Contribution ID: **73**   Type: **not specified**

Challenges of the RDMA subsystem

Monday 9 September 2019 17:45 (45 minutes)

The RDMA subsystem in Linux (drivers/infiniband) is now becoming widely used and deployed outside its traditional use case of HPC. This wider deployment is creating demand for new interactions with the rest of the kernel and many of these topics are challenging.

This talk will include a brief overview of RDMA technology followed by an examination & discussion of the main areas where the subsystem has presented challenges in Linux:

Very complex user API. An overview of the current design, and some reflection on historical poor choices

The DMA from user space programming model and the challenge matching that to the DMA API in Linux

Development of user space drivers along with kernel drivers

Delegation of security decisions to HW

Interaction with file systems, DAX, and the page cache for long term DMA

Inter-operation with GPU, DMABUF, VFIO and other direct DMA subsystems

Growing breadth of networking functionality and overlap with netdev, virtio, and nvme

Fragmentation of wire protocols and resulting HW designs

Placing high performance as paramount and how this results in HW restrictions limiting the architecture and APIs of the subsystem

The advent of new general computation acceleration hardware is seeing new drivers proposed for Linux that have many similar properties to RDMA. These emerging drivers are likely to face these same challenges and can benefit from lessons learned.

RDMA has been a successful mini-conference at the last three LPC events, and this talk is intended to complement the proposed RDMA micro-conference this year. This longer more general topic is intended to engage people unfamiliar with the RDMA subsystem and the detailed topics that would be included in the RDMA track.

The main goal would be to help others in the kernel community have more background for RDMA and its role when making decisions. In part this proposal is motivated by the number of times I heard the word 'RDMA' mentioned at LSF/MM. Often as some opaque consumer of some feature.

Jason Gunthorpe is a Sr. Principal Engineer at Mellanox and has been the co-maintainer for the RDMA subsystem for the last year and a half. He has 20 years' experience working with the Linux kernel and in RDMA and InfiniBand technologies.

I agree to abide by the anti-harassment policy

Yes

Primary author: Mr GUNTHORPE, Jason (Mellanox Technologies)

Presenter: Mr GUNTHORPE, Jason (Mellanox Technologies)

Session Classification: LPC Refereed Track