* **S**hared **M**emory **C**ommunications – **R**emote **D**irect **M**emory **A**ccess (**SMC-R**) is a new communication protocol aimed at providing transparent acceleration for sockets-based TCP/IP applications and middleware.
– RDMA [a] technology provides low latency, high bandwidth, high throughput, low processor utilization attachment between hosts.
- SMC-R utilizes RDMA over Converged Ethernet (RoCE) as the physical transport layer
- SMC-R is built on the following concepts:
  > RDMA enablement of the communications fabric
  > Partitioning a part of OS host real memory into buffers and using RDMA technology to access this memory
  > Establishing an 'out of band' connection over which data is passed to the partner peer using RMDA writes and signaling.
* RDMA support for z/OS over 10GbE RoCE Express (RDMA over Converged Ethernet [b]) through the use of the new SMC-R (Shared Memory Communications - Remote) protocol
- High speed inter communication facilitating data movement between zBC12/zEC12 Systems with z/OS using SMC-R.
- Improves network latency and throughput, reducing CPU overhead, z/OS network congestion and cost related to remote off stack data movement.
* Key attributes of RDMA:
- Enables a host to read or write directly from/to a remote host's memory ***without*** involving the remote host's CPU
- Registering specific memory for RDMA partner use
- Management of RoCE [c] fabrics can now readily be integrated with existing datacenter Ethernet fabrics (prior RDMA networks used InfiniBand)
- Interrupts are still required for notification (i.e. CPU cycles are not completely eliminated)
Note1: Initial deployment limited to z/OS<->z/OS communications, but goal to expand exploitation to additional operating systems and possibly appliances/accelerators.
Note2: RoCE can use existing Ethernet fabric but requires advanced Ethernet hardware (RDMA capable NICs and RoCE capable Ethernet switches)
* SMC-R is a protocol that allows TCP socket 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 (write operations)
– TCP connection remains active (controls SMC-R connection)
– This model preserves many critical existing operational and network management features of TCP/IP

SMC-R and RoCE will *not* support SNA, only TCP.

Each 10GbE RoCE Express feature can only be used by a single LPAR.

Why a **"Hybrid Protocol" ?**
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
– SMC-R (a sockets based solution for RDMA) means that host application software is not required to change, therefore all host application workloads can benefit immediately .

SMC-R will be part of z/OS V2.1 CommServer & there will be no additional charge.

Fast memory to memory data transfers with SMC-R (like Hipersockets between processors)

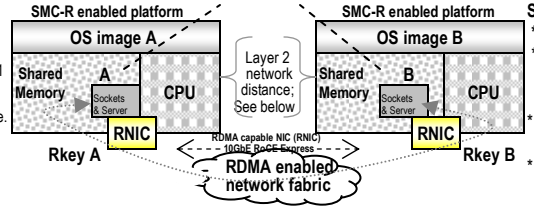
The combination of a new 10GbE RoCE Express feature and the SMC-R protocol deliver compelling improvements in network performance for processor to processor communication with reduced latency, improved throughput, and application transparency.

The Value
IBM SMC-R delivers faster communications with:
- Transparent application use
- Low CPU utilization and latency
- Leverages existing infrastructure
- Preserves TCP/IP security, management
- Standards based

The Users
Data or transaction intensive Installations:
- Ideal for banks, retail, healthcare, or financial institutions
- Superb for service providers
- Any organization needing to move data quickly between processors

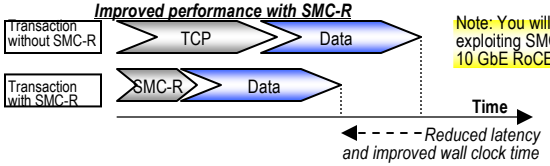
Any z/OS TCP sockets based workload can **seamlessly** use SMC-R without requiring any application changes.

Shared Memory Communications - RDMA



Layer 2 connectivity is required for server-to-server communication.
A communication fabric comprises the transmission media required to create a communication link between two computing nodes in a network.

Improved performance with SMC-R

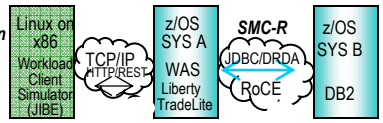


Note: You will still need OSA exploiting SMC-R using the 10 GbE RoCE Express feature.

⚠ No requirement for TCP/IP protocols/stack, sockets, etc. Low level APIs such as uDAPL, MPI or RDMA verbs allow optimized exploitation.
* ***For applications / middleware willing to exploit these interfaces*** *

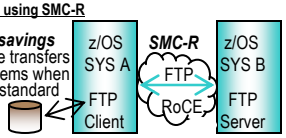
WAS to DB2 using SMC-R

40% reduction in overall transaction response time for WebSphere Application Server v8.5 Liberty profile TradeLite workload accessing z/OS DB2 in another system measured in internal benchmarks [d] .

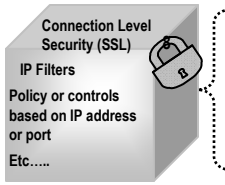


File Transfer (FTP) using SMC-R

Up to ***50% CPU savings*** for FTP binary file transfers across z/OS systems when using SMC-R vs standard TCP/IP [d].



The 10GbE RoCE Express feature is not defined as a CHPID and does not consume a CHPID number. Instead the 10 GbE RoCE Express feature is defined with PCIe-based definitions called PCIe Function IDs or PFIDs.



The hybrid nature of SMC-R (beginning with TCP/IP, then switching to SMC-R) allows all existing IP and TCP layer security features to automatically apply for SMC-R connections:
- Without requiring any changes from a customer perspective
- And without requiring these functions to be retrofitted into a new protocol.

SMC-R preserves existing security model

a. RDMA is the remote memory management capability that allows server to server data movement directly between applications without any CPU involvement.
b. RDMA over Converged Ethernet (RoCE – pronounced rockee) is a mechanism to provide efficient data transfer with low latencies on lossless Ethernet networks.
c. InfiniBand Trade Association (IBTA) standardized RDMA over Converged Ethernet (RoCE) in April 2010.
d. Based on projections and measurements completed in a controlled environment using latest code versions.

Two adapter types are required to exploit RoCE on z/OS V2R1:
- The QDIO OSA port to support the SMC-R protocol
- The RoCE adapter port (the RNIC) to support the data traffic.

- The SMC-R solution was created to meet all of the below objectives:
1. Performance! The primary value point of RDMA is performance advantages.
2. Key attributes - the following "Time to Market / Value" and Total Cost to Deploy / Operate attributes must also be considered:
- Full compatibility with existing socket based applications (no application changes required)
Can not regress key existing TCP/IP network operational or administrative attributes:
> Security (existing IP address based (IP filters) and TCP connection level security (SSL))
> HA (high availability and resiliency across redundant hardware - separate physical adapters)
> LB (server or clustering load balancing)
> Consumable (requirements to configure and deploy)
-. Preserve the existing network IP topology, "IP eco-system" and administration model (minimize disruption and required configuration changes and runtime / ongoing operational cost)
3. Interoperability (the need for a common protocol / solution).
* SMC-R uses a TCP connection to establish or terminate a connection and monitor heartbeat functions.
- Once the initial handshake is complete, communication uses SMC-R sockets-based communication.
- Socket application data is then exchanged via RDMA while the TCP connection remains active.

There are no application changes required. This is all handled inside z/OS' Communications Server.

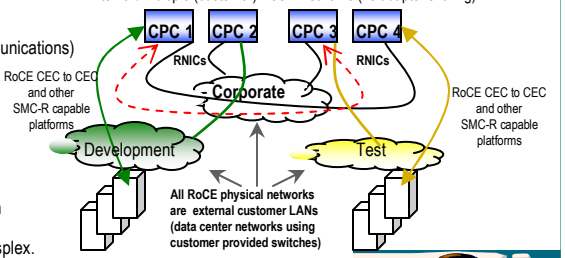
TCP connection load balancers are transparent to SMC-R connectivity.

The 10GbE RoCE Express feature is a new PCIe based network adapter on System z.

The IBM System z 10GbE RoCE Express feature is an RDMA Network Interface Card (RNIC).

SMC-R Can Benefit Many Workloads:
* SMC-R is designed for highly efficient, low latency data transfer. With lower CPU cost to move data, SMC-R can benefit network intensive and transaction oriented workloads. For instance, workloads that benefit include:
- Request response workloads; z/OS transactional workloads that generate frequent z/OS to z/OS interactive network traffic patterns such as:
  >WAS to DB2® and CICS® /CTG)
  >MQ to DB2, CICS and IMS™ (IMS Connect, Soap Gateway)
  >CICS to CICS (CICS IPIC communications)
- z/OS workloads that generate bulk data transfer workloads (FTP) or any transaction that exchanges "large messages"; (i.e. Connect Direct, SFTP, FTP, MQ FT).
- z/OS clients using z/OS Sysplex Distributor with VIPAROUTE when SMC-R is enabled on both z/OS clients and z/OS servers in the sysplex.

SMC-R Supported Configurations - External / Multiple (customer) RoCE Networks (no adapter sharing)



Server requirements
* Exclusive to zEC12 (with Driver 15E) and zBC12
* New 10 GbE RoCE Express feature for PCIe I/O drawer (FC#0411)
- Single port enabled for use by SMC-R
- Each feature must be dedicated to one LPAR
* Recommended minimum configuration two features per LPAR for redundancy
- Up to 16 features supported
* OSA Express – either 1 GbE *or* 10 GbE
- Must be Layer 2 connection
- Does not need to be dedicated to the LPAR
* Standard 10GbE Switch or point to point configuration supported.

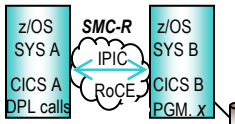
Flow control mechanisms similar to the window sizing and congestion controls of TCP connections exist with the SMC-R protocol.

RDMA technology is now available on Ethernet - RDMA over Converged Ethernet (RoCE).

Exploit RDMA over Converged Ethernet (RoCE) with **qualities of service support** for dynamic failover to redundant hardware.

CICS to CICS IP Communications using SMC-R

Up to 48% reduction in response time and up to 10% CPU savings for CICS transactions using DPL (Distributed Program Link) to invoke programs in remote CICS regions in another z/OS system via CICS IP interconnectivity (IPIC) when using SMC-R vs. standard TCP/IP [d].



Note: a modeled CICS workload driving a CICS transaction that performs 5 DPL calls to a CICS region on a remote z/OS system, using 32K input/output containers.

Note, RDMA is not IP routable. Both peer hosts must be within the same Layer-2 IP network.

SMC-R exploitation of RoCE does not require special Ethernet (RoCE-capable) switches.

The distance from the 10GbE RoCE Express port to the 10GbE switch is limited to 300 meters With OM3 fiber cable. The latency advantages of RDMA are diminished when travelling long distances,And so RDMA performs best when used within datacenter distances. RoCE Express features connected to a single 10GbE switch are preferable. Therefore the distance between two different servers would not exceed 600 meters with a switch in the middle.