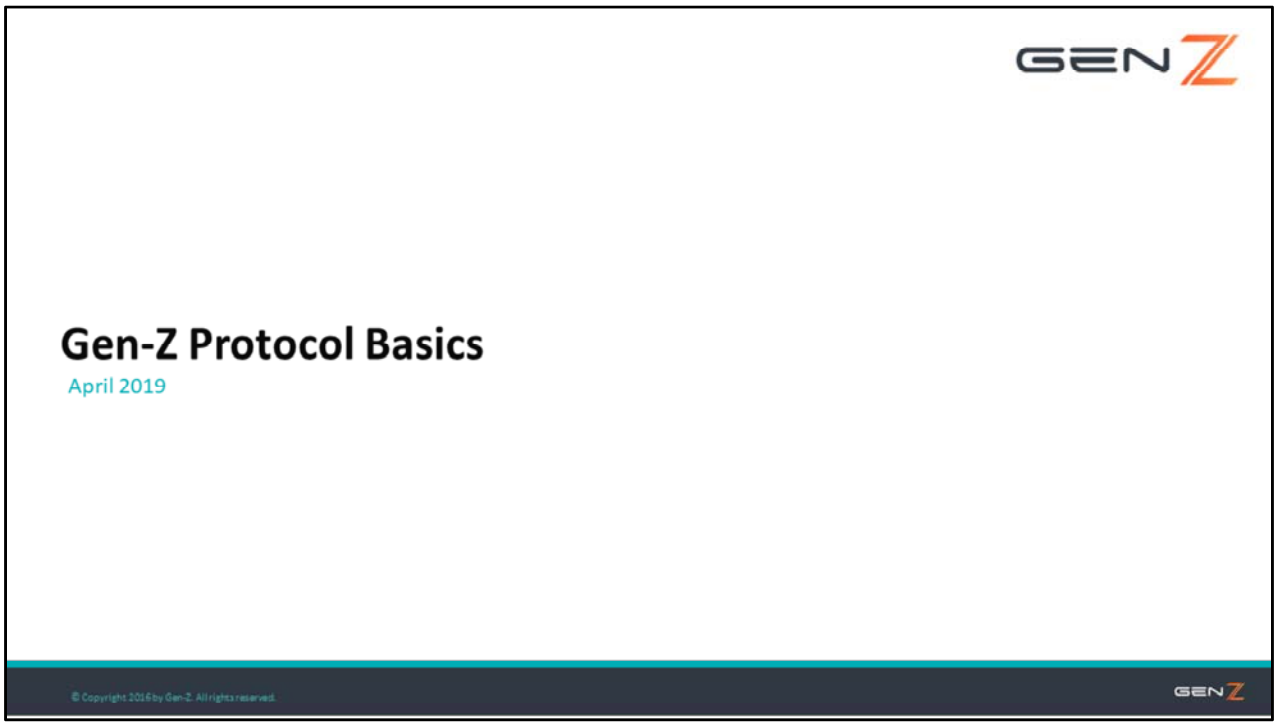


This presentation covers the basic elements of the Gen-Z protocol.

Disclaimer

This document is provided 'as is' with no warranties whatsoever, including any warranty of merchantability, noninfringement, fitness for any particular purpose, or any warranty otherwise arising out of any proposal, specification, or sample. Gen-Z Consortium disclaims all liability for infringement of proprietary rights, relating to use of information in this document. No license, express or implied, by estoppel or otherwise, to any intellectual property rights is granted herein.

Gen-Z is a trademark or registered trademark of the Gen-Z Consortium.

All other product names are trademarks, registered trademarks, or servicemarks of their respective owners.

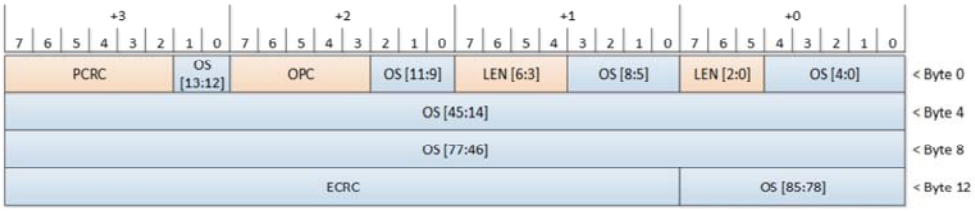
All material is subject to change at any time at the discretion of the Gen-Z Consortium

http://genzconsortium.org/

Gen-Z Protocol Overview

- Gen-Z uses the following basic packet layouts
  - Link-local—small packets used to exchange link-specific operations, e.g., flow control packets
  - P2P 64—end-to-end packets optimized for point-to-point and mesh topologies.
  - P2P-Vendor-defined—end-to-end packets optimized for point-to-point and mesh topologies using vendor-define operations
  - Explicit—end-to-end packets optimized for single or multi-subnet switch topologies, multi-tenancy, and to access up to a 64-bit resource address space per component
    - Explicit OpClass packets may be used in any topology including point-to-point, mesh, and switch-based topologies
    - Gen-Z supports nearly any routing protocol enabling explicit OpClass packets to used at any scale from single enclosure to multi-enclosure / rack-scale solutions

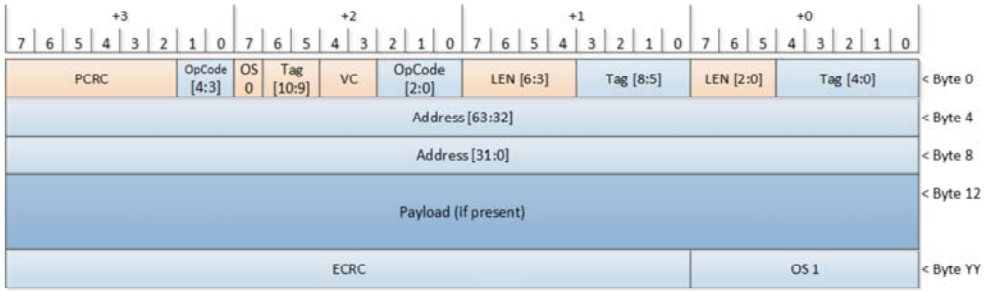
Link-Local Generic Packet Layout



- Length == 0x7F indicates if a link-local or end-to-end packet
- PCRC—protects length and OPC fields
- ECRC—protects entire packet sans ECRC field
- OPC—OpCode
- OS—OpCode-specific fields

Link-local packets are fixed-sized. With the exception of Link Idle, all link-local packets are 128b in length. The packet is protected by two CRCs– the PCRC protects the length and OpCode, and the ECRC protects the entire packet sans the ECRC field.

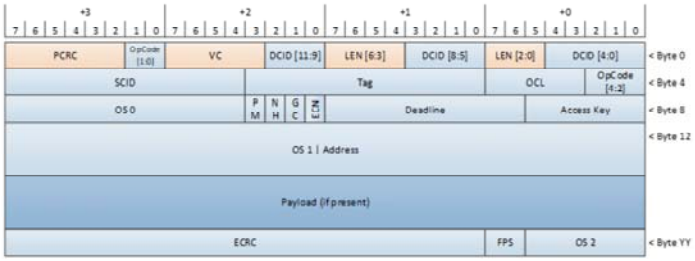
P2P 64 Generic Packet Layout



- PCRC—protects length and VC
- ECRC—protects entire packet
- Tag—an opaque identifier used to correlate request and response packets
- LEN—packet length in 4-byte multiples
- Address—If present, the Address of the target resource
- Payload—If present, the data payload associated with the target resource
- VC—Virtual Channel
- OS—OpCode-specific

GEN Z

Explicit Generic Packet Layout



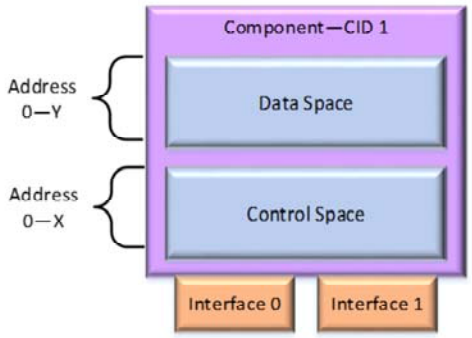
- PCRC—protects length and VC
- ECRC—protects entire packet sans ECRC field
- Tag—an opaque identifier used to correlate request and response packets
- LEN—packet length in 4-byte multiples
- Address—If present, the Address of the target resource
- Payload—If present, the data payload associated with the target resource
- DCID—subnet-local destination component identifier
- OpCode—identifies one of 32 operation types
- SCID-subnet-local source component identifier
- VC—Virtual Channel
- OCL—OpClass Label—identifies 1 of 32 Explicit OpClasses
- Access Key—Access Key used to enforce component-level access control
- Deadline—"Wall Clock" Time-to-live
- ECN—Explicit Congestion Notification Bit
- GC—Indicates if multi-subnet header is present
- NH—Indicates if next header is present (e.g., HMAC authentication or precision timestamp)
- PM—Performance Marker indication
- FPS—Forward progress field

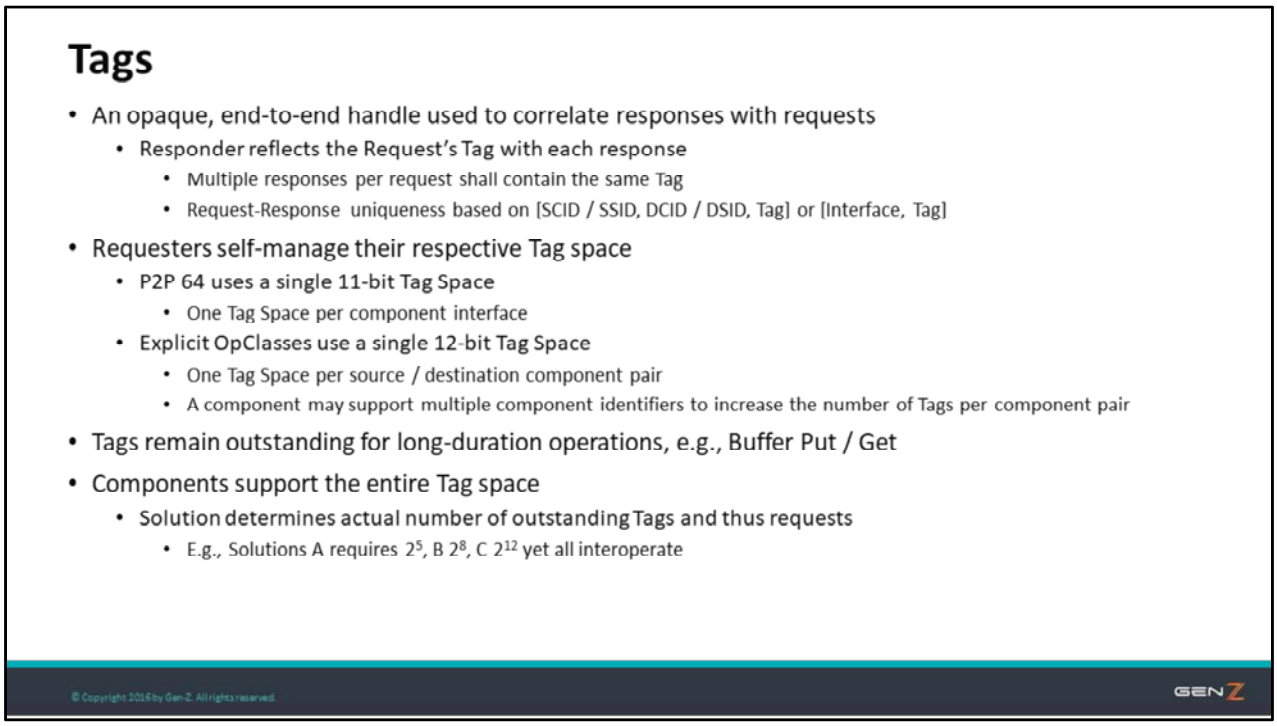
GEN Z

Explicit OpClass packets contain additional fields. These include component identifiers, OpClass labels to identify the specific explicit OpClass, Access Keys to validate whether a packet can communicate across a given interface, an ECN bit to communicate congestion, a packet deadline field to indicate when a packet should be discarded if the value hits zero, a forward progress field (FPS) to indicate if a request packet was previously rejected due to a lack of resources required to execute the request, etc. The packet also includes at least 3 1-bit fields that indicate whether a multi-subnet header is present, whether the Next Header field is present, and whether an application wants performance records to be generated as the packet traverses the topology.

Addressing

- Request Address
  - Address size varies by OpClass
    - P2P 64—up to 2^{64}
    - Explicit OpClasses
      - Non-Control OpClass—up to 2^{64} bytes per component
      - Control OpClass—up to 2^{52} bytes per component
- Two address spaces
  - Data address space—application data
    - No address range "carve outs"
    - Single-byte addressability
  - Control address space—control structures, accelerator executables, protected data, etc.
    - Control space accessed by trusted components / services
    - Single-byte addressability





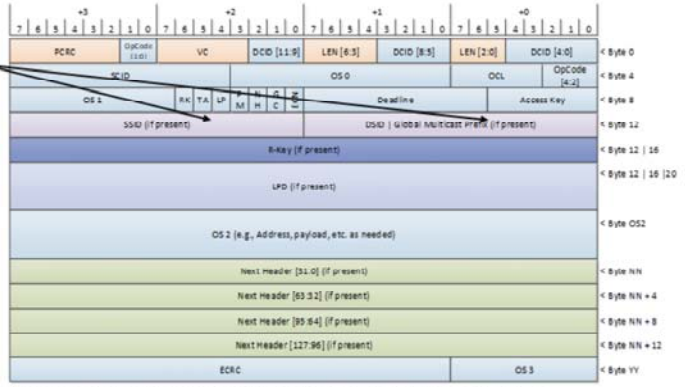
All Tags are opaque, un-encoded handles used to correlate request packets with response packets. Tags are managed solely by the Requester.

P2P 64 uses a single un-encoded 11-bit Tag space per component interface, i.e., up to 2048 outstanding request packets per component interface.

Explicit OpClass packets uses a single un-encoded 12-bit Tag space per source / destination component pair. This enables up to 4096 outstanding request packets per communicating peer component.

GC Field

If GC == 1b, then these fields are present

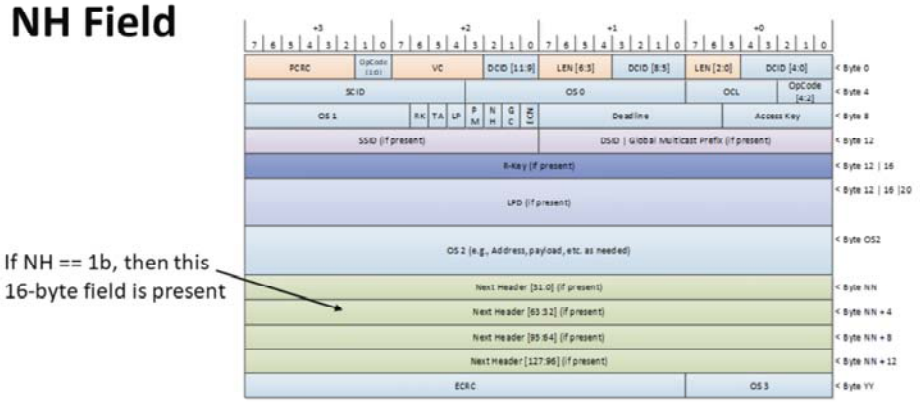


- GC field indicates if the packet can explicitly traverse multiple subnets
  - GC == 0b indicates this is a single subnet packet (or one that can traverse Transparent Routers)
  - GC == 1b indicates this is a multi-subnet packet
    - An additional 32 bits are included. If unicast, then these contain the source and destination subnet identifiers
    - If a multicast packet, then bytes 0 and 1 contain the global multicast prefix instead of the destination subnet identifier

GEN Z

The GC field is present in all Explicit OpClass packets. If GC == 1b, then this indicates an optional 4-byte field is present as illustrated. If a unicast packet, then this 4-byte field contains the source and destination component subnet identifiers (SID). If a multicast packet, then the SSID contains the subnet identifier of the source component, and, instead of a DSID, the field is interpreted as the Global Multicast Prefix.

NH Field



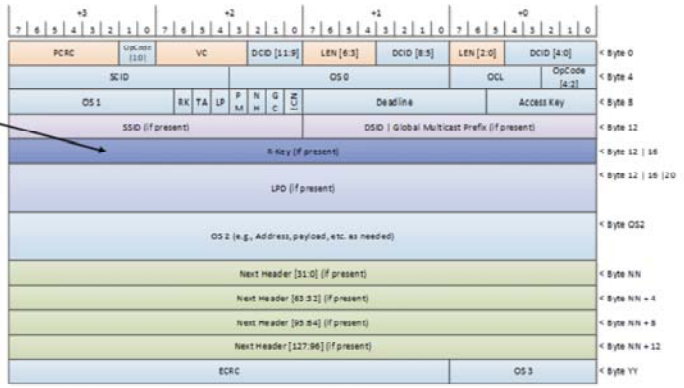
- NH indicates if the Next Header field is present
- If NH == 1b, then the Next Header field is present and located as shown.
- Next Header is used to carry HMAC digest, anti-replay attack value, precision timestamp, etc.

GEN Z

The NH field is present in all explicit OpClass packets. If NH == 1b, then this indicates an additional 16-byte Next Header field is present as illustrated. The Next Header field is used to transport packet authentication fields (HMAC and ART) or a precision timestamp.

RK and R-Key Field

If RK == 1b, then this 32-bit field is present



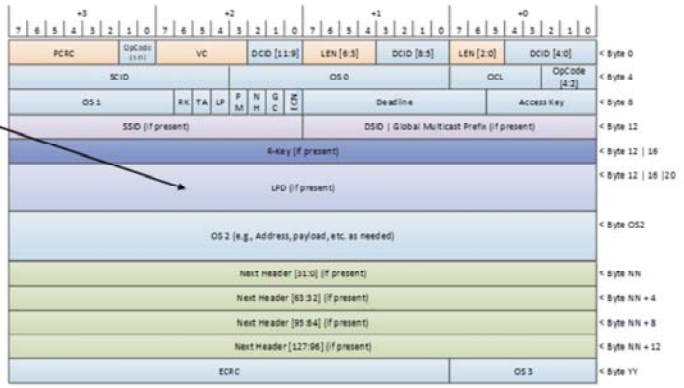
- Some explicit OpClass request packets contain the RK field
- RK field indicates if the R-Key field is present
- If RK == 1b, then the R-Key field is present and located as shown

GEN Z

The RK field is present in some explicit OpClass request packets. If RK == 1b, then it indicates an additional 32-bit R-Key field is present as illustrated. The R-Key field is used to validate access permission to the addressed resource.

LPD Field

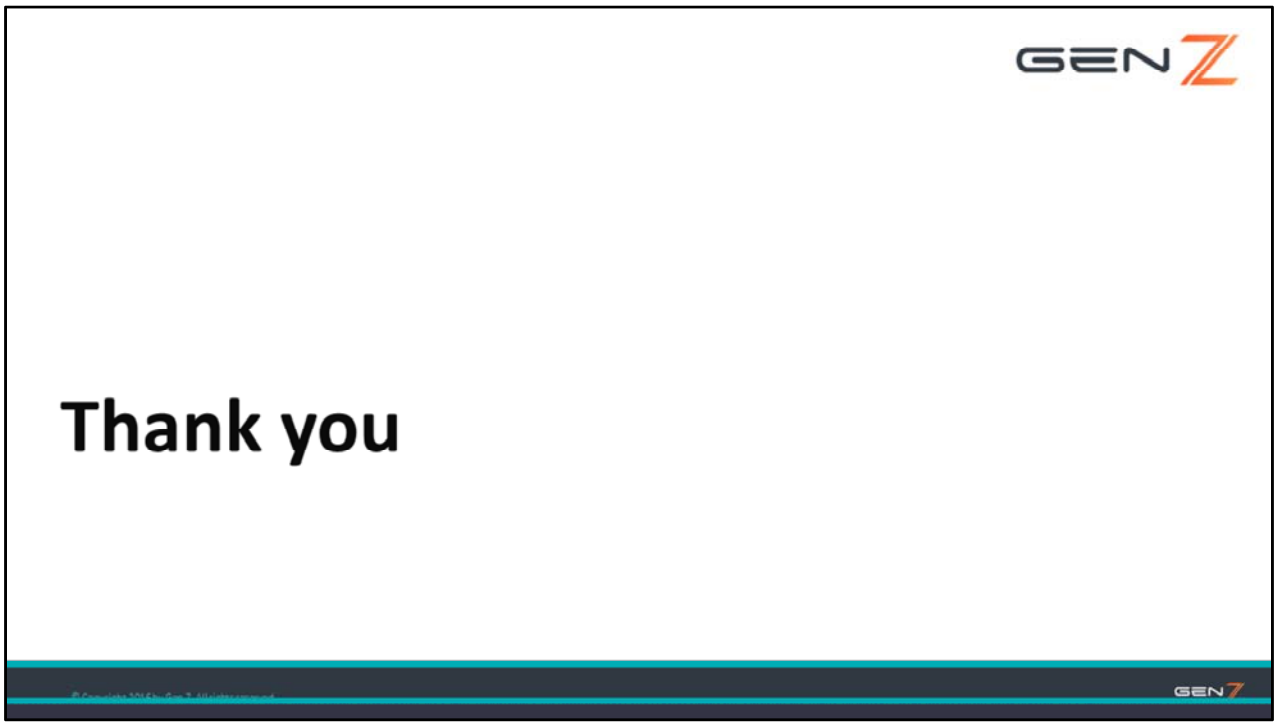
If LP == 1b, then the LPD field is present



- Some explicit OpClass request packets contain the LP field
- LP field indicates if the Logical PCI Device (LPD) field is present
- If LP == 1b, then the LPD is present and located as shown
- LPD field contains the PCI Bus Number, Device Number, Function Number, PASID, etc. fields based on the supported PCIe capabilities

GEN Z

The LP field is present in some explicit OpClass packets. If LP == 1b, then this indicates the LPD (Logical PCI Device) field is present. The LPD field contains sub-fields used to communicate PCI / PCIe-specific information.



This concludes this presentation. Thank you.