

面向信息物理融合系统的多源异构数据交互模型

杨博涵¹, 燕雪峰¹, 郭丽琴²

(1. 南京航空航天大学计算机科学与技术学院, 南京 211106; 2. 北京电子工程总体研究所, 北京 100854)

摘要: 在信息物理融合系统(Cyber-physical system, CPS)中, 传统多源异构数据集成模型难以通过中间件实现异构系统间的概念层关系映射, 存在系统难以扩展和传输性能低等问题。基于上述问题和挑战, 提出了一个面向CPS的异构数据交互模型。设计数据对象模型实现物理系统和仿真系统高层概念映射; 定义监测类和控制类元数据, 针对不同的数据类型使用增量或全量字段更新以降低网络负载; 基于Protobuf协议设计系统的通信模型, 提高系统的扩展能力。基于该数据交互模型和高层体系结构(High level architecture, HLA)/数据分发服务(Data distribution service, DDS)系统中间件实现了一个CPS原型系统, 验证了模型的可用性并对比了报文的压缩性能。

关键词: 信息物理融合系统; 异构数据; 数据集成; 桥接中间件; 对象模型; Protobuf协议

中图分类号: TP391.9

文献标志码: A

Heterogeneous Data Integration Method for Cyber-Physical System

YANG Bohan¹, YAN Xuefeng¹, GUO Liqin²

(1. College of Computer Science and Technology, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China;

2. Beijing Institute of Electronic System Engineering, Beijing 100854, China)

Abstract: In the cyber-physical system (CPS), the traditional multi-source heterogeneous data integration model is difficult to map the conceptual layer between heterogeneous systems through middleware, which has the problems of low transmission performance and difficult system expansion. Due to the challenges above, a CPS-oriented heterogeneous data integration model is proposed. The data object model is designed to realize the high-level concept mapping between physical and simulation systems. The monitoring and control metadata are defined and the incremental or full field updates for different data types are used to reduce network load. A communication model of system is designed based on the Protobuf protocol to improve the system expansion capability. Based on the data interaction model and high level architecture (HLA)/data distribution service (DDS) system middleware, a CPS prototype system is implemented, which verifies the usability of the model and compares the compression performance of the message.

Key words: cyber-physical system (CPS); heterogeneous data; data integration; middleware; object model; Protobuf protocol

基金项目: 国家重点研发计划“云制造数字孪生技术研究”重点课题(2018YFB1702702)。

收稿日期: 2021-06-09; **修订日期:** 2021-09-21

引言

随着各类大型设备及系统复杂度的提高,信息物理融合系统被广泛应用于生产制造流程,实现对复杂装备或系统的建模仿真和实时控制。在云制造数字孪生技术研究课题中,拟设计一套面向智能制造生产线的信息物理融合系统。该系统可以在生产线设备运行过程中采集数据并转发至仿真环境,通过真实数据的驱动修正仿真模型;仿真平台能够实时监测生产线的运行状态并对其进行在线反馈控制,实现虚实系统的双向交互融合。

以数字孪生系统为代表的信息物理融合系统(Cyber-physical system, CPS)由3部分组成^[1]:(1)物理空间的真实设备及架设于其上的采集控制平台;(2)虚拟空间对物理实体的仿真模型;(3)连通虚实系统之间的数据。只要具备以上组成部分及相应的功能,基于不同规范设计的仿真或物理节点均可接入系统并展开交互,提供业务逻辑的模拟或设备数据的采集、传输和控制功能。然而,大量接入的设备在系统内采用不同的数据格式传输,产生的“信息孤岛”间无法直接进行通信和交互^[2]。因此,本文根据CPS的特点,分析了现有异构数据集成方法在异构系统间不利于协同交互的原因,提出了一个面向CPS的多源异构数据交互模型和通信模型。通过在高层体系结构(High level architecture, HLA)和数据分发服务(Data distribution service, DDS)上构建中间件及模型的集成,验证了这一模型的可行性和高效压缩性,为多领域CPS构建提供了解决方案。

1 异构数据集成与交互

1.1 异构数据集成相关研究

异构的数据不仅包括异构的数据存储类型,还包括异构的数据结构,例如结构化、半结构化和非结构化的数据^[3]。近年来,对异构数据集成的研究主要集中在异构数据库的数据融合上,例如,Korobko等^[4]提出了一个SMM(Simplified multidimensional model)三元组模型,实现了对异构数据库的多个关联维度进行映射;Liu等^[5]提出了一个HDIM(Heterogeneous data integration model),通过建立顶层视图,使用全局表和统一键值对建立从高层到底层数据的映射关系,其多层概念适用于多级架构的异构数据融合;Chromiak等^[6]面向多源异构数据库,提出了一个iDAO(interoperable Data access object)互操作数据接入对象模型,在异构数据库之间实现了数据集中与操作;Nigatu等^[7]基于ARDI方法,将JSON(JavaScript object notation)和可扩展标记语言(Extensible markup language, XML)等半结构化数据转换为使用RDFS(Resource description framework schema)模型统一表示的元模型形式。上述对异构数据库的数据集成方法主要以数据提取和可视化为导向,面向多用户提供统一的数据模型视图,在面向多源异构数据库的集成方法方面,针对数据单方向传输场景有一定的实用价值,能够较全面地提取和展示异构数据包含的信息,但此类模型和方法在结构和算法层面难以适用于CPS动态、双向交互数据的集成场景,为了提高CPS异构数据集成和交互的动态性和效率,需针对CPS特点设计多源异构数据的集成和交互模型。

1.2 CPS中的异构数据交互

信息物理融合系统通常由工业现场的传感器、执行器以及控制器和仿真信息系统组成,各组成部分在功能和设计方法上存在较大的异构性。仿真平台不仅能够监控、显示物理设备的状态数据,同时也是物理设备数字化的孪生模型;物理平台对数据进行实时采集和分发,响应仿真平台及控制器指令并向具体设备执行器发出操作命令。异构的程序设计语言、协议规范和设计模式导致CPS中的数据呈现出概念、数据结构、语法规义和数据格式4个方面的异构性。为了连通异构的系统,通常采用中间件或适配器结构并制定相应的转发规则实现数据转发^[8],若使用现有异构数据集成方法实现系统交互,需

要在系统设计阶段于中间件内预先确定数据表字段及格式,而CPS中存在各类设备和数据的动态加入和离开,在系统更换或升级时需要重新编写规则,可扩展性差。因此,需要设计面向CPS的异构数据集成模型,基于模型定义规则以提高系统扩展性。

在交互数据集成模型方面,赵军富等^[9]基于OSLC(Open services for technology and lifecycle collaboration)集成规范和HTTP协议建立Web服务,并采用URL标识物理系统设备,开展了数字化智能工厂监控整合研究,然而使用URL和HTTP请求会产生较大的冗余信息,在设备数量较大或传输速率较高时会成为性能瓶颈,且模型未加入时间参数,无法保证系统间消息的因果一致性;王宁等^[10]提出了一个基于CORBA(Common object request broker architecture)的OIM(Object integration model)模型,包括一个四元组 $\langle \text{OID}, d, t, v \rangle$,分别表示对象标识符、对象名、对象类型和对象值,实现了对异构数据基于面向对象的建模,但在数据类型上,交互对象间不仅存在由数值类型组成的监测类数据,还包括由操作和参数组成的控制类型数据,两类数据在封装格式、更新粒度和传输规则上存在差异,后者更侧重实时的交互控制,而该方法中提出的四元组未实现对控制类和监测类数据的区分,对监测类数据只能采用全量更新,增加了网络传输负载;李小龙^[11]通过对结构化和非结构化数据建模,提出了一个包含元模型和结构模型的统一异构数据表达式,实现了不同结构数据在数字孪生车间中的转发,但在系统发生变动时需要人工修改映射规则,不利于系统维护。在通信模型方面,对异构数据集成主要选取半结构化数据,如XML^[12-13]和JSON^[11,14-15]格式,但此类数据模型无法对系统内数据进行预先定义,在系统构建时缺少直观的全局视图,不利于业务功能的扩展。

综上所述,现有异构数据集成或交互模型未考虑不同交互数据类型的特点,存在通信报文冗余问题,影响系统的实时运行;而通信模型又存在转换效率低,难以支撑系统组件热插拔的问题。基于上述不足和CPS异构数据特点,本文提出了一个面向CPS的异构数据交互模型。该模型显著提高了系统扩展性和数据传输效率,根据模型的结构和交互算法,可以基于任意异构系统构造CPS原型系统。

2 多源异构数据交互模型

2.1 CPS交互数据对象模型

物理系统和仿真系统在设计形式上存在差异,因此数据交互模型需首先满足概念层次上的映射。同类型的所有设备组成一个元设备类,它们具有相同的订阅发布空间,因此可定义数据域主题号作为同类设备的唯一标识。在跨系统数据交互中,仅允许声明相同主题号的设备间通信,减少系统网络负载。同类的虚拟和物理设备成对存在,虚拟设备对应物理世界中某个真实存在的设备组成一个成员对概念,一个成员对在类空间内注册一个唯一的句柄标识。当系统同方向上有多台同类设备同时进行数据更新时,中间件缓存并合并多条数据,根据成员句柄进行区分和转发。

一个元设备类的数据域内存在多种数据类型,对于不同的数据类型定义唯一的类型名,同属一个元设备类的成员对可以对数据域内的各字段数据组合使用。数据类型分为由数值字段组成的监测类数据和具有瞬时特性的控制类型数据,定义交互类型标识对两种数据进行区分,从而能够针对不同类型数据采取不同的存储结构和转发规则。最后,模型还应包括通信的数据实体部分。基于该模型的整体系统功能结构如图1所示。

对概念模型进行描述,首先定义监测类元数据和控制类元数据。监测类元数据格式如式(1)所示。

$$\text{Meta_Data} = \{\text{Msg_Type}, \text{Struct_Data}\} \quad (1)$$

由于同类设备可能发布或订阅多种主题的数据,因此定义string类型的Msg_{Type}元素作为成员对之间传输的某个监测类消息的主题名,交互双方通过该主题名确认消息的类型,从而实现对对应字段的反序列化和获取;Struct_{Data}是监测类数据的实体部分,它是Msg_{Type}主题数据的子集,即可能包含该主题下

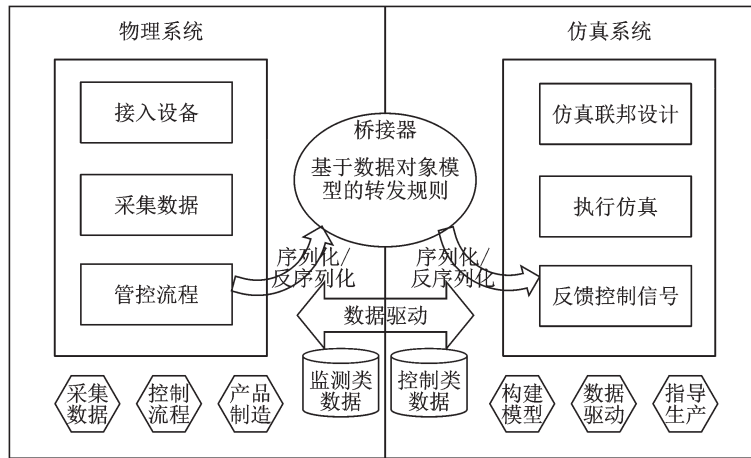


图1 基于CPS的功能结构示意图

Fig.1 Function and structure based on CPS

的全部字段或部分字段,在发送该监测类型消息时可以根据仿真或实际设备的状态实施增量更新,减少通信数据量,提高联合仿真性能。

控制类元数据格式表示为

$$\text{Meta_Data} = \{ \text{Msg_Type}, \text{Enum}, \text{Parameters} \} \quad (2)$$

式中:Enum代表一个由仿真系统或物理设备下达或执行的控制命令,是枚举类型;Parameters是其控制参数,由于控制类型的消息需要完整的控制动作及参数来确保其执行的正确性,因此Parameters包含一个控制命令的全部变量,只能进行全量更新。

包含监测类元数据和控制类元数据的数据对象模型表示为

$$\begin{aligned} \text{Obj_Model} = \{ \text{Obj_id}, \text{Ctrl}, \text{Meta_Data}, \text{TS} \} = \{ \text{Obj_id}, \text{false}, \text{Msg_Type}, \text{Struct_Data}, \text{TS} \} = \\ \{ \text{Obj_id}, \text{true}, \text{Msg_Type}, \text{Enum}, \text{Parameters}, \text{TS} \} \end{aligned} \quad (3)$$

式中:Obj_{id}表示元设备类内一个成员对的标识,是数值类型,在设备加入系统时注册和获取;Ctrl表示该消息是否为控制类型数据,属于布尔类型;true表示该消息是控制类型数据,false则表示该消息为检测类数据;TS表示时间戳,保证各交互消息按照时间顺序转发,维护系统因果一致性。

基于上述对象模型的表达式,可以定义CPS数据对象模型(CPS interaction model, CIM)为

$$\text{CIM} = \{ \text{Cls_id}, \text{Num}, \text{Obj_List} \} \quad (4)$$

式中:Cls_{id}代表系统内一个元设备类的标识;Num表示该条消息包含的对象模型消息数量,也即相同时间区间内发送消息的一类消息数量,是Numeric类型;Obj_List由Obj_Model组成,组合多个设备的数据对象模型数据。

CIM与传统异构数据集成模型相比具有以下3点优势:(1)增加了元设备类的概念,使用一个CIM可以存储相同类的多个对象模型,实现了虚实系统之间基于发布订阅机制的高效通信,系统开发人员只需要设计仿真系统及物理系统中产生和接收的数据类型,并定义设备类型和设备编号,即可通过CIM实现交互,简化系统构建流程;(2)使用类型ID和设备ID能唯一确定一台设备,仿真系统或物理系统根据标识进行数据定点转发,实现了精简数据、提高全系统吞吐量的目标;(3)增加了对数据类型的定义,区分监测类和控制类数据,中间件可根据类型确定不同的转发优先级以提高实时性,同时可实现字段增量更新,减少系统内不必要的通信量。

2.2 基于 Protobuf 协议的通信模型

在建立 CPS 中的数据对象模型后,对异构平台之间的通信数据结构进行建模,实现与数据模型间对应关系映射。与现有半结构化数据集成模型不同,本文采用了 Protobuf 序列化协议对 CPS 通信数据进行建模和传输。

Protobuf 协议是由 Google 开发的结构化数据序列化协议,使用与平台和开发语言无关的序列化机制,适用于通信协议、数据存储等多种领域。Protobuf 协议可高效作用于异构分布式进程之间的数据交互,因此可以构造 CPS 异构数据交互的通信模型。使用 Protobuf 协议进行异构数据交互具有 4 个主要优势:(1)结构上能够与 CIM 在对应概念进行映射;(2)便于系统扩展,修改 proto 文件即可实现设备快速更新;(3)使用简单,支持多种语言和平台,使用官方代码生成器即可生成对应语言的序列化和反序列化方法;(4)序列化方法高效,使用 Base 128 编码方式,黄斌等^[16]论证了其相较于 XML 和 JSON 格式的优势。信息物理融合系统通常要求较高的实时性,而 Protobuf 协议具有高效的编码及转换效率,可以充分压缩通信报文空间,当系统存在大量设备时可以有效减少拥塞并提高通信效率,从而保证 CPS 的实时性;工业现场的设备在生产制造过程中会时常更新或变动,如采用传统通信模型则需要频繁变更转发规则,不利于系统的扩展性和兼容性,使用 Protobuf 协议更新传输规则时,仅需在指定的 proto 文件中添加或删除字段,并生成新的序列化及反序列化方法即可由程序使用,不会影响未更新字段的发送和接收。基于 Protobuf 的通信模型与 CIM 的映射关系如表 1 所示。

一个“.proto”文件内声明了多个数据结构,并使用 Package 元素封装,这对应于一个元设备类。在系统设计初期,可以对系统内的各类设备预先进行发布和订阅数据的声明和定义,当类内数据结构发生变化时可以通过修改文件内数据定义实现快速更新;Message 是由设计人员封装的数据结构,对应监测类数据的数据类型或控制类数据的命令参数;buffer 是经过序列化后的 Protobuf 字节流,封装了 Struct_{Data} 对象数据或 Parameters 参数,包含一个对象模型 Obj_Model 的全部信息。Protobuf 概念与 CIM 模型的对应关系如表 1 所示。

在经过序列化的 buffer 上嵌套一层 Package 结构来表示 CPS 数据对象模型 Cls_{id}、Num 和 Obj_List。其中 Obj_List 用 repeated 类型存储。基于上述通信模型和数据交互模型的映射,对 CPS 系统异构数据设计交互算法,如图 2 所示。

表 1 Protobuf 结构与 CIM 元素映射
Table 1 Protobuf and CIM element mapping

序号	Protobuf 结构	CIM 元素
1	Package	Cls _{id}
2	Message	Msg _{Type} Enum
3	buffer	Struct _{Data} Parameters

```
1 string      ClassName = getPackageName();
2 int         Num       = getObjectNum();
3 repeated    Array[];
4 while (Num != 0) {
5     int      ObjectID  = getObjectID();
6     bool     Ctrl      = getObjCtrl();
7     string   buffer;
8     if (Ctrl == ture) {
9         foreach( para in Parameters ){
10             Set(para);
11         }
12         SerializeToString(buffer);
13     }else{
14         foreach( attr in StructData ){
15             if( Update(attr) ){
16                 Set(attr);
17             }else
18                 continue;
19         }
20         SerializeToString(buffer);
21     }
22     Array.append(buffer);
23     Num -- ;
24 }
25 Set(ClassName);
26 Set(Num);
27 SerializeToString(CIM);
```

图 2 基于 CIM 的异构数据交互算法
Fig.2 Heterogeneous data interaction algorithm based on CIM

基于CIM的异构数据交互算法步骤如下:(1)获取并存储元设备类名和该类的多个对象模型消息,确定对象数量;(2)遍历每个对象模型,获取对象ID并判断是否为控制类消息;(3)如果是控制类消息,则将其控制命令和全部参数序列化为buffer类型;(4)如果是监测类型消息,则对增量更新的字段进行序列化,忽略无需更新的字段;(5)将序列化完成的消息封装到一个数组内,并定义类名和数量,序列化为CIM,各类异构系统即可按照规则进行解析。使用本文所提模型进行CPS交互,只需声明发布或订阅的对象ID,并在proto文件中定义所包含的字段及类型,便于CPS中成员的更新与动态热插拔,而现有CPS交互模型通常采用JSON或XML格式封装模型信息,数据类型不直观且解析过程复杂,随着系统规模的增加,不利于系统维护及实时交互。

3 基于HLA/DDS的异构系统交互设计

3.1 数据模型与HLA/DDS概念映射

HLA是面向分布式仿真的体系结构,基于HLA规范实现的RTI软件提供仿真环境所需的通信及多种管理服务;DDS主要面向传输数据,提供订阅发布机制及多种QoS服务,支持接入多种不同的软硬件设备。基于上述特点,HLA和DDS可以作为CPS的仿真规范和设备数据采集规范。但是二者基于不同的协议,数据定义形式也存在较大差异,需根据异构数据交互模型和算法实现交互。

根据异构中间件的结构和性质,HLA/DDS桥接器应当能够分别与HLA和DDS系统进行通信,兼具仿真联邦成员和发布订阅参与者身份。表2是HLA、DDS和Protobuf协议对应概念模型之间的映射关系,其中:Federation是HLA中的仿真联邦,可以实例化为多个仿真联邦成员Federate;Domain是DDS中的数据域概念,一个数据域内可以包含多个参与者Participant;联邦成员和参与者都可以通过注册一个或多个对象类消息、交互类消息或主题数据来表示通信中的各类消息变量和参数结构;最后是数据实体部分,HLA对数据进行了对象类和交互类的区分,DDS使用Model元素统一表示,双方采用Protobuf序列化结构进行存储。

3.2 HLA/DDS结构与交互流程

HLA和DDS系统进行交互的流程如图3所示。HLA或DDS系统通过仿真推进或采集数据获取运行数据,随后仿真联邦成员或参与者进程根据所模拟或监管的模型和物理设备设置设备类型、设备编号、是否为控制类型消息和时间戳等公共数据部分;在发送数据时,联邦成员或参与者使用Protobuf提供的Set和SerializeToString方法可以将消息序列化为可直接进

表2 HLA、DDS与CIM模型映射关系

Table 2 Mapping relationships of HLA, DDS and CIM models

序号	HLA	CIM	DDS
1	Federation	Cls _{id}	Domain
2	Federate	Obj _{id}	Participant
3	Object	Msg _{Type}	Topic
4	Interaction	Enum	
5	Attribute	Struct _{Data}	Model
6	Parameter	Parameters	

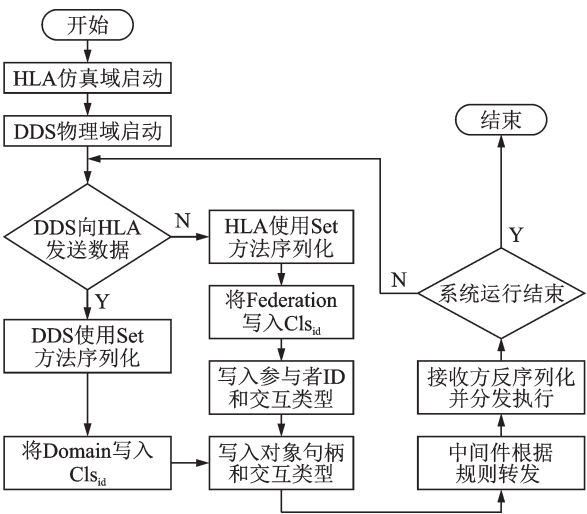


图3 HLA/DDS异构数据交互流程

Fig.3 HLA/DDS heterogeneous data interaction flow

行网络传输的字节流形式并存入字符串中;随后发送进程将公共数据部分封装到字符串前,组成一个 DDS 主题发布至中间件;桥接中间件根据消息类型和时间戳信息执行相应的调度方法,在 HLA 或 DDS 订阅发布域的层级实现消息转发,在此过程中,中间件对发送方和接收方都是透明的;接收成员通过 ParseFromString 方法将消息反序列化,即能够取得接收消息中数据实体部分的各个字段。经过上述步骤,异构的 HLA 和 DDS 系统即可实现交互通信,直至系统被仿真操作人员关闭或所有 HLA 仿真联邦成员、DDS 参与者断开与桥接器的连接。

4 实验验证

4.1 实验系统环境

根据前文对异构系统集成模型、框架和 HLA/DDS 集成方法的设计,实现了一个基于 CERTI 和 OpenDDS 软件 CPS 原型系统。其中, CERTI 为法国航空航天局开源 RTI 项目,使用 CERTI 作为 HLA 仿真中的运行支撑环境,主要负责模型构建及仿真运行,基于 OpenDDS 开发物理系统的仿真节点,从而可以构造一个半实物仿真系统环境。实验物理环境为多台通过网络互连的计算机,在 CentOS 7.4 操作系统下开发桥接中间件 HLA_DDS_Bridge,验证了基于 CIM 集成模型和 Protobuf 通信模型的 HLA 与 DDS 异构系统连通性,并对使用 CIM 模型和其他模型方法传输的报文大小进行了对比。

4.2 CPS 原型系统实现

实验构建了一个家具生产线的原型系统,包括厂区智能设备共 3 类(包括精密锯、镗铣机和立体货架),以及 3 类厂区环保安全检测设备(环保检测仪、电气检测仪和消防设备),各设备通过 OpenDDS 参与者实现平台接入。HLA 仿真平台对设备建模并同步运行仿真,实现对设备的监控、预警和反馈控制。系统共定义交互数据 30 种,其中监测类数据 20 种,控制类数据 10 种。首先对系统交互数据建立 Protobuf 格式的数据声明文件,并使用 protoc 工具生成数据的序列化和反序列化方法。定义 DDS 各参与者订阅或发布的数据域,在 IDL 文件中进行声明,各字段的类型、语义与 Protobuf 格式数据相同。声明 HLA 仿真系统中的对象类属性和交互类参数。最后根据 CIM,在 FED 文件中加入数据对象模型的声明,一侧系统将数据依照算法流程嵌入数据对象模型并经过中间件发送至对侧系统,对侧系统接收并反序列化,即完成一次数据交互。原型系统中所使用的部分数据声明如图 4 所示。

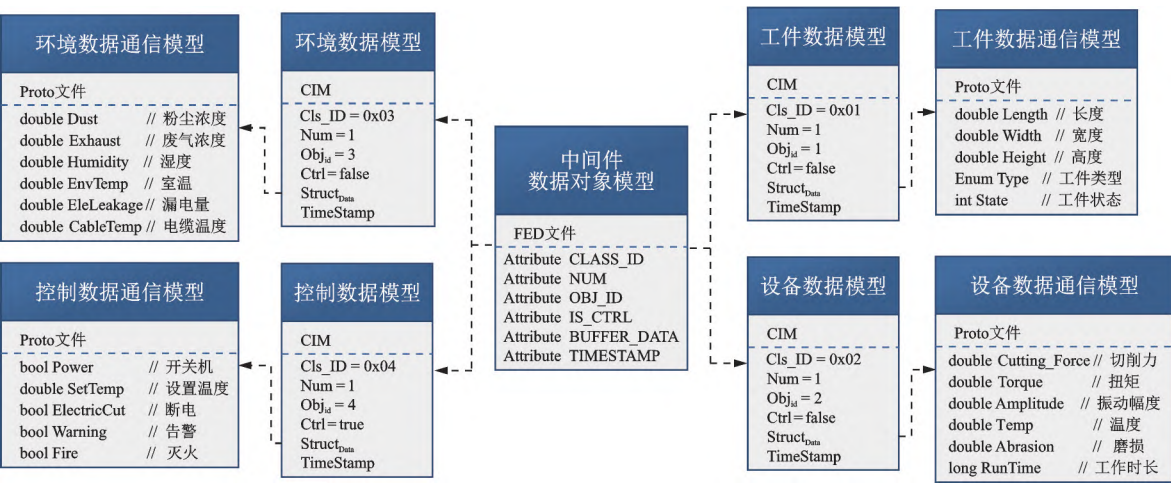


图 4 CPS 原型系统异构数据声明
Fig.4 Data declaration of the CPS prototype

首先启动桥接中间件及其依赖的运行支撑软件 CERTI,将接入物理设备的 OpenDDS 参与者加入至桥接器的订阅发布域中,运行各设备对应的虚拟模型设备并开始仿真。开始仿真后,DDS 参与者将采集到的数据周期性发送至桥接器,桥接器根据收到消息的 $Cl_{s_{id}}$ 和 Obj_{id} 在 HLA 订阅域中发布,对该设备感兴趣的虚拟设备将会收到该消息,通过反序列化操作即可得到数据的主体部分。以环境监测数据为例,使用 DDS 参与者模拟了温度、含氧量的持续下降,并模拟压力均匀上升,联邦成员在发现环境数据异常后会执行相应的反馈控制,保障物理系统维持稳定的环境参数。交互效果如图 5 所示。



图5 仿真联邦成员及物理系统数据交互示例
Fig.5 Federate and physical system data interaction

4.3 数据压缩性能对比

为了验证 CIM 数据交互模型和 Protobuf 的压缩性能,对比了使用 CIM 及 Protobuf 通信模型、仅使用 CIM、仅使用 Protobuf 协议和使用传统数据集成模型并用 JSON 保存的报文长度。实验中所使用的结构体包含 4 个整型、2 个浮点型、2 个布尔类型和 1 个长度为 6 B 的字符串,实验总共分为 15 次,前 14 次采用增量更新逐次增加更新的字段数量,第 15 次采用全量更新,即全部字段均需要更新发送。为平均 Protobuf 协议对数值类型变量的压缩性能,对每次实验更新的整型字段数值在 $0 \sim 2^{32}$ 内进行递增,记录交互过程中压缩报文的长度并计算多次循环平均值作为传输消息的空间占用量。交互数据字段更新组合如表 3 所示。各方法每次传输的报文长度对比如图 6 所示。结果表明,使用 CIM 数据交互模型进行传输时,传输报文在更新字段数较少时占用极小的空间,能够极大地降低网络负载。在全量更新时,由于 Protobuf 协议具有较好的压缩性能,相比使用 JSON 传输也实现了报文空间的压缩。在实际应用场景中,假设以上述实验为例的监测类数据结构在发布更新时,15 种组合方式出现的概率分布均匀,考虑 Protobuf 协议的压缩性能,统计对 $0 \sim 2^{24}$ 之间的较小数值压缩后占字节数平均为 16.5,对 $2^{24} \sim 2^{32}$ 之间的较大数值压缩后占字节数平均为 19.8,而不使用所提模型进行压缩的消息字节数平均为 40,可得使用本文设计的数据交互

表 3 交互数据字段更新组合
Table 3 Combination of update field in interactive data

实验编号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
整型数量	1	2	2	3	3	3	4	2	2	3	2	3	4	4	4
布尔类数量	0	0	1	0	1	2	0	0	1	1	1	1	1	2	2
浮点型数量	0	0	0	0	0	0	0	1	1	1	2	2	2	2	2
字符串数量	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

互模型的网络负载可降至40%~50%。

4.4 序列化及传输性能对比

为了测试所提出模型对交互实时性的提升,在所开发的CPS平台上对CIM模型和传统CPS集成所使用的JSON及XML模型进行了对比。实验分别使用35种不同大小的数据进行交互,记录并比较3种模型的平均消息传输延迟。实验在3台机器上进行,各机器之间使用以太网进行通信,交互流程为DDS参与者将数据通过3种不同的模型进行序列化封装,随后发送序列化数据至桥接器,桥接器接收后根据转发规则发送至HLA仿真联邦成员,联邦成员接收到数据后进行反序列化,对部分字段进行更新并再次序列化,最后通过桥接器发回DDS参与者完成一次交互。交互过程包括2次序列化及反序列化操作和2次消息传输,每种模型进行10万次交互并取平均值,对比3种模型的平均延迟时间,结果如图7所示。实验结果表明,随着交互数据量的增加,使用本文所提模型进行封装解析并传输的平均延迟能够基本保持稳定,而使用传统的JSON和XML模型进行传输的平均延迟则出现显著增长,数据量为5万字节时,CIM传输速度约为JSON的2.4倍,XML的5倍;数据量为30万字节时,CIM传输速度约为JSON的5.9倍,XML的19倍。相比于现有CPS所采用的数据交互模型,本文所提模型可实现快速消息序列化并高效压缩数据,在进行大数据量传输时能够保证CPS的交互实时性。

5 结束语

本文基于CPS特点,提出了一个面向CPS的异构数据交互模型,使用Protobuf协议进行异构数据的集成和传输,并在HLA/DDS联合仿真系统上进行了验证。结果表明:使用本文提出的模型进行CPS数据部分的构造较为直观,能够以面向对象的方式对系统间的异构数据进行集成;实现中间件对整体系统透明,提高了系统扩展性;对监测类数据的增量更新能够降低网络传输负载,提升网络性能;使用Protobuf协议的序列化方法可以快速对异构数据进行包装,提高系统的构建效率,实现系统成员动态热插拔。因此,该模型作为一个可通用的模型,能够集成跨平台的多源异构数据,为CPS构造提供了解决方案,可以提高系统灵活性。本文的下一步工作包括以下两点:(1)面向更复杂的场景,对半结构化与非结构化数据定义统一的集成交互语义,实现视频、图像等信息的交互传输;(2)在大规模CPS中,目前所使用的向量时间会产生较大的空间开销,需在保证因果一致性的前提下设计方案进一步降低消息空间复杂度。

参考文献:

- [1] 杨林瑶,陈思远,王晓,等.数字孪生与平行系统:发展现状、对比及展望[J].自动化学报,2019,45(11):2001-2031.
YANG Linyao, CHEN Siyuan, WANG Xiao, et al. Digital twins and parallel systems: State of the art, comparisons and prospect[J]. Acta Automatica Sinica, 2019, 45(11): 2001-2031.
- [2] 陶飞,张萌,程江峰,等.数字孪生车间——一种未来车间运行新模式[J].计算机集成制造系统,2017,23(1):1-9.
TAO Fei, ZHANG Meng, CHENG Jiangfeng, et al. Digital twin workshop: A new paradigm for future workshop[J]. Comput-

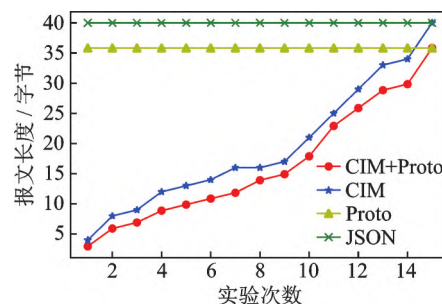


图6 报文长度对比结果

Fig.6 Message length comparison results

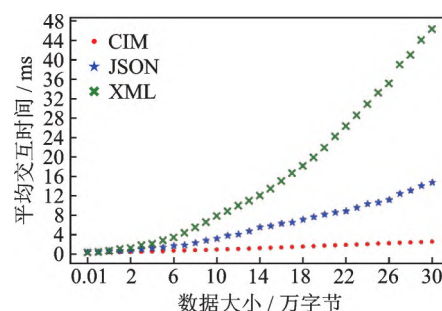


图7 不同数据量下异构数据交互时间对比

Fig.7 Comparison of heterogeneous data interaction time under different data volumes

- er Integrated Manufacturing Systems, 2017, 23(1): 1-9.
- [3] MESITIM, JIMÉNEZ-RUIZ E, SANZ I, et al. XML-based approaches for the integration of heterogeneous bio-molecular data[J]. BMC Bioinformatics, 2009, 10(12): 1-18.
 - [4] KOROBKO A, KOROBKO A. Matching disparate dimensions for analytical integration of heterogeneous data sources[C]//Proceedings of MEDES'19: 11th International Conference on Management of Digital EcoSystems. [S.l.]: ACM, 2019: 66-72.
 - [5] HAI L, LIU Y, WU Q, et al. A heterogeneous data integration model[C]//Proceedings of Geo-informatics in Resource Management and Sustainable Ecosystem. Wuhan, China: Communications in Computer and Information Science, 2013: 298-312.
 - [6] CHROMIAK M, STENCEL K. A data model for heterogeneous data integration architecture[C]//Proceedings of International Conference: Beyond Databases, Architectures and Structures. Ustron, Poland: Communications in Computer and Information Science, 2014: 547-556.
 - [7] NIGATU S T, CRISTINA G, ROMERO O, et al. ARDI: Automatic generation of RDFS models from heterogeneous data sources[C]//Proceedings of International Enterprise Distributed Object Computing Conference. Paris, France: IEEE, 2019: 190-196.
 - [8] 李洪阳, 魏慕恒, 黄洁, 等. 信息物理系统技术综述[J]. 自动化学报, 2019, 45(1): 37-50.
LI Hongyang, WEI Muheng, HUANG Jie, et al. Survey on cyber-physical systems[J]. Acta Automatica Sinica, 2019, 45(1): 37-50.
 - [9] 赵军富, 李建军. 基于物联网数字工厂与智能工厂的整合研究[J]. 制造技术与机床, 2020, 4(4): 20-25.
ZHAO Junfu, LI Jianjun. Integration research of digital factory with smart factory based on Internet of thing[J]. Manufacturing Technology & Machine Tool, 2020, 4(4): 20-25.
 - [10] 王宁, 陈滢, 俞本权, 等. 一个基于CORBA的异构数据源集成系统的设计[J]. 软件学报, 1998, 9(5): 378-382.
WANG Ning, CHEN Ying, YU Benquan, et al. Design of a heterogeneous data integration system based on CORBA[J]. Journal of Software, 1998, 9(5): 378-382.
 - [11] 李小龙. 基于数字孪生的机床加工过程虚拟监控系统研究与实现[D]. 成都: 电子科技大学, 2020.
LI Xiaolong. Research and implementation of virtual monitoring system for machine tools process based on digital twin[D]. Chengdu: University of Electronic Science and Technology of China, 2020.
 - [12] NASSIRI H, MACHKOUR M, HACHIMI M. Integrating XML and relational data[J]. Procedia Computer Science, 2017, 110(1): 422-427.
 - [13] 刘丽娜. 一种基于XML的异构数据集成共享模式研究[J]. 中小企业管理与科技, 2019(18): 191-192.
LIU Lina. Research on an heterogeneous data integration and sharing mode based on XML[J]. Management & Technology of SME, 2019(18): 191-192.
 - [14] 郭磊, 陈兴玉, 张燕龙, 等. 面向智能制造终端的车间生产数据采集与传输方法[J]. 机械与电子, 2019, 37(8): 21-24.
GUO Lei, CHEN Xingyu, ZHANG Yanlong, et al. Workshop production data collection and transmission method for intelligent manufacturing terminals[J]. Machinery and Electronics, 2019, 37(8): 21-24.
 - [15] 梁志宏, 陆歌皓, 黄宇翔, 等. 基于JSON的异构数据集成的研究[J]. 计算机科学与应用, 2018, 8(3): 314-322.
LIANG Zhihong, LU Gehao, HUANG Yuxiang, et al. Research on heterogeneous data integration based on JSON[J]. Computer Science and Application, 2018, 8(3): 314-322.
 - [16] 黄斌, 谢艳新, 唐友, 等. 数据存储信息序列化完整性及效率评估仿真[J]. 计算机仿真, 2020, 37(4): 159-163.
HUANG Bin, XIE Yanxin, TANG You, et al. Data storage information serialization completeness and efficiency evaluation simulation[J]. Computer Simulation, 2020, 37(4): 159-163.

作者简介:



杨博涵(1996-),男,硕士研究生,研究方向:系统建模与仿真, E-mail: yangbohan@nuaa.edu.cn.



燕雪峰(1975-),通信作者,男,教授,博士生导师,研究方向:智能建模、大数据、MBSE、复杂系统建模与仿真实理论和方法, E-mail: yxf@nuaa.edu.cn.



郭丽琴(1976-),女,研究员,研究方向:复杂系统建模与仿真。

(编辑:张黄群)