



重庆交通大学  
CHONGQING JIAOTONG UNIVERSITY

## 本科毕业设计（论文） 相关材料

题目：基于分布式技术的文件存储管理系统  
研究与实现

学    院： 信息科学与工程学院  
专    业： 计算机科学与技术  
学 生 姓 名： 徐博文  
学    号： 631602010429  
指 导 教 师： 邢荣军

重庆交通大学  
CHONGQING JIAOTONG UNIVERSITY

## 2020 届毕业论文（设计）相关材料目录

- 1、毕业论文（设计）任务书
- 2、开题报告
- 3、外文翻译
- 4、指导记录表
- 5、答辩资格审查表

# 重庆交通大学

## 毕业论文（设计）任务书

题    目：基于分布式技术的文件存储  
管理系统研究与实现

学    院：信息科学与工程学院

专    业：计算机科学与技术

学生姓名：徐博文

学    号：631602010429

指导教师：邢荣军

### 主要内容:

- (1) 熟悉、分析、理解此次设计项目所涉及的基本资料并学习相关技术, 包括分布式系统的特点、分布式文件系统的特点、Hadoop 的特点等。
- (2) 利用多台虚拟机搭建集群, 在该集群上搭建分布式 Hadoop 平台、搭建 HDFS 分布式文件系统。
- (3) 深入学习 Hadoop 平台上的 HDFS 分布式文件系统, 了解其工作原理与特性。
- (4) 利用多台虚拟机上搭建 Hadoop 分布式集群进行分布式文件系统相关功能的设计与开发。
- (5) 运用 HDFS 分布式文件系统, 实现分布式文件存储管理系统。
- (6) 利用 H2 嵌入式数据库对用户信息这类结构化数据信息进行存储。
- (7) 运用 Hadoop 上的 HBase 数据库, 实现对实时数据信息的存储与调用。
- (8) 利用 Java 开发可视化网页程序的功能接口。
- (9) 利用 Vue.js 开发用户可视化界面。

### 任务要求:

- (1) 搭建分布式 Hadoop 平台
  1. 在多台服务器上搭建分布式 Hadoop 平台。选择其中一台服务器作为主服务器, 并在整个分布式平台上搭建 HDFS 分布式文件系统。
  2. 以 HDFS 分布式文件系统为基础, 开发基于分布式技术的文件存储管理系统。
- (2) 对分布式系统的工作原理与特性进行研究
  1. 对分布式系统种的元数据管理技术、副本技术、一致性理论等关键技术进行学习与探究, 了解其原理与工作特性, 分析其优劣势。
  2. 对 Hadoop 与 HDFS 的特点进行学习与探究, 了解其基础架构并进行实现。
- (3) 对存储的文件数据进行分析处理
  1. 通过 HDFS 对文件进行分布式存储管理。
  2. 利用 Java 语言开发后端功能, 实现文件上传、下载与删除等文件系统基本功能。
  3. 利用 H2 嵌入式数据库对系统中结构化数据进行存储, 如用户信息。
- (4) 实现用户界面的可视化
  1. 利用 Java 语言开发 web 程序, 为用户操作界面提供功能接口。
  2. 利用 Vue.js 语言开发前端页面, 实现用户的可视化界面。
  3. 注意前后端交互时信息传递的准确性与可靠性。

进度计划:	
周次	主要工作计划 (内容)
1	规范开发文档, 阅读学习相关论文。
2	搭建开发环境, 设计数据库, 实现对相关数据的存储与调用。
3	学习 HDFS 相关知识, 对分布式文件系统的主要原理进行探究。
4	学习 HDFS 相关知识, 对分布式文件系统的主要原理进行探究。
5	利用 Java 语言进行后端功能开发, 实现分布式文件存储系统的主要功能。
6	利用 Java 语言进行后端功能开发, 实现分布式文件存储系统的主要功能。
7	利用 Java 语言进行后端功能开发, 实现分布式文件存储系统的主要功能。
8	实现系统的各项管理功能。
9	实现用户可操作界面。
10	实现用户可操作界面。
11	程序测试。
12	程序测试, 完成毕业论文。
接受任务日期: 2020 年 2 月 25 日                      要求完成日期: 2020 年 5 月 29 日	
学生接受任务 (签名):    指导教师签名:	

**备注:** 1. 此任务书由指导教师填写, 学生接受任务并签名, 并应在规定时限内完成教师布置的毕业设计 (论文) 任务;

2. 此任务书必须在毕业设计 (论文) 开始一周内下达给学生;

3. “主要工作计划 (内容)” 填写各阶段工作的具体内容和要求, 在开展毕业设计 (论文) 过程检查时将作为对学生进度检查的依据。

# 重庆交通大学

## 毕业论文（设计）开题报告

题    目： 基于分布式技术的文件存储  
管理系统研究与实现

学    院： 信息科学与工程学院

专    业： 计算机科学与技术

学生姓名： 徐博文

学    号： 631602010429

指导教师： 邢荣军

## 一、选题背景

此题目为《基于分布式技术的文件存储管理系统研究与实现》，其核心为利用分布式系统实现对相关文件的存储、调用与展示。

如今社会已进入大数据时代，在工作与生活中，我们每时每刻都在面临着大量的数据与文件。这些文件数量繁多而且种类繁杂，对存储工作也提出了巨大的挑战。而且，单机系统处理能力的升级仍存在瓶颈，其稳定性与可用性难以达到升级换代的要求。继续升级单机处理能力的性价比越来越低，分布式系统应运而生。

分布式系统是多个处理机通过通信线路互联而构成的松散耦合的系统，使一组独立的计算机成为一个统一的整体。系统拥有多种通用的物理和逻辑资源，可以动态的分配任务，分散的物理和逻辑资源通过计算机网络实现信息交换。具有分布性、自治性、并行性与全局性四大特征。具备实现资源共享、加快运算速度、可靠性高、通信快捷等优点。

此次，我们将使用 Hadoop 分布式系统基础框架。通过搭建分布式 Hadoop 平台以实现对相关文件的分布式存储，实现对文件的分布式存储、调用与展示，从而完成相应选题的设计与开发。

为完成此选题，我们将从搭建分布式 Hadoop 平台入手。随后，以其自身实现的分布式文件系统（HDFS）为基础，在我们自己的多台服务器上形成分布式集群，实现对相关文件的分布式存储。最后，利用 Java、Python 等高级语言开发相应的 web 程序，实现供用户使用的可视化界面，完成对服务器上的分布式系统的调用与展示。

## 二、国内外现状

自 1946 年世界上第一台电子数字计算机诞生以来，世界进入了崭新的时代。计算机飞速发展，帮助人们解决各种问题，如大规模的计算、大量文件的存储。70 多年来，单台计算机的性能不断提升，向不同方向的发现也满足着人们的各类需求。但随着计算机处理的数据量与文件量越来越大，单机系统处理能力的升级出现了瓶颈，提高硬件的计算性能代价已经越来越高。人们发现在一台超高级计算机上提高运算速度的成本远远大于使用多台普通计算机的成本，分布式系统应运而生。

分布式系统<sup>[1]</sup>诞生于 20 世纪 70 年代，从最开始简单的多机文件夹共享到广义文件共享，从单一计算模型到多种多层次的计算模型，从封闭的局域网络到开放的全球网络，分布式系统已经嫣然演变成了近代计算机系统的基本组织结构，与此同时它自身也脱身与传统的学术产物，逐渐变成商业产品<sup>[2]</sup>。

在现代的计算机模式当中，分布式系统代表了大型的，结构复杂的一类模型<sup>[3]</sup>。

分布式系统及其理论凭借着对系统资源出色的整合与调度，以及成熟的网络技术的支持，完成高性能计算中计算资源与存储资源的统筹规划，进而提供出色的计算能力；同时利用分布式系统资源分散的特点，使的集群系统在管理统一的情况下，具有更高的系统安全性和可靠性。分布式系统凭借其客户端分散的特点，也逐渐成为了云计算时代<sup>[4]</sup>重要的组成部分。

Andrew S. Tanenbaum 给出的分布式系统的定义是：“A distributed system is a collection of independent computers that appear to the users of the system as a single computer.”<sup>[5]</sup>。在一个分布式系统中，每一台计算机都是独立的，但其展现给用户的是一个统一的整体。其具有良好的内聚性和透明性，其中每一个计算机节点都相互独立，具有内聚性的特点，而对于上层应用和用户来说，分布式系统却如同一台性能强劲的“超级计算机”，底层的一切细节对于他们来说都是透明的<sup>[6]</sup>。分布式系统作为由多个处理机通过通信线路互联而构成的松散耦合的系统，拥有多种通用的物理和逻辑资源，可以动态的分配任务，分散的物理和逻辑资源通过计算机网络实现信息交换。同时，分布式系统由硬件和软件共同构成，系统以组件的形式分布在互联的计算机节点上，组件之间的动作协调通过信息传递的方式完成<sup>[7]</sup>。

分布式存储已在行业内处于重要地位。Apache 旗下的 Hadoop<sup>[8]</sup> 是能够运行在廉价硬件设备之上的分布式计算平台。Hadoop 的分布式文件系统组件 HDFS<sup>[9]</sup> 为众多项目提供了分布式存储解决方案，是目前云计算领域研究的焦点。国内很多知名公司都是用 HDFS 建设云计算基础设施，中国移动研究院开发的 Big Cloud, 华为的云计算平台，百度的云计算平台都是以 HDFS 为基础。因此，研究 HDFS 在云计算和云储存领域都具有非常重要的意义<sup>[10]</sup>。此次，我也将以 HDFS 为研究基础，对其提供的云存储解决方案加以实现。

GFS<sup>[11]</sup> (Google File System) 是 Google 为了存储公司内部海量搜索数据而开发的分布式文件系统，HDFS 是 GFS 的开源实现。Google 在其中部署了多个 GFS 的集群来存储数据。在 GFS 的实现中，它采用了数千台廉价的存储设备组成集群来对外服务，而不是使用一台性能超强的打大型机设别，使用了水平扩展而不是垂直扩展。

分布式存储系统仍存在一些问题。因为使用廉价机器，其硬件故障与软件故障几乎无法避免。为使整个系统正常运行，需要在软件上面进行错误处理和数据恢复来保证集群的可靠性，容错技术至关重要。大多数分布式存储系统都会采用多副本的方式来提高系统的容错能力。纠错码<sup>[12]</sup> 是目前分布式存储所采用的比较新的容错技术。此外，副本技术是存储系统中的一种常见的数据管理技术<sup>[13]</sup>。该技术也可增加系统的可



用性与可靠性，又能使得系统负载均匀，提高存储系统的 IO 吞吐量。

由此可见，存储的数据和系统状态信息的持久化是分布式存储系统要面对的主要挑战，保证数据的一致性也是分布式存储系统发展至今的难题之一。Brewer, E 在 2000 年提出了著名的“CAP 理论”<sup>[14,15]</sup>，该理论对分布式系统的设计和实现有着巨大的影响。

### 三、主要内容

(1) 熟悉、分析、理解此次设计项目所涉及的基本资料并学习相关技术，包括分布式系统的特点、分布式文件系统的特点、Hadoop 的特点等。

(2) 利用多台虚拟机搭建集群，在该集群上搭建分布式 Hadoop 平台、搭建 HDFS 分布式文件系统。

(3) 深入学习 Hadoop 平台上的 HDFS 分布式文件系统，了解其工作原理与特性。

(4) 利用多台虚拟机上搭建 Hadoop 分布式集群进行分布式文件系统相关功能的设计与开发。

(5) 运用 HDFS 分布式文件系统，实现分布式文件存储管理系统。

(6) 利用 H2 嵌入式数据库对用户信息这类结构化数据信息进行存储。

(7) 运用 Hadoop 上的 HBase 数据库，实现对实时数据信息的存储与调用。

(8) 利用 Java 开发可视化网页程序的功能接口。

(9) 利用 Vue.js 开发用户可视化界面。

### 四、方案与计划

#### (1) 搭建分布式 Hadoop 平台

1. 在多台服务器上搭建分布式 Hadoop 平台。选择其中一台服务器作为主服务器，并在整个分布式平台上搭建 HDFS 分布式文件系统。

2. 以 HDFS 分布式文件系统为基础，开发基于分布式技术的文件存储管理系统。

#### (2) 对分布式系统的工作原理与特性进行研究

1. 对分布式系统种的元数据管理技术、副本技术、一致性理论等关键技术进行学习与探究，了解其原理与工作特性，分析其优劣势。

2. 对 Hadoop 与 HDFS 的特点进行学习探究，了解其基础架构并进行实现。

#### (3) 对存储的文件数据进行分析处理

1. 通过 HDFS 对文件进行分布式存储管理。

2. 利用 Java 语言开发后端功能，实现文件上传、下载与删除等文件系统基本功能。

3. 利用 H2 嵌入式数据库对系统中结构化数据进行存储，如用户信息。

#### (4) 实现用户界面的可视化

1. 利用 Java 语言开发 web 程序，为用户操作界面提供功能接口。
2. 利用 Vue.js 语言开发前端页面，实现用户的可视化界面。
3. 注意前后端交互时信息传递的准确性与可靠性。

#### 五、主要参考文献

- [1] George Coulouris, Jean Dollimore, Time Kindberg. Distributed Systems Concept and Design Forth Edition[M]. New Jersey: Pearson Education, 2007
- [2]刘玮. 分布式系统性能测试与分析方法研究[D]. 大连:大连理工大学, 2011
- [3]周傲英, 周敏奇, 钱卫宁, 张蓉. 大规模分布式系统中的多属性查询处理[J]. 计算机学报, 2008, 31(9):1563-1572
- [4]Matthias Schmidt, Niels Fallenbeck, Matthew Smith, Bernd Freisleben. Efficient Distribution of Virtual Machines for Cloud Computing[C]//2010 18th Euromicro Conference on Parallel, Distributed and Network-based Processing, Pisa, Italy, 2010:567-574
- [5] Andrew S.Tanenbaum. Distributed Operating Systems[M]. Prentice Hall, 1995
- [6]李西宁. 分布式系统[M]. 北京:科学出版社, 2006
- [7]陆嘉恒, 刘继荣. 分布式系统及云计算概论[M]. 北京:清华大学出版社, 2011
- [8]Hadoop Apache. Hadoop[M]. 2009
- [9]Shvachko K, Kuang H, Radia S, et al. The hadoop distributed file system[C]//Mass Storage Systems and Technologies(MSST), 2010 IEEE 26th Symposium on. IEEE, 2010:1-10
- [10]崔园. 基于 HDFS 的分布式存储系统的研究与实现[D]. 成都:电子科技大学, 2016
- [11]Ghemawat S, Gobioff H, Leung S T. The Google file system[C]//ACM SIGOPS operating system review. ACM, 2003, 37(5):29-43
- [12]Rizzo Luigi. Effective erasure codes for reliable computer communication protocols[J]. ACM SIGCOMM computer communication review, 1997, 27(2):24-36
- [13]Bhagwan Ranjita, Moore David, Savage Stefan, et al. Replication strategies for highly available peer-to-peer storage [M]. Springer, 2003.
- [14]Brewer Eric. CAP twelve years later: How the "rules" have changed[J]. Computer, 2012, 45(2): 23-9.
- [15]Browne Julian. Brewer's CAP theorem[J]. J Browne blog, 2009.

开 题 纪 要			
时 间	2020-3-16	地 点	腾讯会议
参与人员	陈松、邢荣军、徐博文		
开 题 工 作 记 录	<p>简述开题报告：</p> <p>搭建 Hadoop 基础环境，并在 Hadoop 分布式环境下编写相应程序，完成对应程序的编写，实现文件、数据的存储和读取，理解文件分布式存储的原理。</p> <p>分析并理解 HDFS 的相关优缺点，尝试对这些缺点进行改进。</p> <p>在完成 Hadoop 平台的搭建与 HDFS 的实现后，用高级语言编写用户可视化操作界面。</p> <p>学生签名： <span style="float: right;">2020 年 3 月 16 日</span></p>		
指 导 教 师 意 见	<p>受时间与空间影响，将原定于在物理机上搭建平台改为在虚拟机上搭建。并将需要结合实际现场的数据处理、数据建模部分改为对文件的存储与读取。</p> <p><input checked="" type="checkbox"/>同意开题                      <input type="checkbox"/>不同意开题</p> <p>教师签名： <span style="float: right;">2020 年 3 月 16 日</span></p>		

# 重庆交通大学

## 本科毕业论文（设计）

### 外文翻译

译文题目： 一种信息物理系统的元模型  
学 院： 信息科学与工程学院  
专 业： 计算机科学与技术  
学生姓名： 徐博文  
学 号： 631602010429  
指导教师： 邢荣军  
完成时间： 2020年3月13日

# 一、英文原文

出版日期（期刊号）： Volume 41, August 2019, 100930

出版单位（刊物名称）： Advanced Engineering Informatics

A metamodel for cyber-physical systems

Abstract

With the advent of the Internet of Things and Industry 4.0 concepts, cyber-physical systems in civil engineering experience an increasing impact on structural health monitoring (SHM) and control applications. Designing, optimizing, and documenting cyber-physical system on a formal basis require platform-independent and technology-independent metamodels. This study, with emphasis on communication in cyber-physical systems, presents a metamodel for describing cyber-physical systems. First, metamodeling concepts commonly used in computing in civil engineering are reviewed and possibilities and limitations of describing communication-related information are discussed. Next, communication-related properties and behavior of distributed cyber-physical systems applied for SHM and control are explained, and system components relevant to communication are specified. Then, the metamodel to formally describe cyber-physical systems is proposed and mapped into the Industry Foundation Classes (IFC), an open international standard for building information modeling (BIM). Finally, the IFC-based approach is verified using software of the official IFC certification program, and it is validated by BIM-based example modeling of a prototype cyber-physical system, which is physically implemented in the laboratory. As a result, cyber-physical systems applied for SHM and control are described and the information is stored, documented, and exchanged on the formal basis of IFC, facilitating design, optimization, and documentation of cyber-physical systems.

### 3. Communication in cyber-physical systems

To describe cyber-physical systems using metamodeling approaches, system components, including attributes and methods, need to be defined. As a basis to characterize attributes and methods, in this section, network topologies applied for communication in cyber-physical systems are studied. Also, a survey of network communication is provided, and communication protocols suitable for cyber-physical systems are reviewed. Developing a metamodel capturing communication-related information, such as system components related

to communication, network topologies, network communication characteristics, and communication protocols, enables distinct advantages, such as effective planning of SHM and control systems, improved data control and management in the operation and maintenance phase of CPS, or linking of SHM and control system components with communication-related information.

As nodes of wireless sensor networks are spatially distributed, autonomous devices, power consumption and resource management are important criteria in designing communication protocols and network topologies [31]. Topologies typically applied in wireless sensor networks are shown in Fig.3 [9, 32]. Star topologies and mesh topologies are suitable for communication in cyber-physical systems, therefore frequently applied in SHM and control applications. In star topologies, sensor nodes are connected to a central node, e.g. to a base station. Communication is exclusively realized between sensor nodes and the central node. As a consequence, failures of a central node result in failure of the total sensor network. However, star topologies are tolerant to failure of single sensor nodes, rendering star topologies suitable for SHM and control applications. Mesh topologies enable communication across all nodes in a network. Partially connected and fully connected mesh networks can be distinguished. In fully connected mesh networks, nodes directly communicate with other nodes, whereas in partially connected mesh networks several nodes are connected indirectly. Mesh topologies allow calculating optimal routing paths, which contributes to an efficient exploitation of energy resources distributed over a wireless sensor network.

As data exchanged in sensor networks encompasses several layers of abstraction, a layered model of networking has been developed by the International Organization for Standardization (ISO), published in ISO/IEC 7498-1:1994 [33]. The “Open Systems Interconnection (OSI) reference model” defines seven layers of abstraction in network communication, shown in Fig.4, and the function of each layer. Following the OSI reference model, nodes of networks communicate at equivalent layers of abstraction by layer-specific protocols. Every layer of abstraction  $n-1$  provides services to higher abstraction layers  $n$  through service access points specified for each layer [34].

In the OSI reference model, layers 1 to 4 form the transport system of a network, while layers 5 to 7 represent application-oriented layers. On the physical layer (layer 1), bit transmission over a physical medium, such as air or water, is specified. Protocols on the

physical layer encode parameter modulation schema, transmit power, and hop distance of data packets, which contribute to energy consumption of wireless sensor networks. The modulation schema specifies the transformation of the bit values 0 and 1 into electrical quantities. The term “transmit power” denotes the amount of energy required to transmit data packets. Transmit power is in inverse proportion to errors in data transmission and related to the distances between the transmitter and the receiver of communicating nodes, which is called “hop distance” [32, 34]. The data link layer (layer 2) organizes bits of data units into frames, referred to as multiplexing and demultiplexing, respectively. The data link layer is responsible for error detection through check sums added to a data packet, and it provides flow control to manage data transmission rates. Furthermore, medium access control (MAC) is performed on the data link layer. The network layer (layer 3) manages routing, i.e. identifying optimal paths to forward data from sensor nodes to base stations by passing intermediate nodes [32,34]. The fourth layer, or transport layer, establishes communication between nodes defined by the routing process of the network layer. Transport layer protocols transform data units (multiplexing or demultiplexing) to make data accessible to end-system processes. Moreover, transport layer protocols provide services, such as error detection in data units, flow control and service reliability by the authorization of retransmission of a data unit [35]. The session layer (layer 5) enables and controls communication processes for complete data exchange between nodes by managing multiple transport layer connections. Following the session layer, the presentation layer (layer 6) performs data transformations to make data processable by the application layer (layer 7) [32]. The data format is defined by a uniform syntax, referred to as transform syntax, enabling correct representations of data with respect to encoding mechanisms, such as ASCII code or Unicode. Finally, the application layer provides protocols to be used by software applications [34].

On each layer of the OSI reference model, several protocols exist that can be combined to a specific protocol stack. In a protocol stack, data is transmitted from a node to another node by passing every layer of abstraction (Fig. 4). Starting from the application layer of a sending node, control information is successively added to a data unit and removed in inverse direction at a receiving node. Regarding the energy constraints of wireless sensor networks, which are typically composed of battery-powered components, the deployment of cross-layer protocols improves communication efficiency within a network. Cross-layer protocols

integrate several functionalities of different OSI communication layers into one protocol to enable highly reliable communication with minimal energy consumption, adaptive communication decisions, and local flow control [36].

Due to the plenitude of communication protocols, suitable communication protocols for cyber-physical systems are defined when designing a CPS for structural health monitoring and control, following a requirements analysis applying specific criteria, such as the size of data packets to be transmitted, the data transfer frequency, and the transfer range between nodes. Furthermore, limits on the number of communicating nodes and security issues must be considered in selecting communication protocols. Security is of growing importance in cyber-physical systems specially in those applied in Industrial Internet of Things (IIoT) applications [37]. Security objectives in communication between nodes of a network are (i) confidentiality and integrity of data, (ii) authenticity of system elements, and (iii) data access authorization [38]. Data confidentiality denotes restrictions of readability of data to authorized network components, while data integrity describes the verification of sources sending data as well as the detection of data modified and sent by unauthorized sources. Therefore, system elements require unique identifiers (authenticity) and access authorizations to read, to write, and to manipulate data. As security issues are of growing importance in CPS, IoT, and IIoT (however not within the main scope of this study), information describing security-related network properties is formalized along with communication-related information and monitoring-related information.

Table 1 provides an example selection of communication protocols frequently applied for wireless communication in cyber-physical systems, including Bluetooth, ZigBee, Wi-Fi, and the Message Queue Telemetry Transport protocol (MQTT). ZigBee and MQTT are characterized by low power consumption and provide embedded security strategies with a high degree of scalability with respect to the number of nodes. High-level protocols, such as ZigBee and MQTT, are based on implementations of the physical layer and the data link layer of the OSI model (Fig. 4) [36,39]. A number of low-level networking protocols, encompassing functionalities of the physical layer and the data link layer, are standardized in the IEEE 802 standards [40], such as Wi-Fi (IEEE 802.11.a/b/g) and Bluetooth (IEEE 802.15.1). A low-power and low-cost solution for wireless communication is given by the low-rate wireless personal area network (LR-WPAN) standard IEEE 802.15.4 that allows



setting up networks with star and mesh topologies. An overview of frequency ranges and data rates in compliance with local regulations is given in Table 2 [41]. Two physical layers, i.e. the 2.4 GHz band layer and the 868/915 MHz band layer, are defined in the IEEE 802.15.4 standard. The worldwide unlicensed 2.4GHz band layer is characterized by higher data rates compared to the 868/915 MHz band layer, which is due to higher frequencies and the digital modulation schema applied. While for binary phase-shift keying (BPSK) two phases separated by  $180^\circ$  are used for digital modulation, offset quadrature phaseshift keying (O-QPSK) uses four phases to modulate data for high data rates. The offset in the modulation schema is applied for limiting large amplitude fluctuations undesired in communication systems.

## 二、中文译文

一种信息物理系统的元模型

摘要

随着物联网和工业 4.0 概念的出现，土木工程中的信息物理系统在结构健康监测（SHM）和控制应用中的影响越来越大。在正式基准上设计、优化和记录信息物理系统需要平台独立和技术独立的元模型。此研究以信息物理系统中的通信为重点，提出了一个描述信息物理系统的元模型。首先，回顾了土木工程计算中常用的元建模概念，讨论了描述通信相关信息的可能性和局限性。其次，说明了应用于结构健康监测（SHM）和控制的分布式信息物理系统的通信相关特性和行为，并详细说明了与通信相关的系统组件。然后，提出了正式描述信息物理系统的元模型并将其映射到一个开放的国际建筑信息模型（BIM）工业基础类（IFC）中。最后，使用官方 IFC 认证程序的软件对基于 IFC 的方法进行了验证，并通过一个原型信息物理系统的基于 BIM 的实例建模进行了确认，该模型已在实验室中物理实现。综上，本文描述了应用于结构健康监测（SHM）和控制的信息物理系统而且这些信息已经在工业基础类（IFC）的正式基准上进行了存储、记录和交换，促进了信息物理系统的设计、优化和记录。

### 3. 信息物理系统中的通信

为了使用元建模方法描述信息物理系统，需要定义系统组成部分，其中包括属性和方法。作为描述属性和方法的基础，这一节研究了信息物理系统中用于通信的网络拓扑结构。此外，（本节）提供了一个对于网络通信的全面评述，并对适合信息物理系统的通信协议做出了评价。开发捕获通信相关信息的元模型，例如与通信相关的系统组件、

网络拓扑、网络通信特性和通信协议，使其获得明显的优势，比如有效地规划结构健康监测（SHM）和控制系统、改进信息物理系统（CPS）运行和维护阶段的数据控制和管理，或将结构健康监测（SHM）和控制系统部件与通信相关信息连接起来。

由于无线传感器网络的节点是空间分布的，自主设备、功率消耗和资源管理是设计通信协议和网络拓扑的重要标准[31]。通常应用于无线传感器网络的拓扑结构如图 3 所示[9, 32]。星型拓扑结构和网状拓扑结构适用于信息物理系统中的通信，因此频繁地被应用于结构健康监测（SHM）和控制应用中。在星型拓扑结构中，传感器节点被连接到中心节点，例如基站。传感器节点和中心节点之间的通信是专用的。因此，中心节点的故障会导致整个传感器网络的故障。但是，星形拓扑结构可以容忍单个传感器节点的故障，因此星形拓扑结构适合结构健康监测（SHM）和控制应用程序。网状拓扑结构支持网络中所有节点之间的通信。部分连接和完全连接的网状网络是可被区分的。在完全连接的网状网络中，节点直接与其他节点通信，而在部分连接的网状网络中，几个节点是间接连接的。网状拓扑结构允许计算最佳路由路径，这有助于有效利用分布在无线传感器网络上的能量资源。

由于交换数据的传感器网络中包含多个抽象层，国际标准化组织（ISO）已经开发了一个网络分层模型，该模型发表在 ISO/IEC 7498-1:1994[33]中。“开放系统互连(OSI)参考模型”定义了网络通信中的七个抽象层，如图 4 所示，以及每一层的功能。遵循开放系统互连（OSI）参考模型，网络节点通过特定层的协议在等价的抽象层进行通信。每个抽象层  $n-1$  通过每一层明确规定的服务访问点向更高的抽象层  $n$  提供服务[34]。

在开放系统互连（OSI）参考模型中，1 到 4 层构成网络的传输系统，而 5 到 7 层代表面向应用的层。在物理层（1 层）上，比特流在指定的物理介质上传输，比如空气和水。物理层上的协议对参数调制方案，传输功率和数据包的跳距进行编码，这会增加无线传感器网络的能量消耗。这种调制模式将 0 和 1 的位值变化明确规定为电量。术语“传输功率”表示传输数据包所需的能量总量。传输功率与数据传输中的错误成反比并且与被称为“跳距”[32, 34]的通信节点发射机和接收机之间的距离有关。数据链路层（2 层）将数据单元的位组织成帧，分别称为多路复用和多路分解。数据链路层通过检查被添加到数据包中的数据和来负责错误检测，并且它提供流控制来管理数据传输速率。此外，数据链路层上可执行介质访问控制（MAC）。网络层（3 层）管理路由，即通过传递中间节点来确定从传感器节点向基站转发数据的最佳路径[32, 34]。第四层，或称为传输层，在通过网络层的路由过程定义的节点之间建立通信。传输层协议通过转换数据单元（多路复用或多路分解）使数据可供最终系统进程访问。此外，传输层协议通过授权

一个数据单元的重传来提供服务，例如数据单元中的错误检测、流控制和服务可靠性[35]。会话层（5层）通过管理很多个传输层连接来启用和控制节点之间完整数据交换的通信过程。在会话层之后，表示层（6层）执行数据转换使得数据可由应用层（7层）处理[32]。数据格式由统一语法定义，被称为转换语法，允许对使用如 ASCII 代码或 Unicode 编码等编码机制的数据进行正确表示。最后，应用层提供了被软件应用程序使用的协议[34]。

在开放系统互连（OSI）参考模型的每一层上，都存在一些可以组合成特定协议栈的协议。在协议栈中，由通过每个抽象层来从一个节点向另一个节点传输数据（图4）。从发送节点的应用层开始，控制信息被连续地添加到数据单元并在接收节点处以相反方向进行删除。考虑到无线传感器网络的能量限制，它们通常由电池供电的组件组成，跨层协议的部署提高了网络内的通信效率。跨层协议将不同开放系统互连（OSI）通信层的多个功能合并到一个协议中从而以极少的能耗、自适应通信决策和本地流控制实现高可靠性的通信[36]。

由于有大量的通信协议，在设计用于结构健康监测和控制的信息物理系统（CPS）时，在应用特定标准，如要传输的数据包大小、数据传输频率，进行需求分析之后，定义了适用于信息物理系统的合适的通信协议，以及节点之间的传输范围。此外，在选择通信协议时必须考虑通信节点的数量限制和安全问题。安全在信息物理系统中越来越重要，特别是在工业物联网（IIoT）应用中[37]。网络节点间通信的安全目标是（i）数据的机密性和完整性，（ii）系统元素的真实性，以及（iii）数据访问授权[38]。数据机密性表示对授权网络组件的数据可读性的限制，而数据完整性描述的是对源发送数据的检查以及对未授权的源修改和发送的数据的检测。因此，系统元素需要唯一的标识符（真实性）和访问授权来读取、写入和操作数据。由于安全问题在信息物理系统（CPS）、物联网（IoT）和工业物联网（IIoT）中越来越重要（但不在本研究的主要范围内），描述安全相关网络属性的信息与通信相关信息和监控相关信息一起被正式化。

表1提供了一个在信息物理系统中频繁应用于无线通信的通信协议的示例选择，包括蓝牙(Bluetooth)，紫蜂(ZigBee)，无线上网(Wi-Fi)和消息队列遥测传输协议(MQTT)。紫蜂(ZigBee)和消息队列遥测传输协议(MQTT)具有低功耗的特点，并且在节点数量方面提供了高度可扩展的嵌入式安全策略。高级协议，例如紫蜂(ZigBee)和消息队列遥测传输协议(MQTT)，是基于开放系统互连(OSI)模型的物理层和数据链路层的实现(图4)[36, 39]。一些低级网络协议，包含了物理层和数据链路层的功能已在IEEE 802标准[40]中标准化，例如无线上网(Wi-Fi)(IEEE 802.11.a/b/g)和蓝牙(Bluetooth)

(IEEE 802.15.1)。低速率无线个人局域网 (LR-WPAN) 标准 IEEE 802.15.4 给出了一种低功耗、低成本的无线通信解决方案, 该标准允许建立具有星型和网状拓扑结构的网络。表 2[41]概述了符合当地法规的频率范围和数据速率。IEEE 802.15.4 标准中定义了两个物理层, 即 2.4ghz 频带层和 868/915mhz 频带层。在世界范围内未经许可的 2.4GHz 频带层的特征是其具有比 868/915mhz 频带层更高的数据速率, 这是由于其较高的频率和应用的数字调制模式。对于二进制相移键控 (BPSK), 数字调制使用相隔  $180^\circ$  的两个相位, 而偏移正交相移键控 (O-QPSK) 则使用四个相位来调制高数据速率的数据。调制模式中的偏移量用于限制通信系统中不希望出现的大幅度波动。

指导教师 (签字):

2020 年 3 月 13 日