High Level Design

Overview

This design attempts to add two new features

1. Allow Lustre clients to continue using servers even if they change their NIDs during a boot cycle
2. Allow new Lustre clients to mount the file system on networks which have been added dynamically

Requirements

Test Case ID	Requirement ID	Requirement Description
test_dynamic_nids_06	LU-10360-07	The system shall allow mounting a client on a dynamically added netowrk.
test_dynamic_nids_04	LU-10360-06	All MGCs must have dynamic_nids set in order for the feature to work. Test dynamic_nids set on Client but not on MDT
test_dynamic_nids_05	LU-10360-06	All MGCs must have dynamic_nids set in order for the feature to work. Test dynamic_nids set on MDT but not the client
test_dynamic_nids_07	LU-10360-08	Disabling discovery on the client while keeping discovery on on the servers, shall not impact the functionality of this feature.
test_dynamic_nids_08	LU-10360-09	Disabling discovery on the client and server will disable the dynamic mount feature.
test_dynamic_nids_09	LU-10360-10	Dynamically adding and deleting NIDs should result in an updated IR log propagated to existing clients
test_dynamic_nids_01	LU-10360-01	Eanble/Disable dynamic NIDs via dynamic_nids parameter
test_dynamic_nids_02	LU-10360-03	Allow clients to continue using servers which have changed their IP address during a boot cycle
test_dynamic_nids_03	LU-10360-05	All MGCs must have dynamic_nids set in order for the feature to work.

Feature One Overview

When a Lustre client mounts a server it receives the lustre log describing the servers and their NIDs. It will then attempt to connect to each one of the servers on the first reachable NID. If any of these connections fail, the mount fails. Subsequently it is expected that the server NIDs will remain static. However, this is not true in a dynamically assigned IP address environment. For example, if a server reboots and its NIC gets assigned a different IP address, when lustre comes up the NID it will use will be different. The MGS will send an Imperative Recovery message to the client informing it of the new server NID. The client, however, will not use this server because there is a discrepancy between the NID in the lustre log and the one reported by the Imperative Recovery protocol.

Solution

- An IR log is sent with the current NID information of the server. When the client receives the IR log it checks the entry there against what it has already stored from the llog.
  - If the entry is not there, then add a new connection to the import
  - If the entry is there but the NID list is different, then update the NID information with the latest NID information provided in the IR log.
- Since allowing new servers NIDs previously unknown during the initial mount to be used, it could be considered a security risk on some sites.
  - Add a new File system level module parameter to enable this feature. The feature is disabled by default.
    - `lctl set_param mgc.*.dynamic_nids=1`

Feature Two Overview

LNet allows the addition of new NIDs and new Network. For example, if a server initially starts with tcp449. Clients which are only on that network can mount the server. However, during the lifetime of the server other networks can be added, ex: tcp450, tcp451, etc. Since these NIDs were not part of the initial configuration, they will not be recorded in the llog. Client which are only on these networks will not be able to mount the File System, since they will not know how to reach them based on the NIDs provided in the llog.

This feature is intended to allow this scenario. It can be used in a multi-tenancy environments, where clients need to be segregated, with no traffic allowed between the different clients.

Solution

- When a NID (possibly for a new network) is added dynamically on a server via lnetctl utility, all peers including the MGS are informed of the addition of the NID.
    - Update the MGS internal IR log with the new NID information
    - send the IR log to the currently mounted clients.
- When a new Client mounts the server:
    - the MGS will send it the IR log.
    - Prior to this patch the server will only check the llog. The IR log is sent and processed but no new connections are added.
        - This feature allows the client to process the IR log and create new connections only at mount time.
        - After mount, new IR log notifications from the MGS will not alter the connections created unless dynamic_nids is enabled on the servers and the clients.
    - The client will update its own internal NID database with the NIDs provided in the IR log and will attempt to connect to the servers on the first reachable NID.
        - If the client is restricted on a specific network, then only NIDs reported in the IR log which are on that network are processed.
        - If the client has LNet discovery disabled, then it will only use the first reachable NID reported in the IR log

Description of Behavior

Test Case ID	Requirement ID	Design
test_dynamic_nids_06	LU-10360-07	This feature adds the ability for clients to mount servers on networks or NIDs which has been dynamically added post server bring up. As an example, servers can initially be configured with a tcp network. Then a new network can be added after the servers are mounted using lnetctl: lnetctl net add --net tcp441 --if eth1 A new client which has not mounted the file system yet and only exists on the tcp441 network, can then successfully mount the file system on tcp441 network. This is accomplished as follows: The MGS registers with LNet for updates whenever a newly added network or NID are dynamically added to LNet or when through discovery LNet finds out about an update to a peer's NID list. The notification causes the MGS to update its Imperative Recovery log with the new NIDs. After updating the IR log, a notification is sent to the clients currently mounting the FS. When new clients connect to the MGS, the IR log is sent to the client. The client then uses that IR log to create connections instead of relying only on the llog, which would be out of date at this point. NOTE: This feature relies on LNet dynamic discovery feature.
test_dynamic_nids_04	LU-10360-06	
test_dynamic_nids_05	LU-10360-06	
test_dynamic_nids_07	LU-10360-08	When a NID or a network is added on the servers, since discovery is on, it'll advertise the updated list of NIDs to all its peers. This will result in the MGS updating its IR log. When a client mounts the FS, even though it doesn't have discovery enabled, it will receive the updated IR log which will contain all the server NIDs. It will then iterate through the NIDs and use the first one it can reach. If network restriction is on using the -o network option, then it will only attempt to create connections to NIDs on that network.
test_dynamic_nids_08	LU-10360-09	The new NIDs added dynamically will not be reported to the MGS. The MGS's IR log will therefore not be updated. When a client mounts the MGS, the IR log will not contain the dynamically added NIDs for the other servers, OSS, MDS, etc. And the client will not be able to establish communication with these causing a mount failure. The only exception to this is if all the servers are configured on the same node. In this case discovery has no impact on this feature.
test_dynamic_nids_09	LU-10360-10	

test_dynamic_nids_01	LU-10360-01	Standard Lustre sysfs module parameters setting
test_dynamic_nids_02	LU-10360-03	An IR log is sent with the current NID information of the server. When the client receives the IR log it checks the entry there against what it has already stored from the llog. If the entry is not there, then add a new connection to the import If the entry is there but the NID list is different, then update the NID information with the latest NID information provided in the IR log. Since allowing new servers NIDs previously unknown during the initial mount to be used, it could be considered a security risk on some sites. Add a new File system level module parameter to enable this feature. The feature is disabled by default. To Enable the feature: lctl set_param mgc.*.dynamic_nids=1
test_dynamic_nids_03	LU-10360-05	