Routing and MR integration

Overview

Multi-Rail has changed how LNet views the world. Prior to Multi-Rail, each NID represented one unique peer on a network. There was no concept that multiple NIDs can identify the same peer. After Multi-Rail a peer can have multiple NIDs on the same or different network, and LNet has become aware that these NIDs reference the same peer. This creates a disconnect with the routing infrastructure currently in place. This is highlighted in two recent LUs

(at the time of this writing): U-11143 - Multi-Rail/Dynamic Discovery break LNet router checker and asymmetric route failure detection CLOSED

and C LU-11144 - Dynamic Discovery is not triggered for router peers CLOSED

The routing infrastructure needs to deal at the peer level and not the peer NI level.

When adding a route it takes the following form:

lnetctl route add --net <remote network> --gateway <local gateway NID>

The code currently adds this gateway as a standard peer, which is also kept on a gateway list.

Multi-Rail changes the way we deal with peers such that a peer is composed of multiple peer_nis. However, this infrastructure doesn't extend to the routing logic.

Proposed Changes

These are a set of proposed changes to align LNet's routing infrastructure with Multi-Rail.

- 1. When a route is added the gateway should be automatically discovered, the same as when a peer is first communicated with. This will allow the node to know all the interfaces it can reach the gateway on.
- 2. When a gateway aliveness is changed from dead back to alive again, it should be rediscovered, in case its interface list has changed.
- Instead of maintaining a separate mechanism to check if a gateway is alive, currently implemented in the router checker code, use the discovery mechanism.
 - a. This will consolidate the ping generation code to one area instead of having duplicate code doing the same functionality.
- Modify the router code to deal with each gateway as a peer with multiple interfaces. This will involve modifying the code which checks if the router has dead interfaces.
 - a. if a router has multiple interfaces on the same network, and one of them is down, the router is still usable.
 - b. The router is not usable only if it can not route a message from one network to another. This is not equivalent to having one of the interfaces down.
 - i. This implies the avoid_asym_router_failure logic needs to be reworked.
- 5. Allow multiple routes to the same remote network over multiple gateways.

Conclusion

These changes will integrate the router handling more closely with the Multi-Rail code and will avoid issues where an MR router is not discovered properly or identifying that a router is dead when it really is not.