IETF-84 TCPM Work Group
TCP Option for SMC-R

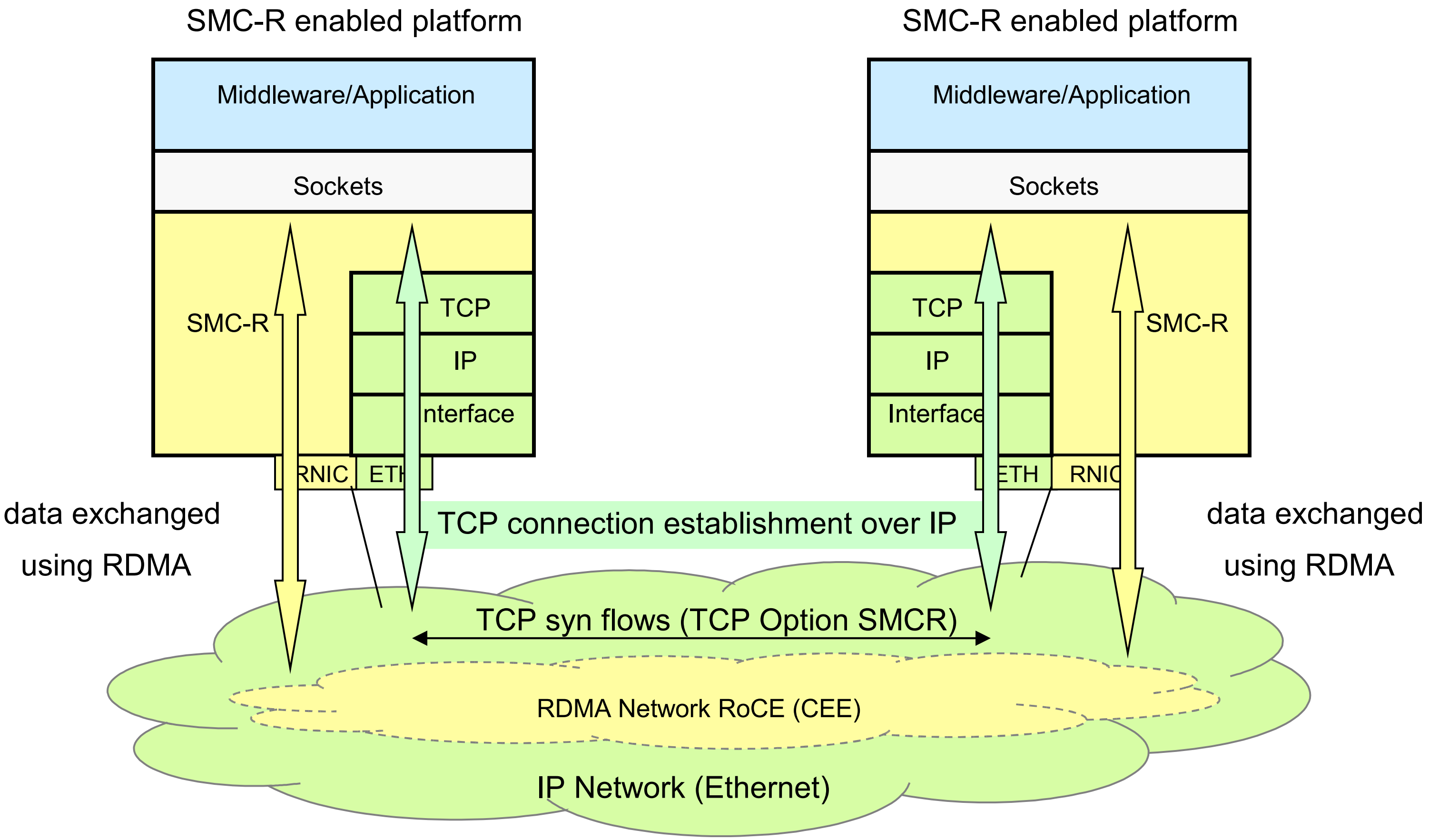
Reference Drafts:
draft-ietf-tcpm-experimental-options-01.txt
draft-fox-tcpm-shared-memory-rdma-00.txt

Jerry Stevens (sjerry@us.ibm.com)

Background

- RDMA over Converged Ethernet (RoCE) allows conventional IP traffic and RDMA protocols to flow over the same wire (Ethernet)
- Shared Memory Communications over RDMA (SMC-R) is a protocol that allows TCP sockets applications to transparently exploit RDMA (RoCE)
- SMC-R is a hybrid solution that:
  - Uses TCP connection (3-way handshake) to establish SMC-R connection
  - Switching from TCP to “out of band” SMC-R is controlled by a TCP Option (Experimental Option “magic number”)
  - SMC-R “rendezvous” (RDMA attributes) information is then exchanged within the TCP data stream
  - Socket application data is exchanged via RDMA
  - TCP connection remains active (controls SMC-R connection)
  - This model preserves many critical existing operational and network management features of TCP/IP (see backup charts)

Dynamic Transition from TCP to SMC-R



Dynamic (in-line) negotiation for SMC-R is initiated by presence of TCP Option (SMCR)

TCP connection transitions to SMC-R allowing application data to be exchanged using RDMA

Requirement

TCP Option Indicating SMC-R Capability

- Need the capability to communicate new TCP option during TCP/IP 3-way handshake ("in-line" syn flows)
- Must preserve "in-line" (TCP data stream) negotiation model (see backup)[1]
- Current SMC-R implementation uses TCP experimental option (253) with magic number "SMCR" (in EBCDIC) per draft-ietf-tcpm-experimental-options-01.txt)
- Would like to ensure that there are no other collisions with other uses of option 253

[1] Alternative approaches (e.g. static config, connection Mgr, etc.) were considered but would significantly diminish the TCP/IP operational and network management features of SMC-R

Objective

Exploit Final (Standard) TCP Experimental Options

- SMC-R has a dependency on draft-ietf-tcpm-experimental-options-01.txt
- Advocating to finalize and publish TCP Experimental Options “magic number” draft RFC (draft-ietf-tcpm-experimental-options-01.txt) **preferably as standards track**
- SMC-R to exploit (and will adopt or adjust to) the final standard
- As SMC-R evolves (based on adoption / acceptance) a request for a standard TCP option code point may be made in the future[1]

[1] Also interested in comments and feedback regarding SMC-R Information RFC draft-fox-tcpm-shared-memory-rdma-00.txt

Backup

References:

1. draft-ietf-tcpm-experimental-options-01.txt
2. draft-fox-tcpm-shared-memory-rdma-00.txt

What is RDMA?

- Remote Direct Memory Access is a technology that allows computers in a network to exchange data without involving the processor, cache or operating system of either computer.
- SMC-R is a “byte stream” protocol similar to the Transmission Control Protocol that exploits RDMA technology for TCP sockets based applications

How & Why TCP Connectivity

- Follows standard TCP/IP connection setup
- Dynamically switch to RDMA (SMC-R)
- TCP connection remains active (idle) and is used to control SMC-R connection
- Preserves critical operational and network management TCP/IP features such as:
  - Minimal (or zero) IP topology changes
  - Compatibility with TCP connection level load balancers
  - Preserves existing IP security model (e.g. IP filters, policy, VLANs, SSL etc.)
  - Minimal network admin / management changes