Management Component Transport Protocol (MCTP) Overview White Paper

Version 1.0.0a

Status: Informational

Publication Date: July, 2007

DSP2016

Version 1.0.0a

Publication Date: June, 2007

DSP2016

Status: Informational

Abstract

The Management Component Transport Protocol (MCTP) is a protocol defined by the DMTF Platform Management Component Intercommunications sub-team of the DMTF Pre-OS Workgroup. MCTP is designed to support communications between different intelligent hardware components that make up a platform management subsystem that is provides monitoring and control functions inside a managed system.

This document provides an overview of the architectural goals and general operation of MCTP.

Table of Contents

List of Figures

1 Introduction

The lack of standard interfaces among platform management components impedes the ability for platform developers to be able quickly and economically develop and deploy platform management subsystems that can be accessed using DMTF Common Information Model (CIM) Profiles and access protocols. The DMTF Platform Management Component Intercommunications (PMCI) sub-team of the DMTF Pre-OS Workgroup seeks to address this problem by defining interfaces and data models that facilitate interoperability and interchangeability among platform management components, enabling devices from different vendors to be more readily linked together into a cohesive manageability subsystem.

The Management Component Transport Protocol (MCTP) supports the PMCI goals by defining a media-independent transport protocol that enables communications between different intelligent hardware components that make up a platform management subsystem that provides monitoring and control functions inside a managed system.

1.1 Target Audience

The intended target audience for this document is readers that develop or utilize platform management subsystems that are formed using management controllers and intelligent management devices and are interested in obtaining an overview of the MCTP specifications that define a common intercommunication mechanism for those components.

1.2 Related Documents

[1] DMTF, DSP0236, *Management Component Transport Protocol (MCTP) Base Specification*

[2] DMTF, DSP0237, *Management Component Transport Protocol (MCTP) SMBus / I^2C Transporting Binding Specification*

[3] DMTF, DSP0238, *Management Component Transport Protocol (MCTP) PCIe VDM Transport Binding Specification*

[4] DMTF, DSP0239, *Management Component Transport Protocol (MCTP) IDs and Codes*

[5] DMTF, DSP2014, *Systems Management Architecture for Mobile and Desktop Hardware White Paper*

1.3 Terminology

Term	Definition
ATCA	Advanced Telecommunications Computing Architecture. ATCA is a specification effort PCI Industrial Computer Manufacturers Group (PICMG). Known as AdvancedTCA™ (also often abbreviated as ATCA), the official specification designation is PICMG 3.x. AdvancedTCA requirements of "carrier grade" communications computer systems. ATCA includes specifications for platform management subsystem elements within and between blades in a blade-based computer chassis.
I^2C	Name of a multi-master, two-wire, serial bus originally developed by Philips Semiconductor.
In-Band	Management that operates with the support of hardware components that are critical to and used by the operating system

Term	Definition
Intelligent Management Device	A management device that is typically implemented using a microcontroller and accessed through a messaging protocol. Management parameter access provided by an IMD is typically accomplished using an abstracted interface and data model rather than through direct "register level" accesses.
IPMB	Intelligent Platform Management Bus. Name for the architecture, protocol, and implementation of an I2C bus that provides a communications path between "management controllers" in IPMI -based systems.
IPMI	A set of specifications defining interfaces and protocols originally developed for server platform management by the IPMI Promoters Group: Intel, Dell, HP, and NEC
Manageability Access Point (MAP)	A collection of services of a system that provides management in accordance to specifications published under the DMTF Server Management Architecture for Server Hardware initiative.
Management Controller	A microcontroller or processor that aggregates management parameters from one or more management devices and makes access to those parameters available to local or remote software, or to other management controllers, through one or more management data models. Management controllers may also interpret and process management-related data, and initiate management-related actions on management devices. While a native data model is defined for PMCI, it is designed to be capable of supporting other data models, such as CIM, IPMI, and vendor-specific data models. The microcontroller or processor that serves as a management controller can also incorporate the functions of a management device.
Managed Element	The finest granularity of addressing which can be the target of commands or messages, or a collection thereof.
Out-of-Band	Management that operates with hardware resources and components that are independent of the operating systems control
RMII	A reduced signal count MAC to PHY interface, based on the IEEE Media Independent Interface (MII), which was specified by the RMII Consortium (3Com Corporation; AMD Inc.; Bay Networks, Inc.; Broadcom Corp.; National Semiconductor Corp.; and Texas Instruments Inc.)
SMBus	Name of a multi-master, two-wire, serial bus specified by the Smart Battery Systems Implementer's Forum.

1.4 Acronyms and Abbreviations

Term	Definition
MAP	Manageability Access Point
MCTP	Management Component Transport Protocol
PCIe	PCI Express™
PMCI	Platform Management Component Intercommunications. The name of the sub-team of the DMTF Pre-OS Working Group that developed the MCTP and other platform management hardware -related specifications.
USB	Universal Serial Bus

2 MCTP Highlights

The Management Component Transport Protocol (MCTP) is a media-independent protocol for intercommunication among intelligent devices within the platform management subsystem of a managed computer system. This protocol is independent of the underlying physical bus properties, as well as the "data-link" layer messaging used on the bus.

The physical and data-link layer methods for MCTP communication across a given medium are defined by companion "transport binding" specifications, such as MCTP over PCIe® Vendor Defined Messaging and MCTP over SMBus/ I^2C. This approach enables future transport bindings to be defined to support additional buses such as USB, RMII, and others, without affecting the base MCTP specification.

MCTP has been designed to carry multiple types of manageability-related traffic across a common medium. The base MCTP specifications define message types for supporting the initialization and configuration of MCTP itself, and to support vendor-specific messages over MCTP. Other message types, such as message types to support a Platform Level Data Model (PLDM), network controller sideband communications, and so on, are planned to be specified in the future by the DMTF PMCI workgroup.

An example of a basic platform management subsystem topology that utilizes MCTP is shown in Figure 1 on page 13.

2.1 Goals

MCTP has been designed to address the following goals:

- Provide efficient communications between the following parties in a platform management subsystem:
    - Management Controllers and Intelligent Management Devices
    - Management Controllers and other Management Controllers
    - Management Controllers and system firmware (e.g. BIOS)
    - Management Controllers and Network Controllers
- Support multiple media types: for example, SMBus/I^2C, PCIe VDM, USB, and others.
- Support multiple message types on a common media.
- Be suitable for different classes of computer system such as server, desktop, mobile, and communications systems.
- Use compact, byte-efficient formats and simple protocols that are suitable for implementation using low-cost microcontrollers.
- Capture and integrate learning's from other management bus protocols such as IPMB, SMBus, ATCA, ASF, by providing:
    - Improved addressing
    - Simplified message routing
    - Support for device discovery and hot plug devices

- Be cleanly extensible and maintainable
  - Provide clear points in the specifications and protocol for supporting new messages types, media types, protocol version information, and so on.
  - Provide support for Vendor-specific message types to allow value-added functions to use MCTP as a framework without conflicting with base MCTP messaging.

2.2 MCTP Specification Documents

MCTP is defined through a set of DMTF specifications. The following documents make up the MCTP specifications:

2.2.1 Common Specifications

These documents are the generic documents for the MCTP.

DSP0236, *Management Component Transport Protocol (MCTP) Base Specification*

This document describes the aspects of the MCTP, message routing, and MCTP initialization that are common across all MCTP implementations. The specification consists of three main sections: MCTP base protocol, MCTP Control Protocol, and MCTP Control Commands.

DSP0239, *Management Component Transport Protocol (MCTP) IDs and Codes*

This document presents a collection of IDs and codes that are used across the Management Component Transport Protocol (MCTP) and transport binding specifications.

2.2.2 Transport Binding Specifications

These documents describe how MCTP is implemented on a particular physical medium. There are two transport binding specifications as of this writing:

DSP0237, *Management Component Transport Protocol (MCTP) SMBus / I^2C Transport Binding Specification*

This document defines how MCTP is carried using SMBus or I^2C as the transport medium.

DSP0238, *Management Component Transport Protocol (MCTP) PCIe VDM Transport Binding Specification*

This document defines how MCTP is carried using PCI Express™ Vendor Defined Messages (VDMs) as the transport medium.

2.2.3 Message Type Specifications

MCTP Message Type specifications define how a particular class of payload is carried using MCTP. As of this writing, only MCTP Control and Vendor Defined message types are specified. The specification of these message types is incorporated in the MCTP Base Specification rather than as separate documents. In the future, message types for carrying traffic such as the PMCI Platform Level Data Model will be defined and provided as companion specifications to the present MCTP specification documents.

3 MCTP Technology Overview

This section provides a more detailed overview of the elements and operation of MCTP.

3.1 MCTP Base Protocol

The basic unit of data transfer in MCTP is the “MCTP packet”. One or more MCTP packets are used to transfer an “MCTP message”. The base protocol defines the common fields for MCTP packets and how they are used. This includes defining fields such as source and destination address fields, fields that identify which packets belong to a particular MCTP message, and fields that define what type of communication traffic is being carried in the MCTP message. The base protocol also defines the processes used for assembling MCTP messages, routing MCTP packets, and handling error conditions such as dropped or missing packets.

3.2 MCTP Endpoints and EIDs

An endpoint is the function within a device that terminates the communication protocol of MCTP and handles MCTP Control commands. MCTP uses a logical address called the endpoint ID (EID) for addressing and routing MCTP packets to and from endpoints.

3.3 MCTP Busses

In MCTP a bus is defined as an interconnect between platform components that share a common physical layer address space. A bus may be made up of multiple segments. A bus segment is a portion of a bus that is electrically separated from other segments that form a bus, but still shares a common physical address space with other segments.

3.4 MCTP Bus Owners and Bridges

Each MCTP bus has a bus owner. Bus Owners are responsible for assigning EIDs to any MCTP devices on that bus. A bus owner may also have additional media-specific responsibilities, such as device discovery and assignment of physical addresses.

MCTP Bridges are devices that connect to two or more MCTP busses and are responsible for routing MCTP packets between those busses. A Bridge will typically also be the bus owner for at least one of the busses to which it connects.

MCTP allows multiple bridges, busses, and bus owners to be interconnected to form an “MCTP network”. Because bus owners are responsible for assigning EIDs to any devices that are on the bus that it owns, MCTP provides a mechanism that enables bus owners to be allocated a pool of endpoint IDs that can subsequently be assigned or allocated to other devices. The ultimate source of EIDs for the entire MCTP network comes from what is referred to as the “topmost bus owner”.

MCTP bus owners and bridges are also responsible for providing a service that enables resolving EIDs into physical addresses so that the originator of an MCTP packet knows what destination physical address needs to be used to route a message to a given EID.

3.5 MCTP Packet Routing

MCTP packets are routed based on their EIDs. MCTP bridges maintain a set of information referred to as the “routing table” that tracks the relationship between the physical addresses and

bus with which an EID is associated. When an incoming packet is received, this information is used to route the packet to the appropriate bus. If the source and target busses use different physical media the bridge is also responsible for translating the physical layer formatting of the packet as required by the target medium.

The information in the routing table is also used for handling requests to resolve an EID into a physical address and to support a capability to query individual bridges and bus owners for their routing information. This latter capability provides a way to obtain a snapshot of the MCTP network's routing configuration.

The MCTP bridging function simply forwards packets based on endpoint addressing information on a per packet basis. Otherwise, bridging does not interpret message content, or handle message type-specific protocol behavior for routed packets. Bridging does not do intermediate assembly or disassembly of routed packets. Message assembly and disassembly is handled solely by the destination and source endpoints, respectively.

MCTP packet routing between busses does not include a broadcast capability, although individual busses may support broadcast as required by the particular medium to support device discovery. This eliminates the complexity of having bridges replicate and transmit packets across multiple busses.

3.6 Message Types

The type of communication payload in an MCTP Message is identified by an enumeration called the "MCTP Message Type". MCTP is designed to be able to carry packets for multiple message types across a common communications medium. The MCTP base protocol specification includes definition of a message type for MCTP Control messages and message types that support Vendor-defined MCTP messages. Other message types, such as a future message type to support Platform Level Data Model (PLDM) are defined in separate specifications.

Different message types are identified by a message type field that is carried in the header of an MCTP message. The value in this field is specified in the specification that defines the message type. In addition, DSP0239, the Management Component Transport Protocol (MCTP) IDs and Codes specification, provides a centralized collection of the message types values defined for MCTP.

3.7 MCTP Control Messages and Control Protocol

The MCTP Base Specification includes the definition of a message type for "MCTP control messages". These messages are used by bus owners to initialize and maintain the addressing and routing used in an MCTP network. MCTP control messages also include messages to help determine the basic MCTP messaging capabilities of a endpoint, such as what MCTP version and message types it supports. Control messages that initiate or request an action are also referred to as "commands".

The MCTP Control Protocol defines the format of MCTP control messages and how they are delivered and acknowledged. Almost all control messages are delivered using a request / response mechanism where a request message is delivered to a target or "responder" endpoint using an MCTP message transfer, and then later the responder delivers a response message back to

the requester. Response messages provide a positive acknowledgement of the receipt and handling of the request and return parametric data (if any) according to the request.

The control protocol also supports delivering messages as unacknowledged “datagrams” and as broadcasts on the bus. These mechanisms are made available to support control messages that can be used for device discovery as required by the medium.

The control protocol includes a retry mechanism to help ensure reliable delivery of requests, and to cover cases where a responder may be briefly off-line. The control protocol defines fields that are used to identify requests and responses, to identify whether a message is a datagram, and to identify new instances of requests from retried instances.

3.8 Transport Binding Specifications

The Transport Binding specifications are documents that describe how MCTP is implemented on a particular physical medium. This includes the definition of MCTP packet formatting for the medium, source and destination physical addressing handling, physical layer data integrity, and medium-specific timing parameters.

The transport binding specifications include the definition of physical layer retry mechanisms for MCTP packets (if any).

4 Example MCTP Topology

The following example illustrates how MCTP can be used within a hypothetical platform management subsystem implementation. More complex topologies, with multi-levels of bridges and greater numbers of busses and devices can be readily supported by MCTP as required.

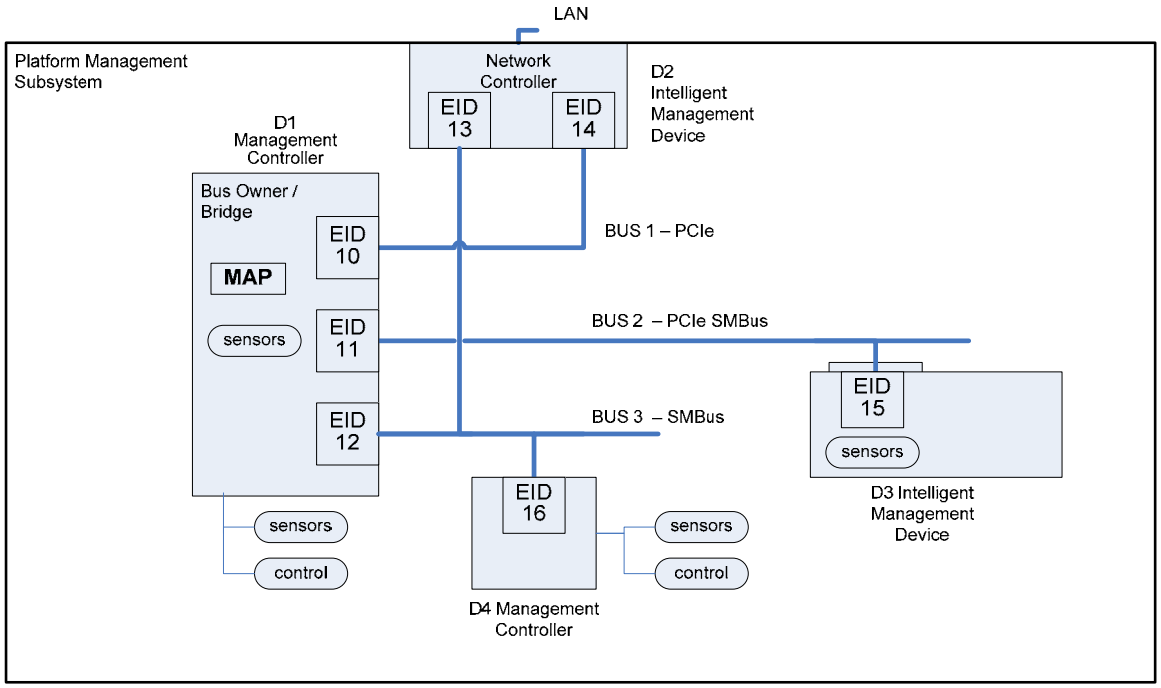

Figure 1 - Example MCTP Topology

The following table describes the hardware elements that make up the subsystem shown in Figure 1.

Element	Description
Bus 1	a PCIe bus that provides a high bandwidth connection between Management Controller D1 and Network Controller D2
Bus 2	an SMBus that connects to the PCI Express slots in the system
Bus 3	an SMBus that provides a separate connection for use by D2 and D4
D1	a microcontroller that provides MCTP bus owner and bridge functions for the motherboard and also provides the intelligence for a DMTF Management Access Point (MAP). (see DSP2014)
D2	a network controller that contains a sideband interface function that enables network traffic to be transferred using MCTP over either bus 1 or bus 2
D3	a PCI Express add-in card that contains sensors (e.g. error status, temperature, and fan sensors)
D4	in this example, D4 represents a management controller that also provides the intelligence for a DMTF Management Access Point (MAP). (see DSP2014)

In this example, D1 is a management controller that serves as intelligence for a DMTF MAP that enables system sensors and control functions such as temperature, voltages, fan speeds, power on/off/reset, etc. to be accessed using DMTF CIM data models. Network controller D2 provides

the network sideband interface that is used to transfer packets between D1 and a local area network.

Network controller D2 has two connections: an SMBus connection via bus 3 and a PCIe connection via bus 1. The SMBus connection is a low-power connection that is available in all system power states, but is low-bandwidth (< ~100kbps) while the PCIe connection is higher bandwidth, but is only available when the system is fully powered and after the BIOS/OS have initialized PCIe. The SMBus connection thus provides a way to communicate with the MAP while the system is powered down or in a sleep state, and the PCIe connection provides a connection that can deliver performance for functions such as media redirection, firmware updates, and so on, when the system is powered up.

D3 represents a PCIe add-in card that is connected to an SMBus, bus 2, that is routed across the PCIe connectors in the system. In this example, bus 2 is separated from bus 3 in order to prevent any possibility that erroneous traffic or hardware errors on bus 2 will affect the SMBus communication between D4, D2, and D1.

D4 represents a management controller that is dedicated to monitoring and control functions that are related to thermal management of the system. It uses the MCTP connections to access temperature sensors and fan speed data maintained by D1, D3, and itself.

5 Conclusion

MCTP provides a common, flexible, media-independent, and byte-efficient protocol to enable the interconnection of devices from multiple vendors within a platform management subsystem.

MCTP supports message types that will be defined by the DMTF PMCI workgroup to provide common access and abstraction of the low-level monitoring and control functions within the platform management subsystem that will in turn support the representation and access of platform manageability functions via higher level DMTF CIM data models, transports, and profiles.