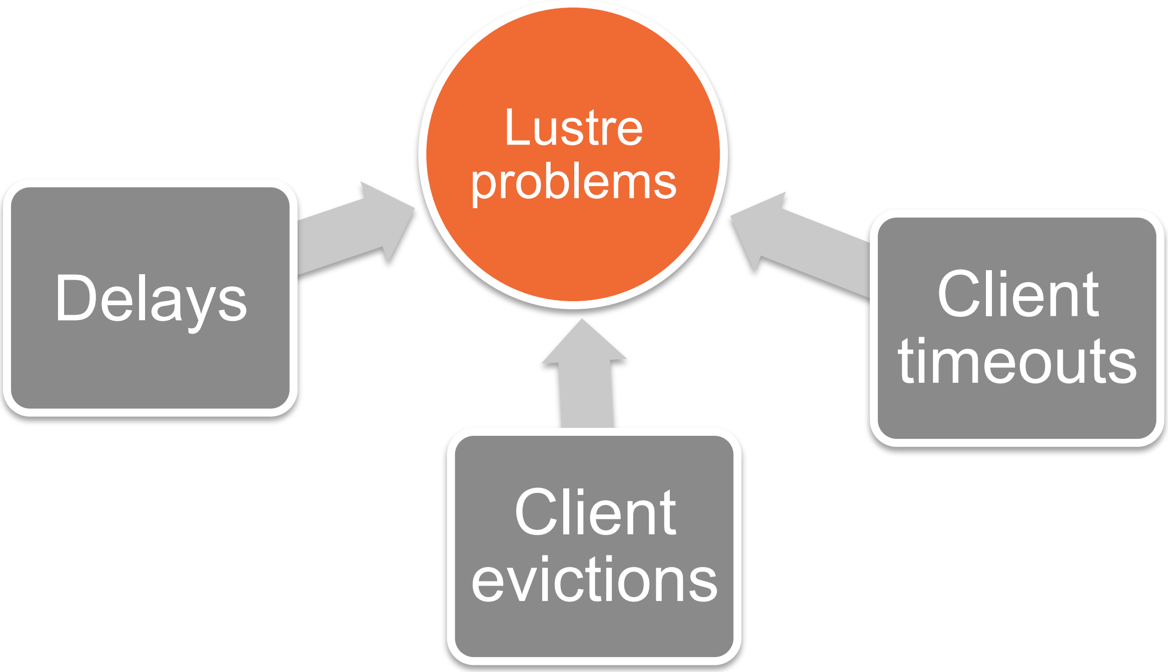
Agenda

- Problem description
- Why it seen ?
- Parameters to help
- Additional fixes and future work

Problem description

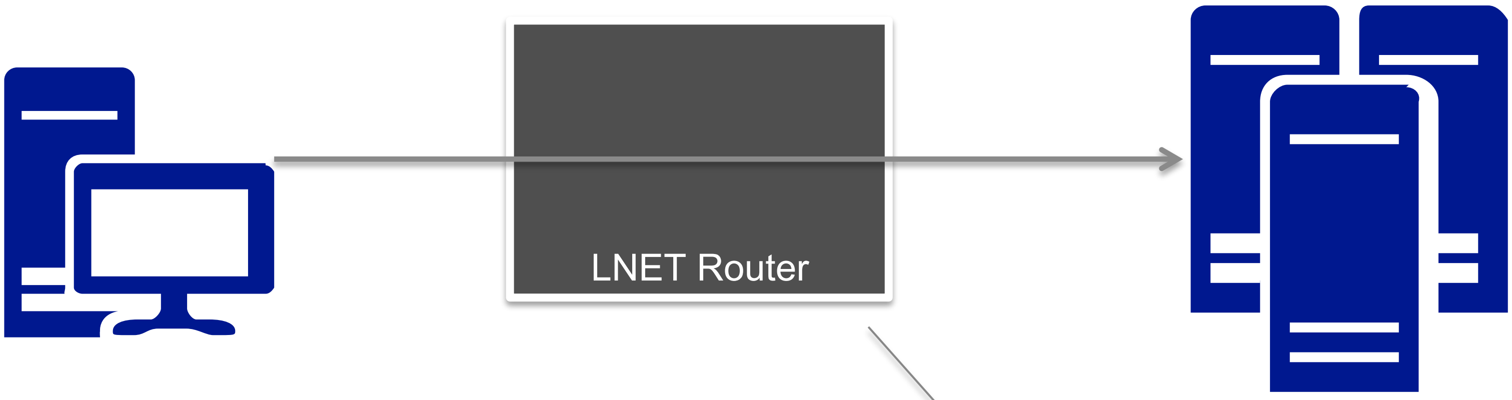
Users may seen a network fails number increased in case LNet router used between client and server.



Why it is seen?

Why it seen ?

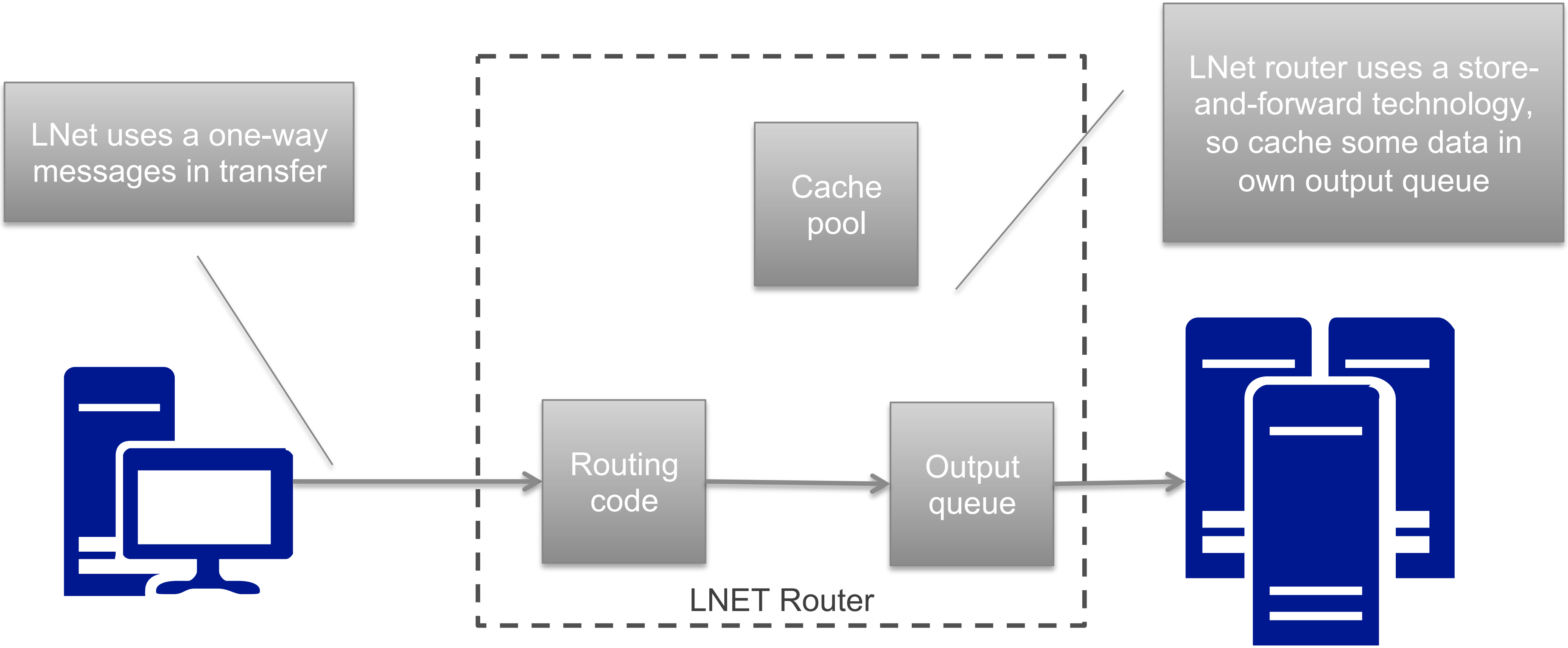
PtlRPC view



PtlRPC know just about destination, router is invisible on that layer

Why it seen ?

LNet router internals



LNet uses a one-way messages in transfer

Cache pool

LNet router uses a store-and-forward technology, so cache some data in own output queue

Routing code

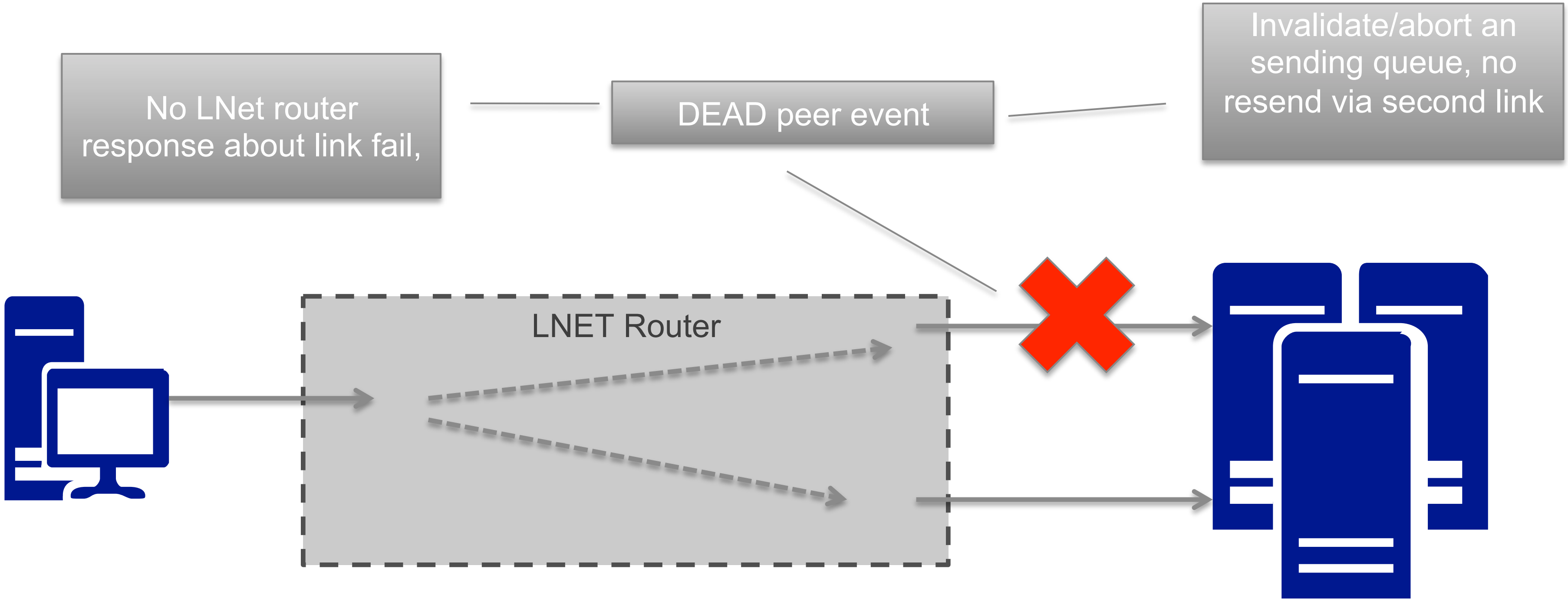
Output queue

LNET Router

* Backside effect of it situation is PtlRPC timeout covers two transfers and router queue.

Why it seen ?

LNet router view



No LNet router response about link fail,

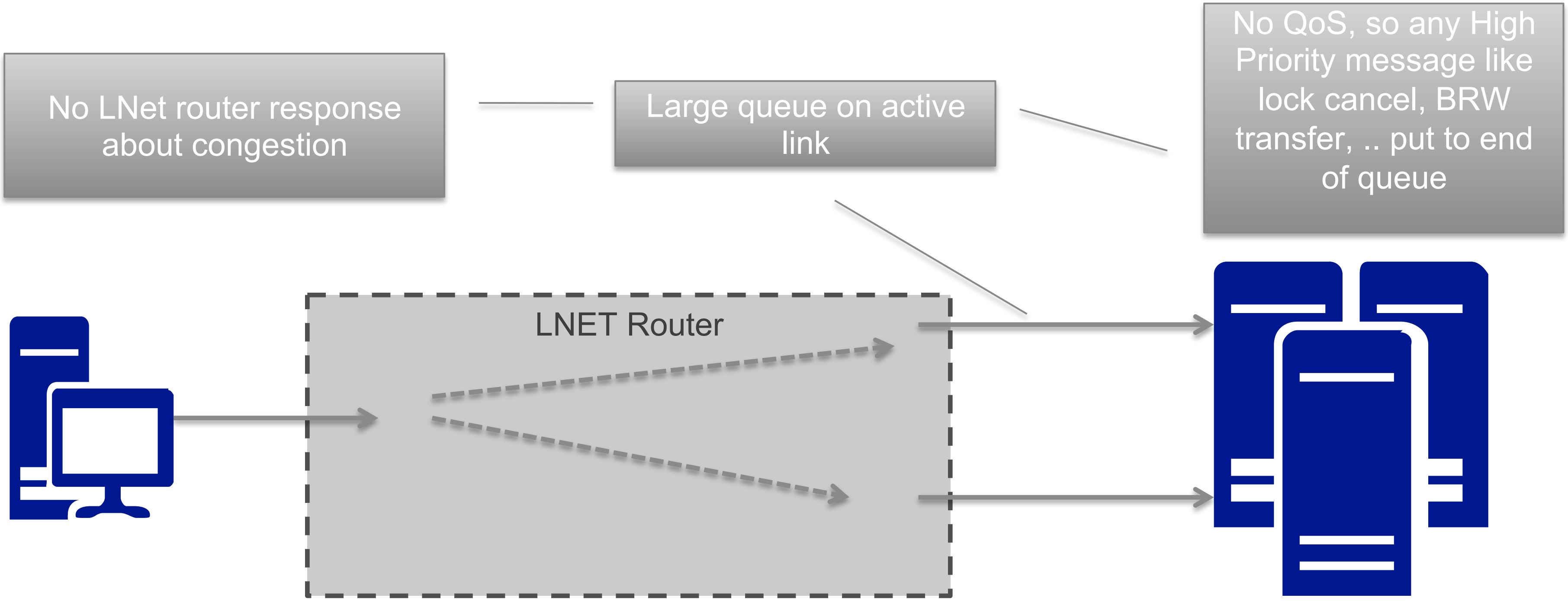
DEAD peer event

Invalidate/abort an sending queue, no resend via second link

Black hole

Why it seen ?

LNet router view



No LNet router response about congestion

Large queue on active link

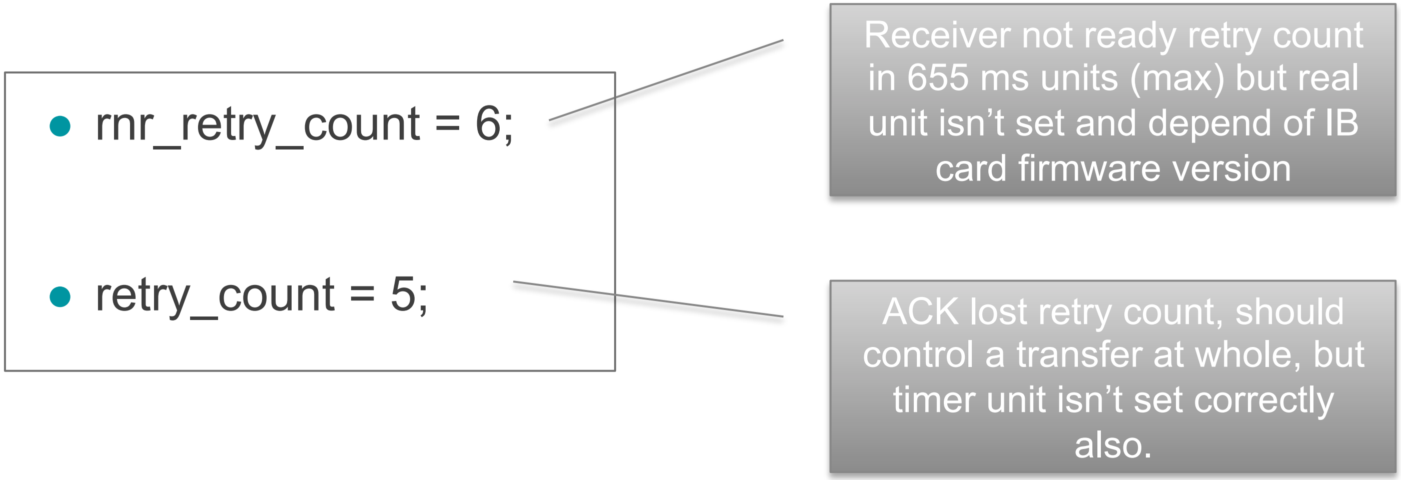
No QoS, so any High Priority message like lock cancel, BRW transfer, .. put to end of queue

Large router queue (potentially) and lack of congestion (flow) control will produce an unpredictable timeouts.

Why it seen?

Unpredictable timeouts - IB LND issues

OFED uses a an QP (queue pair) as object similar to the socket, so timeouts set in QP base.



- rnr_retry_count = 6;
- retry_count = 5;

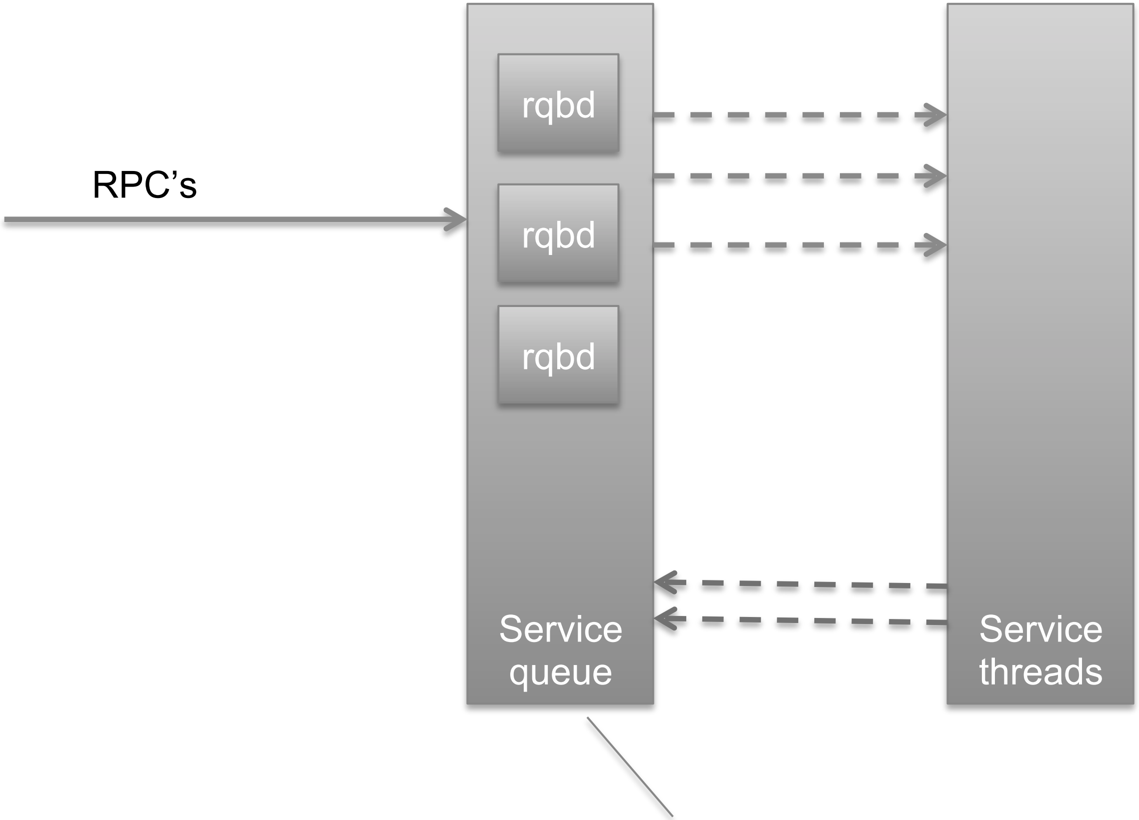
Receiver not ready retry count in 655 ms units (max) but real unit isn't set and depend of IB card firmware version

ACK lost retry count, should control a transfer at whole, but timer unit isn't set correctly also.

As both timers don't an initialize correctly, these timeouts unpredictable.

Why it seen?

Unpredictable timeouts - Server side over buffering problem



Service queue may grow dramatically in case service threads stick in processing (like a lock cancel waiting), so system cache will flushed.

How a Lustre Resilience may increased ?

Parameters to help

LNet operates with two timeouts

- **timeout** (default 50sec) - Transmission timeout: LND finds this link failure and triggers reconnect; Tx descriptors will be aborted if re-connect fails.
    - **peer_timeout** (default 180sec). global peer alive timeout, if the last alive (succeeded) event + peer_timeout exceeded, marks peer dead, returns an error to upper layer.

Parameters to help

Upper layers have a several timeout settings

- **at_min -**This is the minimum processing time that a server will report back to a client.
    - **at_max -** This is the maximum amount of time that a server can take to process a request. If a server has reached this value then the RPC times out.

Parameters to help

Timeout settings unbalanced now. Routing change event needs a “DEAD PEER” event hit, and client should send a new request after it. So we need to lower a peer_timeout and increase a at_min to cover link dead situation.

Seagate suggestion is peer_timeout should be set to 2xtimeout and at_min should be covered a peer_timeout with some external time. Based on maximal IB resend timer it should be something like

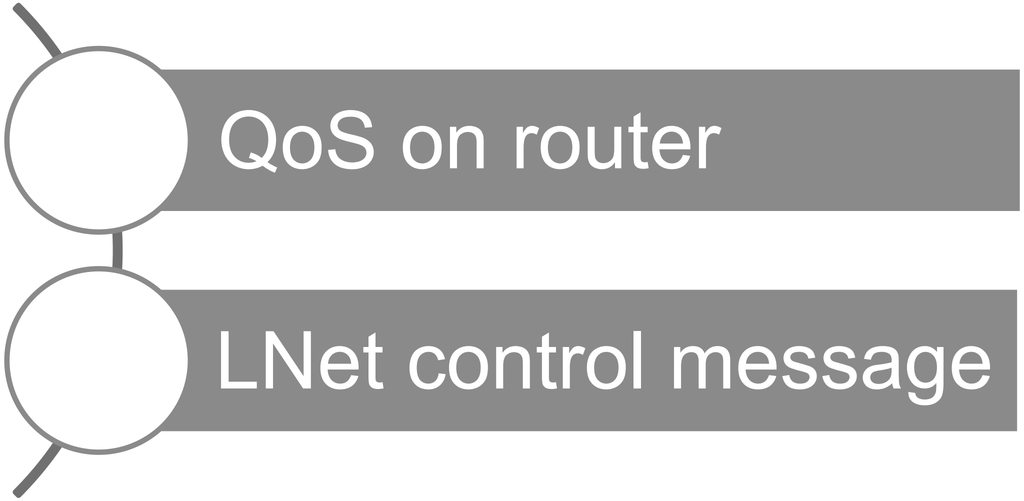
- as **IB timeout** maximal time is 6*0.65s = **3.5s**, so **lnd timeout** should be set to **4s** it will cover resend or rnr timeouts and have chance to deliver a message.
- **peer_timeout** is 2.5xLND timeout to have chance to have one reconnect to same peer and resend own messages or have a decision about peer dead. So it’s need to set **10s**.
- **at_min** – should covered an one “DEAD peer” event and have chance to send a reconnect request to different link as ptlrpc designed to use same connection first. So it should be set to **peer_timeout** + **lnd_timeout** to cover LNet reconnect = **15-20s**.
- **ldlm_enqueue_min** covers a blocking ast timeout and a refresh a lock timeout, should to be set to **85s**

Parameters to help

Additional network settings

- As client will reconnect to same NID first, we need choose a different routing priority for different network links.
- Next NID in network parameters should be reachable via different router.
- Peer_timeout need reduced if link fail mostly permanent or want to include an admin to fix.

Additional fixes and future work

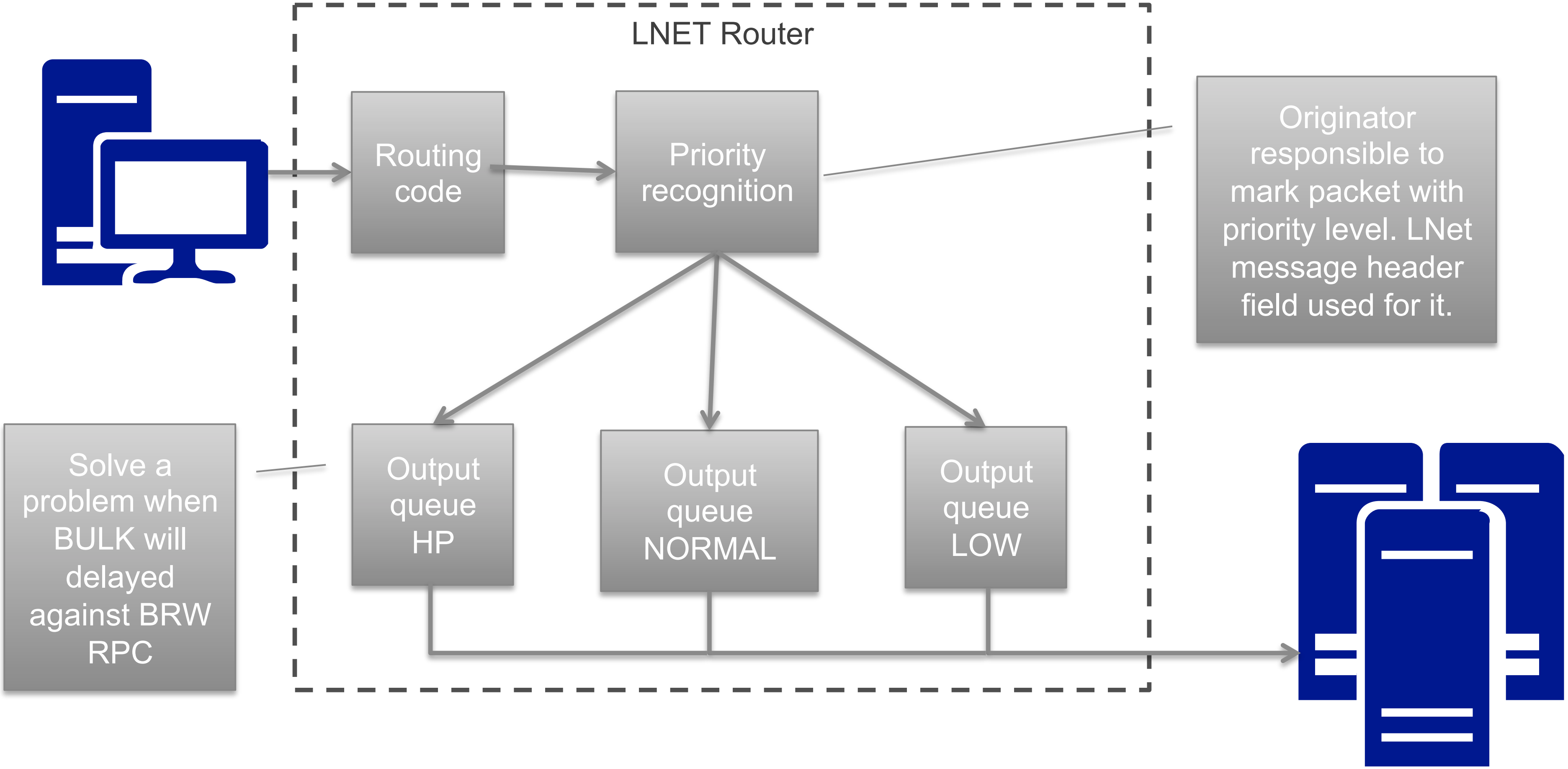


QoS on router

LNet control message

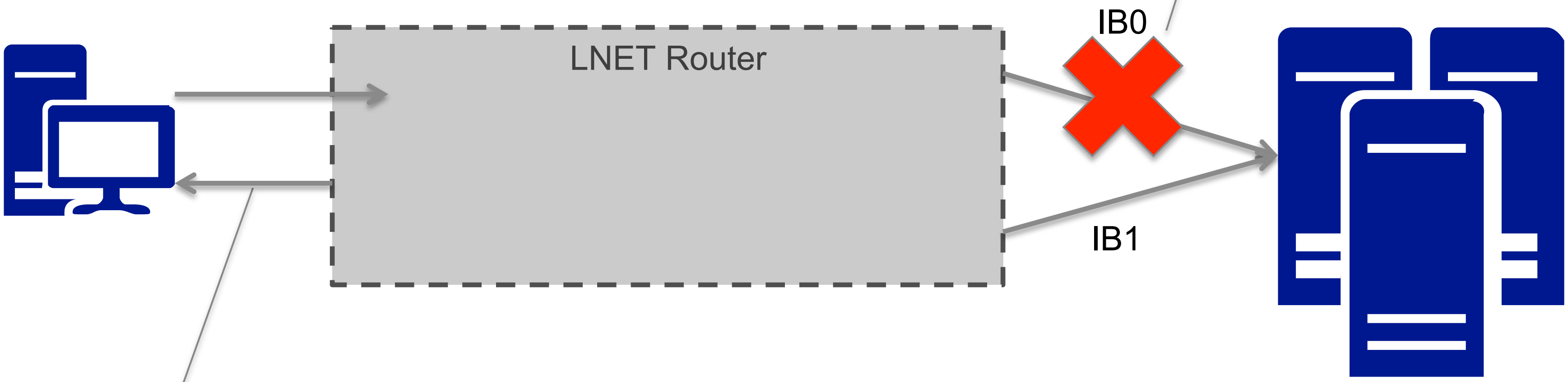
Additional fixes and future work

QoS on router


LNET Router
Routing code
Priority recognition
Originator responsible to mark packet with priority level. LNet message header field used for it.
Solve a problem when BULK will delayed against BRW RPC
Output queue HP
Output queue NORMAL
Output queue LOW

Additional fixes and future work

LNet control message



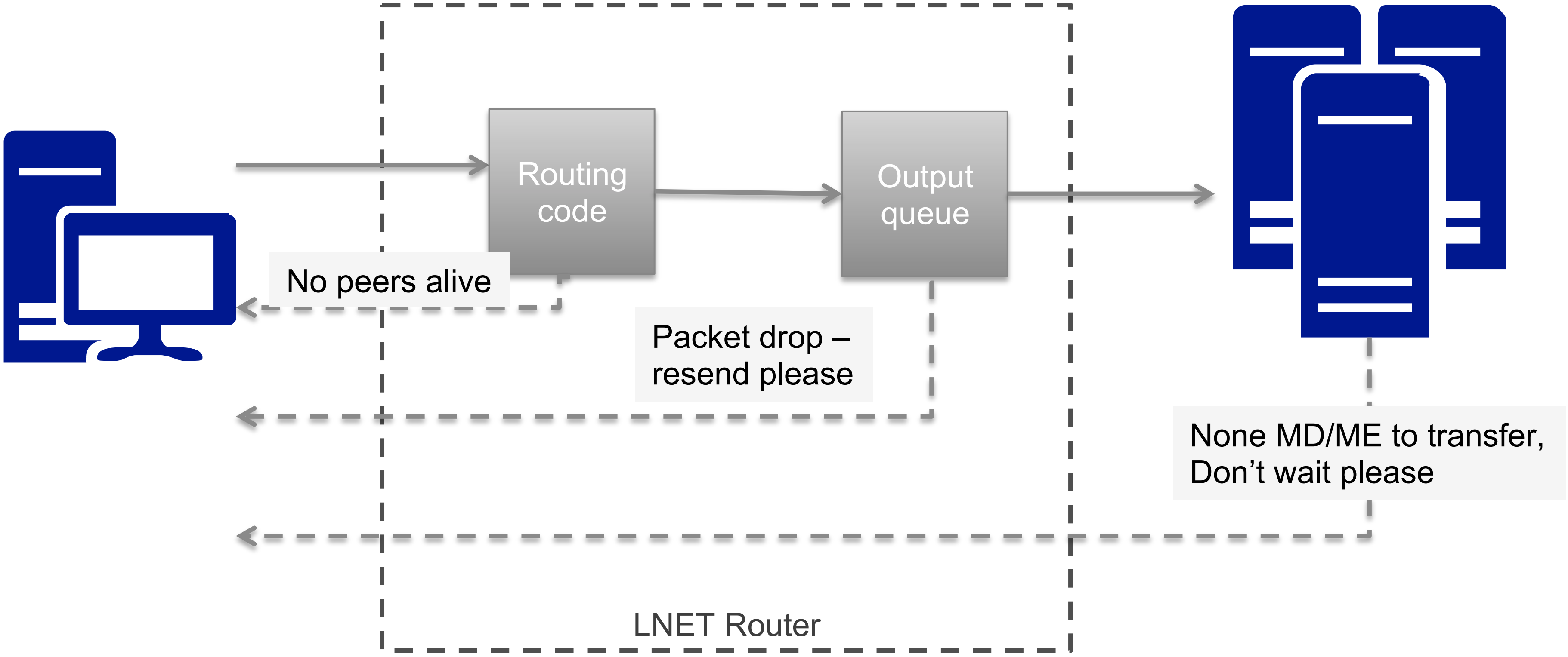
Link failed and alternative route needs choose

LNet control message – “ROUTE unreachable”, So sender need c

Additional fixes and future work

LNet control message – design notes

LNet control message similar to the ICMP protocol in TCP stack used to report about network issues or routing changes


Routing code
Output queue
No peers alive
Packet drop – resend please
None MD/ME to transfer, Don’t wait please
LNET Router

Implementation quite simple – new LNet msg generated in lnet_finalize() function in similar to the ACK.

Questions ?