 seconds and gathers data for the following counters:
   phy_state,logical_state,link_speed_active,link_width_active,fec_mode_active, raw_ber,eff_ber,symbol_ber,phy_raw_errors_lane0,phy_raw_errors_lane1,phy_raw_er phy_raw_errors_lane3,phy_effective_errors,phy_symbol_errors,time_since_last_clear, hist0,hist1,hist2,hist3,hist4,switch_temperature,CableInfo.temperature,link_down_ev plr_rcv_codes,plr_rcv_code_err,plr_rcv_uncorrectable_code,plr_xmit_codes,plr_xmit_r plr_xmit_retry_events,plr_sync_events,hi_retransmission_rate,fast_link_up_status, time_to_link_up,status_opcode,status_message,down_blame,local_reason_opcode, remote_reason_opcode,e2e_reason_opcode,num_of_ber_alarams,PortRcvRemotePhy PortRcvErrorsExtended,PortXmitDiscardsExtended,PortRcvSwitchRelayErrorsExtende VL15DroppedExtended,PortXmitWaitExtended,PortXmitDataExtended,PortRcvDataE PortRcvPktsExtended,PortUniCastXmitPktsExtended,PortUniCastRcvPktsExtended,Pc

3. The collected counters can be configurable and customized to suit your requirements. The counters can be found at /opt/ufm/conf/plugins/alm/counters.cfg

```
root@r-ufm116:~# cat  /opt/ufm/conf/plugins/alm/counters.cfg
[HighFreq]
phy_state = last_update_value
logical_state = last_update_value
link_speed_active = last_update_value
link_width_active = last_update_value
fec_mode_active = last_update_value
raw_ber = last_update_value
eff_ber = last_update_value
symbol_ber = last_update_value
phy_raw_errors_lane0 = delta
phy_raw_errors_lane1 = delta
phy_raw_errors_lane2 = delta
phy_raw_errors_lane3 = delta
phy_effective_errors = delta
phy_symbol_errors = delta
time_since_last_clear = last_update_value
hist0 = delta
hist1 = delta
hist2 = delta
hist3 = delta
hist4 = delta
switch_temperature = last_update_value
CableInfo.Temperature = last_update_value
link_down_events = delta
plr_rcv_codes = delta
plr_rcv_code_err = delta
plr_rcv_uncorrectable_code = delta
plr_xmit_codes = delta
plr_xmit_retry_codes = delta
plr_xmit_retry_events = delta
plr_sync_events = delta
hi_retransmission_rate = delta
fast_link_up_status = last_update_value
time_to_link_up = last_update_value
status_opcode = last_update_value
status_message = last_update_value
down_blame = last_update_value
local_reason_opcode = last_update_value
remote_reason_opcode = last_update_value
e2e_reason_opcode = last_update_value
num_of_ber_alarams = delta
PortRcvRemotePhysicalErrorsExtended = delta
PortRcvErrorsExtended = delta
PortXmitDiscardsExtended = delta
PortRcvSwitchRelayErrorsExtended = delta
```

ALM Configuration

The ALM configuration is used for controlling isolation/de-isolation. The configuration can be found under /opt/ufm/cyber-ai/conf/cyberai.cfg.
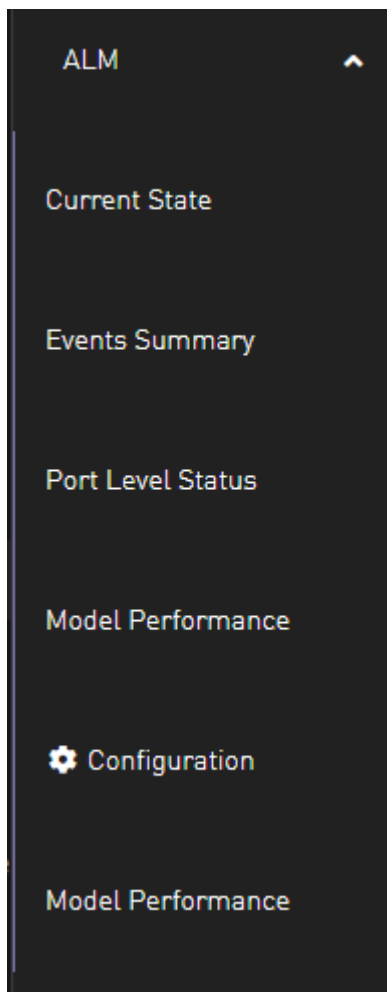
Name	Section name	Description
mode	Prediction	The mode can be either "active" or "shadow." In active mode, the ALM will enforce isolation/deisolation rules on all ports which predict to fail except those listed in the "expect" list.

Name	Section name	Description
		In shadow mode, the ALM will enforce isolation/deisolation rules on the ports listed in the "except" list, and predict to fail.
except_list	Prediction	Includes the ports that receive the opposite treatment compared to the mode. the expect list saved in location /opt/ufm/files/conf/plugin/alm/predict.csv Format: port_guid,port_number 0x1070fd03001769b4,1 0x1070fd03001769b4,3
mode	NOC	The mode can be either "active" or "shadow." In active mode, the ALM will enforce isolation/deisolation rules on all ports that considered as out of nominal condition except those listed in the "expect" list. In shadow mode, the ALM will enforce isolation/deisolation rules on the ports listed in the "except" list.
except_list	NOC	Includes the ports that receive the opposite treatment compared to the mode. the expect list saved in location /opt/ufm/files/conf/plugin/alm/noc.csv Format: port_guid,port_number 0x1070fd03001769b4,1 0x1070fd03001769b4,3
max_per_hour	Isolation	The maximum number of ports that can be isolated in a hour
max_per_week	Isolation	Maximum number of ports that can be isolated in a week
max_per_month	Isolation	Maximum number of the ports that can be isolated in a month
min_links_per_switch_pair	Isolation	Minimum links between two switches to perform isolation
min_active_ports_per_switch	Isolation	Minimum number of active ports per switch before perform isolation

Name	Section name	Description
Deisolation_time	DeIsolation	The waiting time before deisolate the isolated port
max_per_hour	DeIsolation	The maximum number of deisolated port per hour
absolute_threshold_of_isolated_ports	Isolation	The maximum number of ports than can be isolated in one sample

ALM UI

After the successful deployment of the plugin, a new item is shown in the UFM side menu for the ALM plugin:

ALM ∧

Current State

Events Summary

Port Level Status

Model Performance

⚙ Configuration

Model Performance

Current State

This page displays a table presenting the current cluster status, outlining the following counts:

1. Number of ports

2. Number of isolated ports

3. Number of ports in active/shadow prediction mode

4. Number of ports in active/shadow NOC mode

5. Number of ports out of NOC

Current State

	Switch to Switch	Switch to Host	Total, currently
Number of ports	8(57.14%)	6(42.86%)	14(100%)
Number of isolated ports	1(7.14%)	0(0%)	1(7.14%)
Number of ports in active prediction mode	8(57.14%)	6(42.86%)	14(100%)
Number of ports in shadow prediction mode	0(0%)	0(0%)	0(0%)
Number of ports in active NOC mode	0(0%)	0(0%)	0(0%)
Number of ports in shadow NOC mode	8(57.14%)	6(42.86%)	14(100%)
Out of NOC	0(0%)	0(0%)	0(0%)

Viewing **1-7** of **7**

Events Summary

This page displays a table presenting a port count summary, outlining the following counts:

1. Number of isolated ports in the past hour, week, and month for 'host to switch' and 'switch to switch'.

2. Number of de-isolated ports in the past hour, week, and month for 'host to switch' and 'switch to switch'.

3. Number of isolation actions **not** taken from prediction by ALM in the past hour, week, and month for 'host to switch' and 'switch to switch'.

4. Number of isolation actions **not** taken from NOC by ALM in the past hour, week, and month for 'host to switch' and 'switch to switch'.

Events Summary

	Switch to Switch			Switch to Host		
	Past Hour	Past Week	Past Month	Past Hour	Past Week	Past Month
Number of isolation	1	2	2	0	0	0
Number of de-isolation	1	2	2	0	0	0
Number of isolation actions not taken from prediction	3	4	4	0	0	0
Number of isolation actions not taken from NOC	0	0	0	0	0	0

Viewing 1-4 of 4

Port Level Status

This page displays a table presenting the cluster ports.

Port Level Status

Node ID	Port Number	Prediction Mode	NOC Mode	Isolation Status	Type	Last Recommendation	Last Recommendation Time	Last Action	Last Action Time	Last Reason
0x248a070300bee960	9	active	shadow	Health	switch_switch	NA	N/A	NA	N/A	NA
0x043f7203001949c0	3	active	shadow	Health	switch_switch	NA	N/A	NA	N/A	NA
0x043f7203001949c0	5	active	shadow	Health	switch_switch	NA	N/A	NA	N/A	NA
0x98039b0300dfef80	13	active	shadow	De_Isolate	switch_switch	NA	N/A	De-Isolation	N/A	UFM_User
0x0c42a103005bf438	1	active	shadow	Health	host_switch	NA	N/A	NA	N/A	NA
0x1070fd03001769b4	1	active	shadow	Health	host_switch	NA	N/A	NA	N/A	NA
0x98039b030000e5c6	1	active	shadow	Health	host_switch	NA	N/A	NA	N/A	NA
0x98039b0300dfef80	11	active	shadow	Health	switch_switch	NA	N/A	NA	N/A	NA
0x043f72030020b630	1	active	shadow	Health	switch_switch	NA	N/A	NA	N/A	NA
0x248a070300bee960	7	active	shadow	Health	switch_switch	NA	N/A	NA	N/A	NA

Viewing 1-10 of 14

Model Performance

Different metrics can measure the performance and accuracy of the ALM model, these metrics will be exposed to the user through the table as follows

Model Performance

<div>Displayed Columns ▾ | CSV ▾</div>

| | Switch to Switch | | | | | | | Switch to Host | | | | | | |
	Precision	Recall	True Positives	False Positives	False Negatives	Undetermined	High PLR events	Precision	Recall	True Positives	False Positives	False Negatives	Undetermined	High PLR events
all time	0.14	0.8	4	25	1	66	7	0	0	0	0	0	0	5
past month	0	0	0	3	0	0	7	0	0	0	0	0	0	5

Viewing **1-2** of **2** |◀ ◀ ▶ ▶| 10 ⌄

Configuration

This page displays ALM plugin configuration update method.

The ALM configuration is divided into four sections:

- General Configurations

General Configurations ⌄

Max Number of Isolated Ports by ALM

`100`

Max Number of Isolated by ALM Per Time Window

`5` Hour `20` Day `40` Week `60` Month

Max Number of De-isolated by ALM Per Time Window

`20` Hour

Minimum Time Port Should be Healthy Before De-isolation

`1` Minutes

Minimum Number of Links Between Two Switches

`1`

Minimum Number of Active Ports Per Switch

`4`

- Prediction Mode

- NOC Mode

- ML Model Configurations

ALM Jobs

The table presented below displays the names and descriptions of ALM jobs. These jobs are designed to predict the ports that require isolation/de-isolation. Upon enabling the ALM plugin, these ALM jobs run periodically.

ALM Job Name	Description	Frequency
Port_hist	By using **the** low frequency bit error histogram counters, the ALM job identifies the ports that will be monitored at high frequency in the next time interval. The job generates an output file that is later read by the high frequency telemetry monitoring job. It prioritizes **l**inks that are more susceptible to failure**.**	7 mins
Low_freq_predict	Predicts the likelihood of a port failure by analyzing input data from low frequency telemetry, while only utilizing physical layer counters. The prediction works for isolated ports as well. The resulting output from this task serves as a critical input for determining whether to isolate or de-isolate ports.	7 mins