SDC
STORAGE DEVELOPER CONFERENCE
SNIA ■ SANTA CLARA, 2014

iSCSI Protocol Advancements from IETF Storm WG

Mallikarjun Chadalapaka Microsoft	Fred Knight NetApp	David Black EMC
Julian Satran	Kalman Meth IBM	

Agenda

- IETF Context – Why are we here
- iSCSI Protocol – Recap the key concepts
- iSCSI Standards – Navigational aid
- RFC 7144 – What did we add
- RFC 7143 – What did we tweak
- Wrap-up – So you have questions

IETF Context – Why are we here

Storm

- IETF “Storm” (Storage maintenance) Working Group was chartered to make minor adjustments/improvements to multiple storage-related standards (“RFCs”) that were originally published in “IPS” (IP Storage) Working Group.
  - Significant Changes or v2 protocol work was out of scope
  - Preserving backwards compatibility to existing specs was an explicit charter objective
- As of June 2014, Storm WG completed its planned work
- Note that what authors present here is really a product from the entire Storm WG

RFCs from Storm WG & Today's Focus

RFC	Title	Publication Date
RFC 6172 (was draft-ietf-storm-ifcp-ipn133-updates)	Deprecation of the Internet Fibre Channel Protocol (iFCP) Address Translation Mode	2011-03
RFC 6173 (was draft-ietf-storm-ifcpmib)	Definitions of Managed Objects for the Internet Fibre Channel Protocol (iFCP)	2011-03
RFC 6580 (was draft-ietf-storm-rddp-registries)	IANA Registries for the Remote Direct Data Placement (RDDP) Protocols	2012-04
RFC 6581 (was draft-ietf-storm-mpa-peer-connect)	Enhanced Remote Direct Memory Access (RDMA) Connection Establishment	2012-04
RFC 7143 (was draft-ietf-storm-iscsi-cons)	Internet Small Computer System Interface (iSCSI) Protocol (Consolidated)	2014-04
RFC 7144 (was draft-ietf-storm-iscsi-sam)	Internet Small Computer System Interface (iSCSI) SCSI Features Update	2014-04
RFC 7145 (was draft-ietf-storm-iser)	Internet Small Computer System Interface (iSCSI) Extensions for the Remote Direct Memory Access (RDMA) Specification	2014-04
RFC 7146 (was draft-ietf-storm-ipsec-ips-update)	Securing Block Storage Protocols over IP: RFC 3723 Requirements Update for IPsec v3	2014-04
RFC 7147 (was draft-ietf-storm-iscsimib)	Definitions of Managed Objects for the Internet Small Computer System Interface (iSCSI)	2014-04
RFC 7306 (was draft-ietf-storm-rdmap-ext)	Remote Direct Memory Access (RDMA) Protocol Extensions	2014-06

RFC 7143 & 7144: Goals

- 7143: iSCSI spec consolidation
  - Goal: pulling together about half a dozen older RFCs into one coherent spec, making “minor” modifications to improve interop, and obsoleting a few specific unimplemented features

1. 7144: SAM-5 compliance of iSCSI
   - Goal: Extending iSCSI protocol to be a SAM-5-compliant storage transport protocol, negotiable at a session granularity

iSCSI Protocol – Recap the key concepts

iSCSI?

iSCSI is a client-server SCSI transport protocol, just like FCP*

iSCSI can run on any physical network that TCP/IP can run on – Ethernet, IB*,..

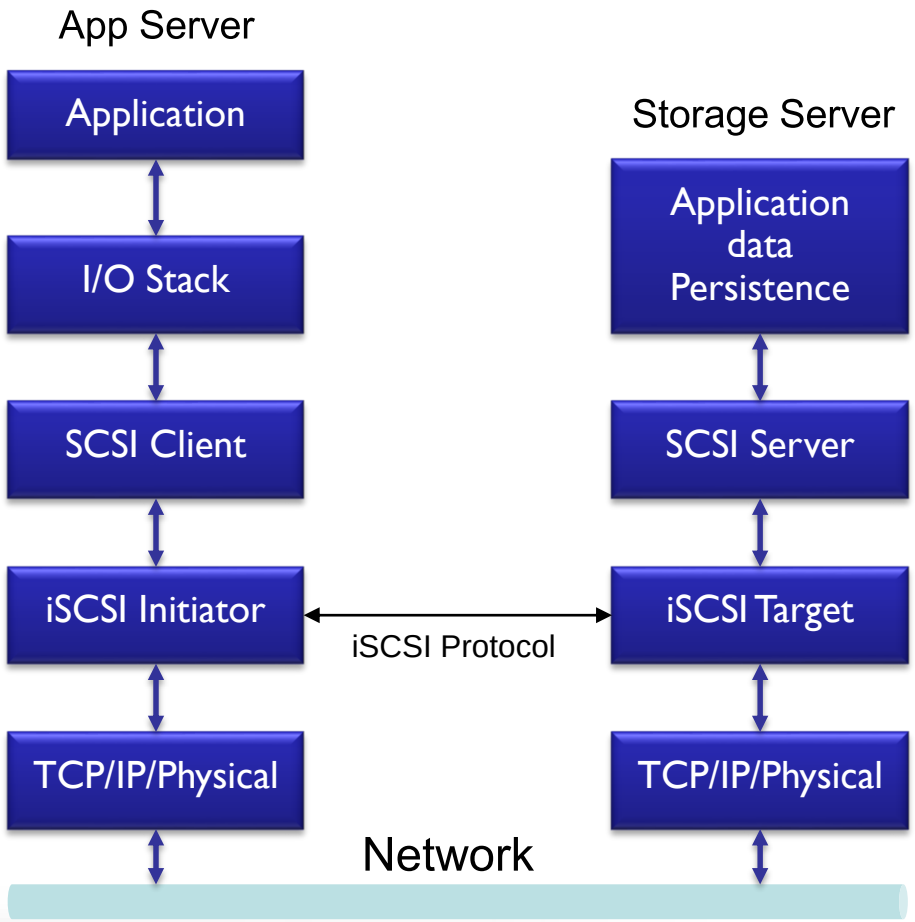
Any type of SCSI device can be accessed over iSCSI

Block Storage is the most typical (and the only supported on Windows Server)

Original protocol spec is RFC 3720

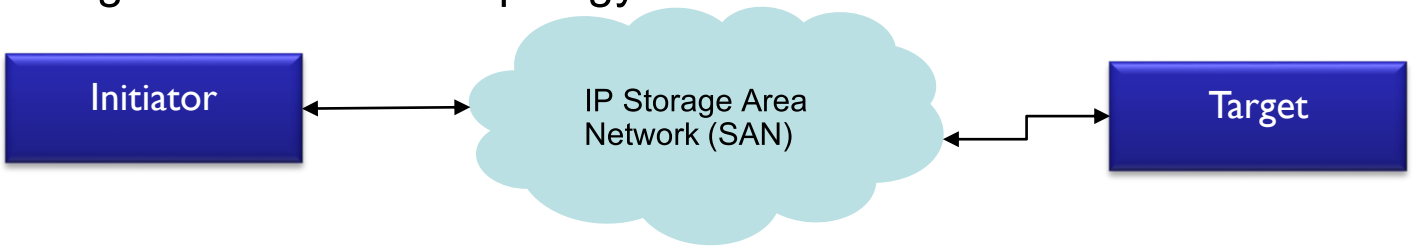
RFC 5048 corrects/clarifies the original

* FCP: Fibre Channel Protocol; IB: InfiniBand

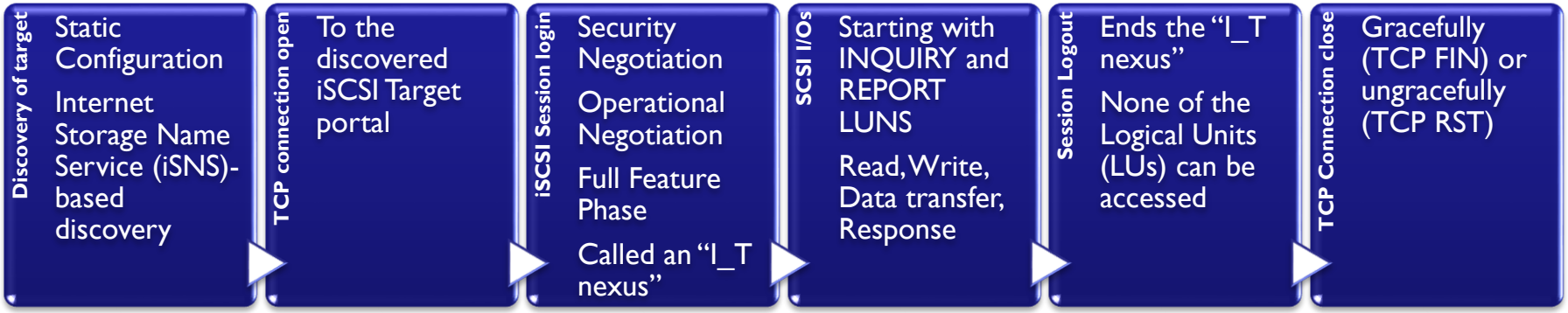


iSCSI terms: Initiator, SAN, Target, Session

Storage Area Network topology

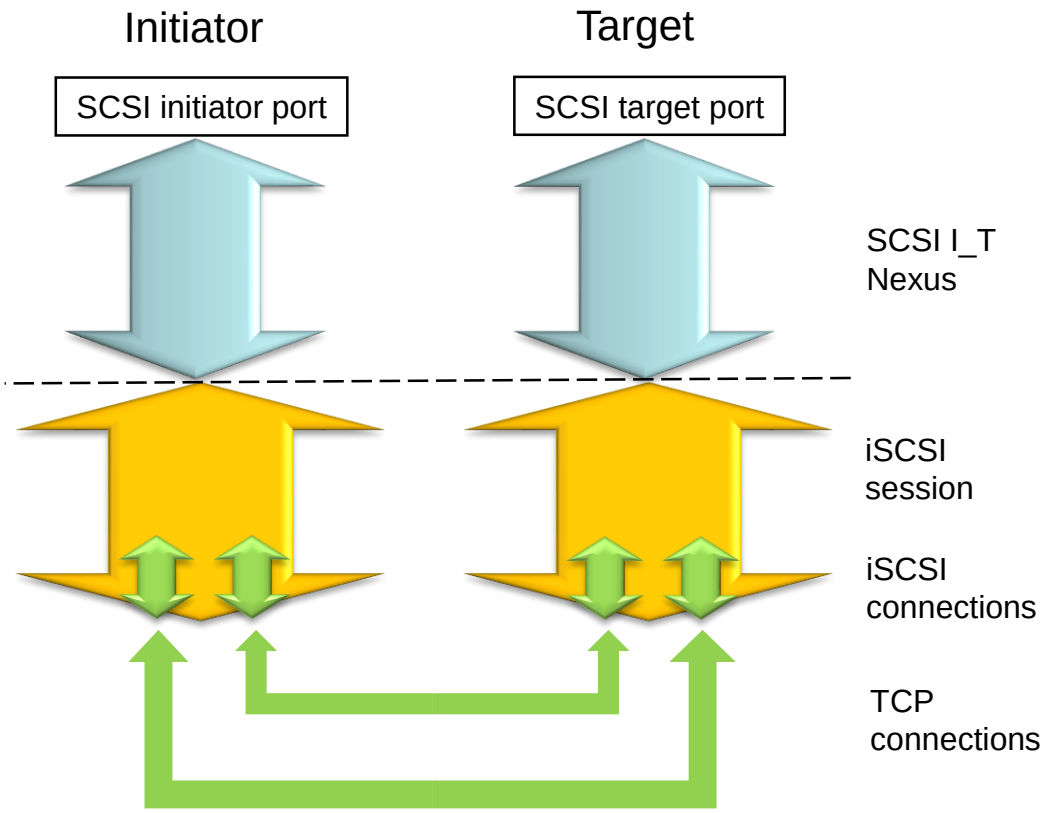


How communication occurs between a target and an initiator

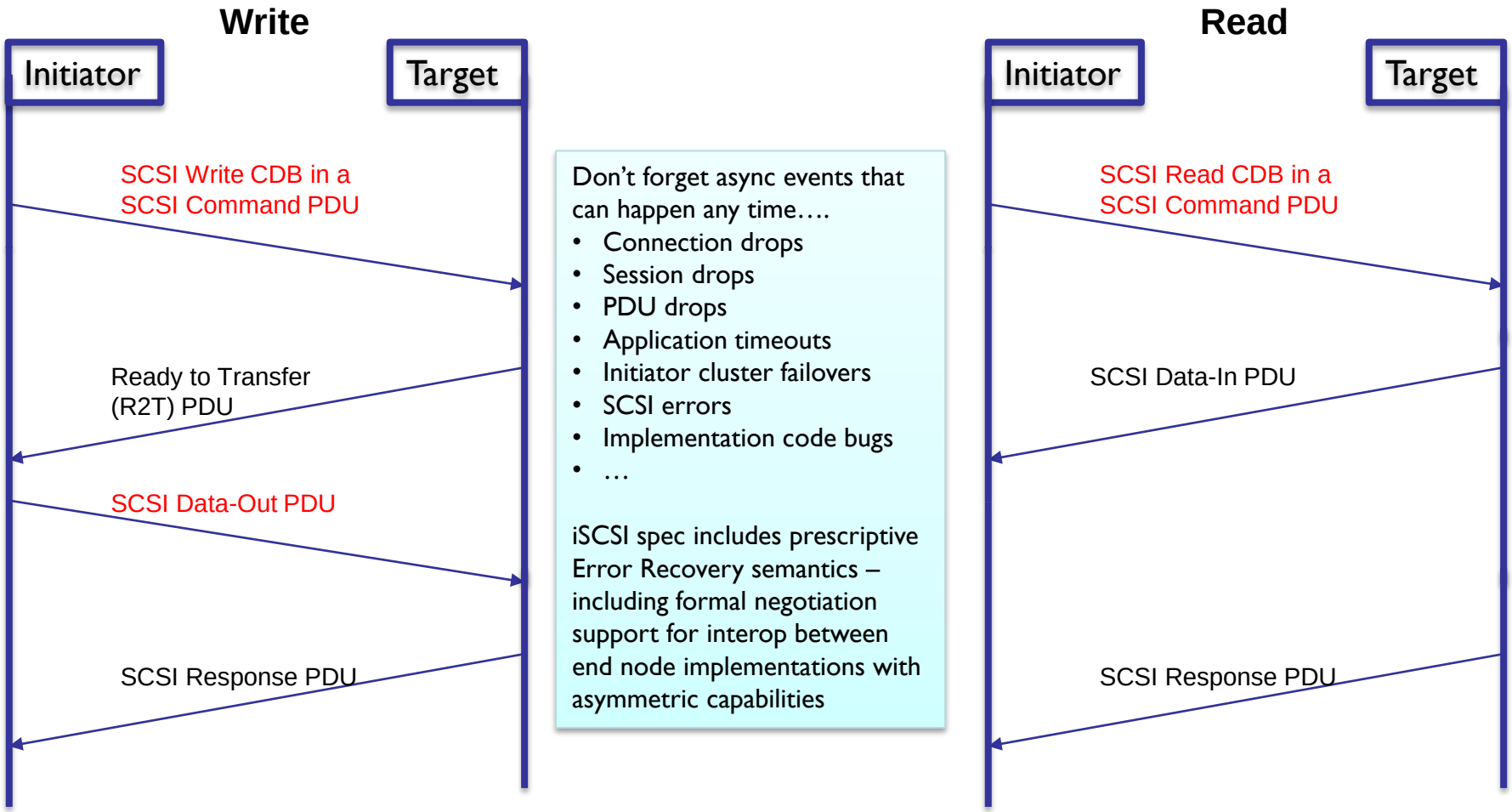


I_T nexus & multi-connection sessions

- iSCSI has native protocol support for combining multiple reliable transport connections into a single iSCSI session.
  - "Connection allegiance" for each I/O
  - Scaling throughput with multiple NICs
  - Load balancing and connection failure resiliency for I/Os in progress
- iSCSI is a "SCSI transport protocol"
  - iSCSI in turn relies on a *different network transport protocol* agnostic to SCSI semantics



SCSI I/O Mapping onto iSCSI transport

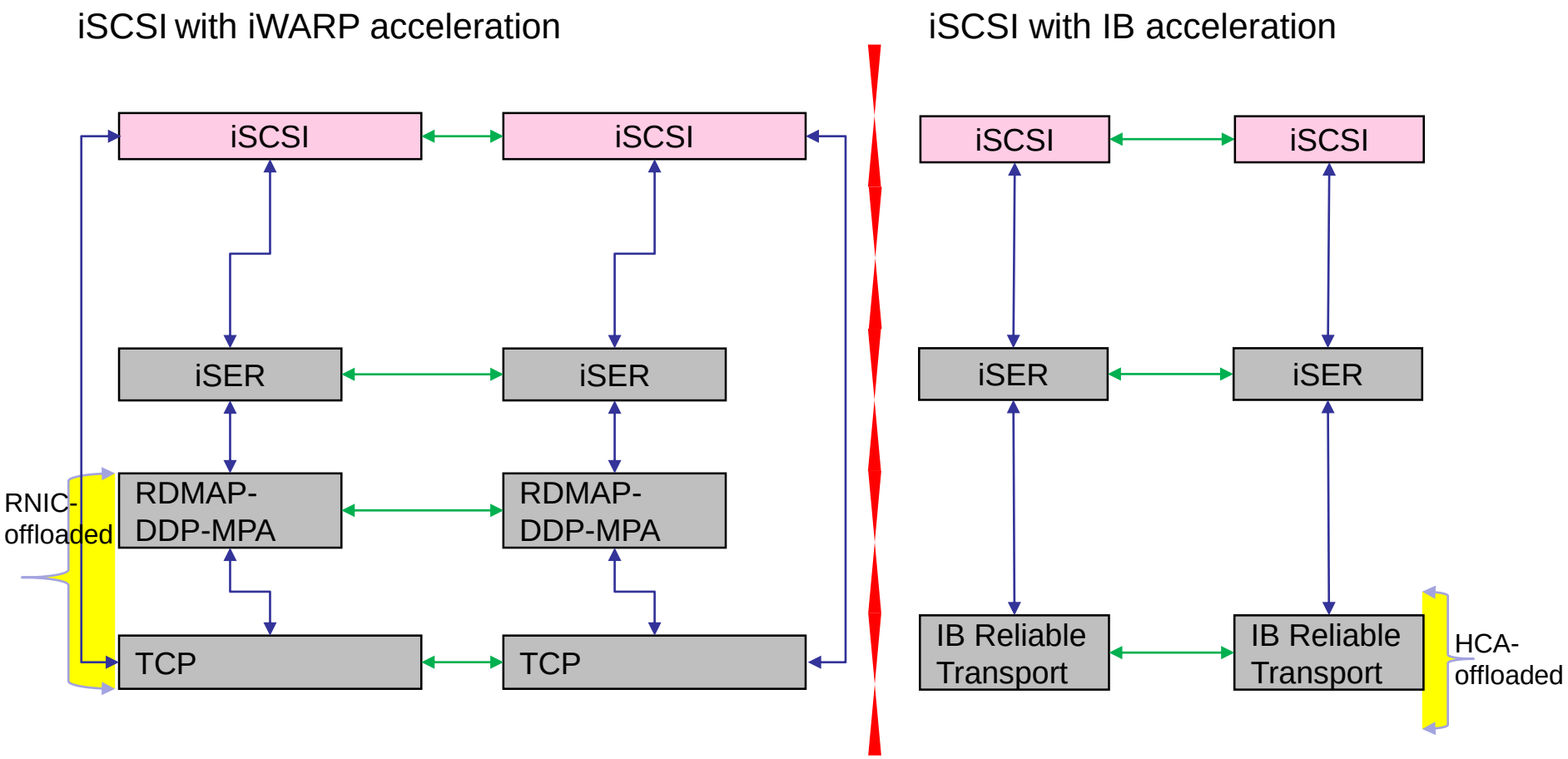


Don't forget async events that can happen any time….

- Connection drops
- Session drops
- PDU drops
- Application timeouts
- Initiator cluster failovers
- SCSI errors
- Implementation code bugs
- …

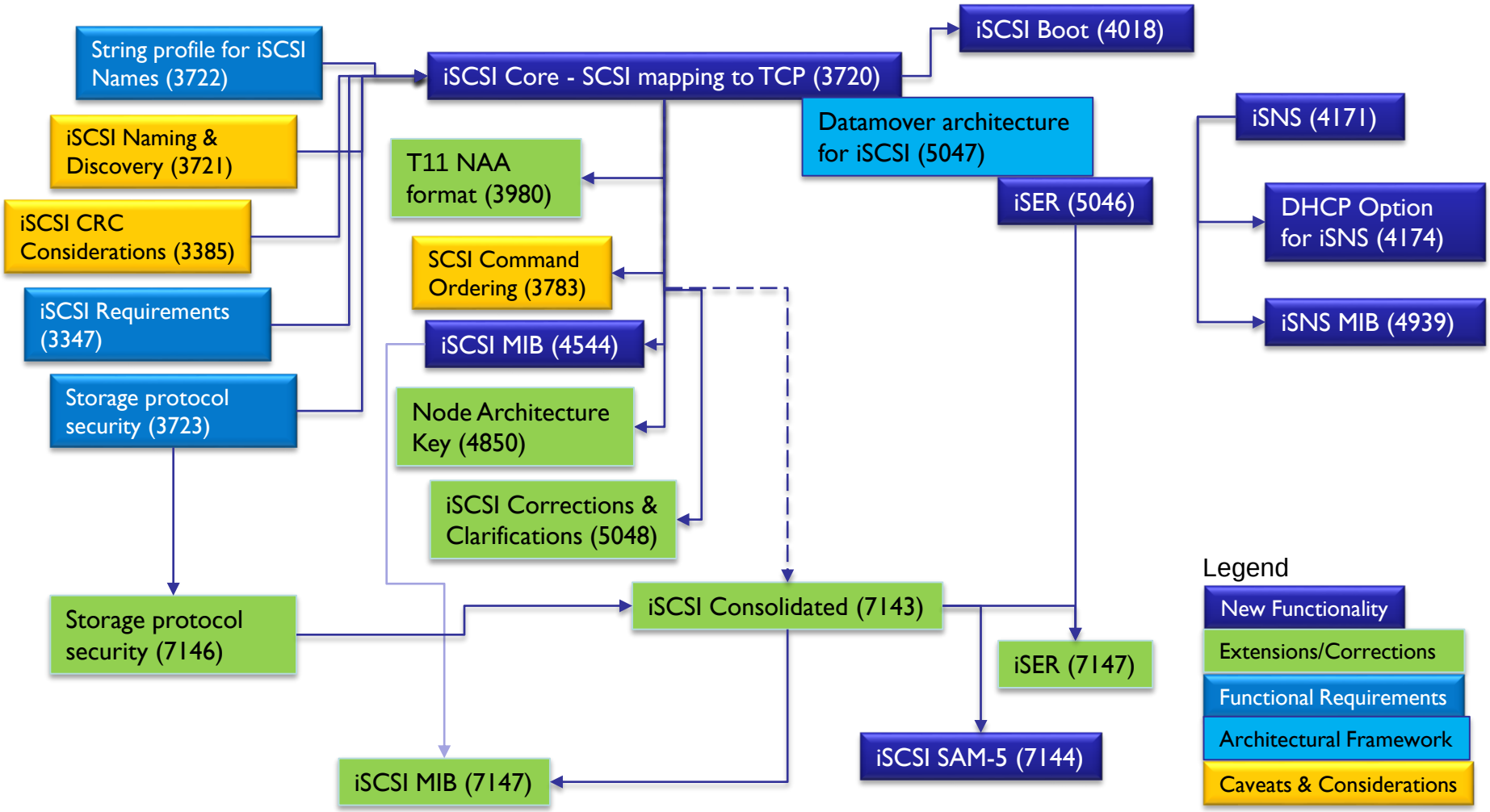
iSCSI spec includes prescriptive Error Recovery semantics – including formal negotiation support for interop between end node implementations with asymmetric capabilities

RDMA with iSCSI



iSCSI Standards – Navigational aid

Making Sense of iSCSI Spec Landscape



RFC 7144 – What did we add

SCSI PDU Updates

- Command Priority
  - An IN argument to the SAM-5 Execute Command () procedure call model
  - Indicates the relative scheduling importance of this task in comparison to other SIMPLE tasks
  - SCSI Command PDU addition (4-bits)
- Status Qualifier
  - An OUT argument to the SAM-5 Execute Command () procedure call model
  - Status qualifier provides additional information about the reason for the status code
  - SCSI Response PDU addition (2 bytes)

Sense Data

- Allowance for sense data
  - Typically, Sense Data is in DataSegment if the status is CHECK CONDITION
  - New draft explicitly allows Sense Data to be present anytime, independent of status

TMF Additions

- Following new Function Codes are now allowed in an iSCSI TMF Request PDU
  - `QUERY TASK (9)`: is the Referenced Task Tag present in the task set?
  - `QUERY TASK SET (10)`: is there a task from "my" I_T_L nexus in the task set?
  - `I_T NEXUS RESET (11)`: perform an I_T nexus loss function for all LUs accessible via "my" I_T nexus
  - `QUERY ASYNCHRONOUS EVENT(12)`: is there a unit attention condition or a deferred error pending for "my" I_T_L nexus?
- New TMF Response "Function succeeded" (equivalent to the FUNCTION SUCCEEDED SAM-5 service response)

iSCSIProtocolLevel

- New session-scoped (LO) text key
- iSCSIProtocolLevel negotiation decides the iSCSI protocol features that may be used on the session
- Plan is that each new standards-track RFC with protocol features will “claim” a new value
- Higher negotiated value → implicit support for lower numbered values
- Current legal values
  - 0: no version claimed
  - 1: iSCSI Consolidated RFC compliance
  - 2: iSCSI SAM-5 RFC compliance
- Key negotiation also causes the right version descriptor values for standard inquiry data to be reported (starting at byte 58)

RFC 7143 – What did we tweak

Key Change List…

1. Consolidates RFCs 3720, 3980, 4850 and 5048, and made the necessary editorial changes
2. Claims a value for the new iSCSIProtocolLevel
3. Removes Markers and related keys
4. Removes SPKM authentication and related keys
5. Explicitly allows initiator+target implementations, including the composite device naming
6. Clarifies that SLP-based discovery cannot be relied on for interoperability
7. Specifies formal protocol artifact relationships via UML diagrams

Key Change List… (contd)

8. Makes FastAbort implementation a "SHOULD" from the previous "MUST"
9. Requires implementing IPsec, 2400-series RFCs (IPsec v2, IKEv1); and SHOULD implement IPsec, 4300-series RFCs (IPsec v3, IKEv2).
10. Restricts the usage of X#, Y# and Z# name prefixes
11. Provides guidance on minimal number of text negotiation responses
12. Provides guidance on Kerberos authentication, OCSP usage, and extended sequence numbers (ESNs)

Now let’s take a whirlwind tour of a few changes…

iSCSI Spec Consolidation

- Following specs are consolidated into the new draft:
  - RFC 3720: base iSCSI protocol
  - RFC 3721: iSCSI naming and discovery considerations (selective updates only)
  - RFC 3980: T11 NAA naming format addition
  - RFC 4850: Public extension key addition
  - RFC 5048: Clarifications and corrections

Markers

- RFC 3720 defined a “sync-and-steering layer” to help in direct data placement of inbound iSCSI data, in the presence of dropped packets
  - Targets reassembly memory, copying overhead etc.
  - “Markers” are recurring pointers embedded in the data stream to help with direct data placement
- Subsequent protocol advances, notably iWARP on TCP/IP, have made RDMA more broadly available beyond iSCSI; and iSER adapts iSCSI to run on iWARP.
- Practically, there have been almost no implementations
- So…. Markers concept, along with related text keys, is now removed

SPKM Authentication

- Simple Public-Key Mechanism (SPKM) support for iSCSI authentication is now removed in RFC 7143
  - Means both SPKM-1 and SPKM-2
- No iSCSI implementation adoption

SLP-based Discovery

- Original iSCSI discovery approach was three-pronged
  - SendTargets-based (in-band/Discovery sessions)
  - Service Location Protocol (SLP)-based
  - iSNS-based
- SLP-based discovery did not pick up wide adoption; Naming & Discovery RFC wording however implies a stronger SLP requirement (“SHOULD”)
- Recommended approach now: try SendTargets, and then try iSNS; don’t rely on SLP-based discovery

FastAbort Requirement

- Multi-task aborts: LU Reset, Clear Task Set, Target Reset etc.
- RFC 3720 semantics: some multi-initiator scenarios where multi-task aborts could cause target deadlocks, waiting on initiators on third-party sessions that may never respond
- “Clarified” semantics in RFC 5048 address deadlock issues, but still not optimal (may cause initiator timeouts & error recovery escalations)
- FastAbort semantics: targets can provide accelerated responses; they can deal with book-keeping/quiescing operations in a lazy fashion (which are the real culprits that trigger timeouts)
- RFC 7143 makes this functionality now a “SHOULD” (it’s a “Good Thing” to implement, but not an absolute requirement) – so implementations may get away with “Clarified” semantics, although not recommended

IPsec v2 vs. IPsec v3

- iSCSI nodes must implement IPSec for data authentication and integrity (run-time usage is up to SAN administrator)
- MUST provide data authentication and integrity by implementing IPsec v2 [RFC2401] with ESPv2 [RFC2406] in tunnel mode
- SHOULD implement data authentication and integrity by implementing IPsec v3 [RFC4301] with ESPv3 [RFC4303] in tunnel mode
- Either IPsec v2 or v3
  - Authentication & integrity with ESP in transport mode is optional
  - Crypto algorithm changes, e.g., AES CBC replaces 3DES CBC as the “MUST implement” encryption algorithm.
- Finally, 1Gbps and higher implementations “MUST implement and SHOULD use” extended sequence numbers in ESP to avoid frequent rekeying

Kerberos authentication & OCSP

- iSCSI uses “raw” Kerberos authentication (only), without GSS-API
  - iSCSI implementations with Kerberos support then must be aware of Kerberos payloads
  - New guidance expands on the specifics – how KRB_AP_REQ and KRB_API_REP are to be handled
  - New guidance also strongly recommends mutual authentication whenever Kerberos is used (caveat: mutual authentication still *does not* guarantee it’s the desired service principal – use iSNS for service discovery!)

- Online Certificate Status Protocol (OCSP) usage
  - New text calls out that OCSP may be used in addition to CRL (Certificate Revocation List) for certificate-based IKE authentication

How many “Texts” are too many?

- Text negotiation is used during Login or during Full Feature Phase
- RFC 3720 allows an implementation to drop a connection if a negotiation does not converge after “a reasonable number” of text exchanges
- RFC 7143 makes it crisper – six text exchanges SHOULD be supported

Wrap-up – So you have questions, ☺

For any follow-up questions

- Decide if it is an implementation question or a protocol question
  - If it's the former, your iSCSI vendor will answer it for you
- If it's a protocol question, start with the IETF storm WG DL
  - storm@ietf.org (you must be a DL member)
- If you do *really* need to directly reach the RFC authors (also the presenters today)
  - Get a hold of us right after this session!
  - If not, check out RFC 7143 and RFC 7144 (author email addresses at the end of the RFCs)
  - Whichever RFC you have a question on, make sure to include all authors for that RFC on your email

Thank You.