

边缘计算:万物互联时代新型计算模型

施巍松¹ 孙辉² 曹杰¹ 张权¹ 刘伟²

¹(韦恩州立大学计算机科学系 美国底特律 48202)

²(安徽大学计算机科学与技术学院 合肥 230601)

(weisong@wayne.edu)

Edge Computing—An Emerging Computing Model for the Internet of Everything Era

Shi Weisong¹, Sun Hui², Cao Jie¹, Zhang Quan¹, and Liu Wei²

¹(Department of Computer Science, Wayne State University, Detroit, USA 48202)

²(School of Computer Science and Technology, Anhui University, Hefei 230601)

Abstract With the proliferation of Internet of things (IoT) and the burgeoning of 4G/5G network, we have seen the dawning of the IoE (Internet of everything) era, where there will be a huge volume of data generated by things that are immersed in our daily life, and hundreds of applications will be deployed at the edge to consume these data. Cloud computing as the de facto centralized big data processing platform is not efficient enough to support these applications emerging in IoE era, i. e., 1) the computing capacity available in the centralized cloud cannot keep up with the explosive growing computational needs of massive data generated at the edge of the network; 2) longer user-perceived latency caused by the data movement between the edge and the cloud; 3) privacy and security concerns from data owners in the edge; 4) energy constraints of edge devices. These issues in the centralized big data processing era have pushed the horizon of a new computing paradigm, edge computing, which calls for processing the data at the edge of the network. Leveraging the power of cloud computing, edge computing has the potential to address the limitation of computing capability, the concerns of response time requirement, bandwidth cost saving, data safety and privacy, as well as battery life constraint. “Edge” in edge computing is defined as any computing and network resources along the path between data sources and cloud data centers. In this paper, we introduce the definition of edge computing, followed by several case studies, ranging from cloud offloading to smart home and city, as well as collaborative edge to materialize the concept of edge computing. Finally, we present several challenges and opportunities in the field of edge computing, and hope this paper will gain attention from the community and inspire more research in this direction.

Key words edge computing; cloud computing; Internet of everything (IoE); smart home and city; collaborative edge

摘要 随着物联网的快速发展和 4G/5G 无线网络的普及,万物互联的时代已经到来,网络边缘设备数量的迅速增加,使得该类设备所产生的数据已达到泽字节(ZB)级别.以云计算模型为核心的集中式大

收稿日期:2016-12-09;修回日期:2017-02-09

基金项目:国家自然科学基金面上项目(61572001);安徽大学 2016 年博士科研启动经费项目(J01003214)

This work was supported by the General Program of the National Natural Science Foundation of China (61572001) and Anhui University Fund for Doctoral Research (J01003214).

通信作者:孙辉(sunhui@ahu.edu.cn)

数据处理时代,其关键技术已经不能高效处理边缘设备所产生的数据,主要表现在:1)线性增长的集中式云计算能力无法匹配爆炸式增长的海量边缘数据;2)从网络边缘设备传输海量数据到云中心致使网络传输带宽的负载量急剧增加,造成较长的网络延迟;3)网络边缘数据涉及个人隐私,使得隐私安全问题变得尤为突出;4)有限电能的网络边缘设备传输数据到云中心消耗较大电能.为此,以边缘计算模型为核心的面向网络边缘设备所产生海量数据计算的边缘式大数据处理应运而生,其与现有以云计算模型为核心的集中式大数据处理相结合,即二者相辅相成,应用于云中心和网络边缘端的大数据处理,较好地解决了万物互联时代大数据处理中所存在的上述问题.边缘计算中的“边缘”是个相对的概念,指从数据源到云计算中心数据路径之间的任意计算资源和网络资源.边缘计算的基本理念是将计算任务在接近数据源的计算资源上运行.首先系统地介绍边缘计算的概念和原理;其次,通过现有研究作为案例(即云计算任务迁移、视频分析、智能家居、智慧城市、智能交通以及协同边缘),实例化边缘计算的概念;最后,提出边缘计算领域所存在的挑战.该文希望能让学界和产业界了解和关注边缘计算,并能够启发更多的学者开展边缘式大数据处理时代边缘计算模型的研究.

关键词 边缘计算;云计算;万物互联;智能家居和城市;协同边缘

中图法分类号 TP391; TP393

自2005年,云计算^[1]的提出和广泛应用已经改变了人们日常工作和生活的方式,如软件即服务(software as a service, SaaS)被广泛应用到谷歌、Twitter、Facebook等著名IT企业的数据中心.可扩展的基础设施和支持云服务的处理引擎技术已对应用服务程序的运行方式产生了巨大影响,如谷歌的文件系统Google File System^[2]、MapReduce编程模型^[3]、Apache基金会开发的分布式文件系统Hadoop^[4]、加州大学伯克利分校AMP实验室开发的内存计算框架Spark^[5]等.

物联网技术(Internet of things, IoT)^[6]旨在利用射频识别技术、无线数据通信技术构造一种全球物品信息实时共享的实物互联网.随后,“无人参与的计算机信息感知”的概念开始逐渐应用到可穿戴医疗、智能家居、环境感知和智能运输系统中^[7-8].如今,我们已经从物联网时代迈进万物互联(Internet of everything, IoE)^[9]的时代,相比物联网而言,万物互联除了“物”与“物”的互联,还增加了更高级别的“人”与“物”的互联,其突出特点是任何“物”都将具有语境感知的功能、更强的计算能力和感知能力.将人和信息融入到互联网中,网络将具有数十亿甚至数万亿的连接节点.万物互联以物理网络为基础,增加了网络智能,在互联网的“万物”之间实现融合、协同以及可视化的功能.

随着万物互联的飞速发展及广泛应用,边缘设备正在从以数据消费者为主的单一角色转变为兼顾数据生产者和数据消费者的双重角色,同时网络边缘设备逐渐具有利用收集的实时数据进行模式识

别、执行预测分析或优化、智能处理等功能.目前,大数据处理已经从以云计算为中心的集中式处理时代(本文中我们把2005—2015这10年称之为集中式大数据处理时代)正在跨入以万物互联为核心的边缘计算时代(本文中我们称之为边缘式大数据处理时代).集中式大数据处理时代,更多的是集中式存储和处理大数据,其采取的方式是建造云计算中心,并利用云计算中心超强的计算能力来集中式解决计算和存储问题.相比而言,在边缘式大数据处理时代,网络边缘设备会产生海量实时数据;并且,这些边缘设备将部署支持实时数据处理的边缘计算平台为用户提供大量服务或功能接口,用户可通过调用这些接口来获取所需的边缘计算服务.

根据思科全球云指数的预估^[10],到2019年,物联网产生数据的45%将在网络边缘存储、处理、分析,而全球数据中心总数据流量预计将达到10.4泽字节(zettabyte, ZB).据思科互联网业务解决方案集团预测^[11],到2020年,连接到网络的无线设备数量将达到500亿台.基于万物互联平台的应用服务需要更短的响应时间,同时也会产生大量涉及个人隐私的数据.在此情况下,传统云计算模式将不能高效地支持基于万物互联的应用服务程序,而边缘式大数据处理时代下的边缘计算模型则可较好地解决这些问题.在边缘计算模型中,网络边缘设备已经具有足够的计算能力来实现源数据的本地处理,并将结果发送给云计算中心.边缘计算模型不仅可降低数据传输带宽,同时能较好地保护隐私数据,降低终端敏感数据隐私泄露的风险.因此,随着万物互联的

发展,边缘计算模型将成为新兴万物互联应用的支撑平台。

万物互联技术发展,将使云计算中心的部分应用服务程序迁移到网络边缘设备。边缘设备兼顾数据消费者和生产者。本文从概念、原理以及挑战等方面对边缘式大数据处理时代的支撑平台,即边缘计算模型展开介绍。首先阐述边缘计算的背景;其次介绍边缘计算的概念;为进一步解释边缘计算的概念,结合我们研究成果,提出未来边缘计算的6种典型案例,包括云计算任务迁移、视频分析、智能家居、智慧城市、智能交通以及协同边缘;最后,讨论边缘计算所面临的挑战,如编程模型、命名规则、数据抽象、服务管理、数据隐私保护及安全、理论基础及商业模型。通过这些阐述,旨在为边缘计算研究者提供未来的研究方向。

1 边缘计算的兴起

我们将以云计算模型为核心的大数据处理阶段称为集中式大数据处理时代,该阶段特征主要表现为大数据的计算和存储均在云计算中心(数据中心)采用集中方式执行,因为云计算中心具有较强的计算和存储能力。这种资源集中的大数据处理方式可以为用户节省大量开销,创造出有效的规模经济效益。但是,云计算中心的集中式处理模式在万物互联的时代表现出其固有的问题,如万物互联背景下,网络边缘设备所产生的数据已达到海量级别:1)线性增长的集中式云计算能力无法匹配爆炸式增长的海量边缘数据;2)从网络边缘设备传输到云数据中心的海量数据增加了传输带宽的负载量,造成网络延迟时间较长;3)边缘设备数据涉及个人隐私和安全的问题变得尤为突出;4)边缘设备具有有限电能,数据传输造成终端设备电能消耗较大等。

针对于此,万物互联时代如果仍采用集中式大数据处理模式下的云计算,现有的云计算相关技术并不能完全高效地处理网络边缘设备所产生的海量数据。因此,学术界和产业界开始对新的计算模型进行深入地研究,如微数据中心^[12]、移动边缘计算^[13]、雾计算^[14]、Cloudlet^[15]、中国科学院的海云计算^[16]等。万物互联应用需求的发展催生了边缘式大数据处理模式,即边缘计算模型^[17],其能在网络边缘设备上增加执行任务计算和数据分析的处理能力,将原有云计算模型的部分或全部计算任务迁移到网络边缘设备上,降低云计算中心的计算负载,减缓网络

带宽的压力,提高万物互联时代数据的处理效率。本节中,我们首先介绍边缘计算产生的动机,然后给出边缘计算的定义及理解。

1.1 边缘计算产生的动机

1.1.1 云计算服务的不足

云计算大多采用集中式管理的方法,这使云服务创造出较高的经济效益,而在万物互联的背景下,应用服务需要低延时、高可靠性以及数据安全,而传统云计算无法满足这些需求,主要归因于3个方面:

1) 实时性。万物互联环境下,边缘设备产生大量实时数据,云计算性能^[18]正逐渐达到瓶颈。据IDC预测^[19],到2020年,全球数据总量将大于40 ZB。随着边缘设备数据量的增加,网络带宽正逐渐成为云计算的另一瓶颈。仅提高网络带宽并不能满足新兴万物互联应用对延迟时间的要求,例如,波音787每秒产生的数据超过5 GB^[20],但飞机与卫星之间的带宽不足以支持实时传输。装载在无人驾驶汽车^[21]上的传感器和摄像头实时捕捉路况信息,每秒产生约1 GB数据。根据IHS预测^[22],到2035年,全球将有5 400万辆无人驾驶汽车,如何实现较短延时将是未来主要研究方向。为此,在接近数据源的边缘设备上执行部分或全部计算是适应万物互联应用需求的新兴计算模式。

2) 隐私保护。当用户使用电子购物网站、搜索引擎、社交网络等时,用户的隐私数据将被上传至云中心^[23]。Cortes等人研究出一种基于运动追踪的医疗体育应用服务^[24],其包含用户隐私数据,如从路由起点信息可以查找到用户的家庭地址。随着智能家居的普及,许多家庭在屋内安装网络摄像头,如果直接将视频数据上传至云数据中心,视频数据的传输不仅会占用带宽资源,还增加了泄露用户隐私数据的风险。为此,针对现有云计算模型的数据安全问题,边缘计算模型为这类敏感数据提供了较好的隐私保护机制,一方面,用户的源数据在上传至云数据中心之前,首先利用近数据端的边缘结点直接对数据源进行处理,以实现对一些敏感数据的保护与隔离;另一方面,边缘节点与云数据之间建立功能接口,即边缘节点仅接收来自云计算中心的请求,并将处理的结果反馈给云计算中心。这种方法可以显著地降低隐私泄露的风险。

3) 能耗。针对云数据中心的能耗问题,许多研究者进行了深入的调查研究^[25-26]。Sverdlik的研究结果表明^[27],到2020年美国所有数据中心的总能耗将增长4%,在2020年将达到约730亿千瓦时。

在我国,环境 360 报告表明^[28],仅我国数据中心所消耗的电能已经超过匈牙利和希腊两国用电的总和.随着在云计算中心运行的用户应用程序越来越多,未来大规模数据中心对能耗的需求将难以满足^[29].在云计算中心的能耗优化方面,现有的研究内容主要集中在如何提高能源使用效率^[30-31]和动态资源管理策略方面^[32-33],以达到减缓能耗增速、最大程度的节能.然而,仅提高能效水平等策略,虽然可达到节能的目的,但仍不能解决数据中心巨大能耗的问题,这一问题在万物互联环境下将更加突出.为解决这一能耗难题,边缘计算模型提出将原有云数据中心上运行的一些计算任务进行分解,然后将分解的计算任务迁移到边缘节点进行处理,以此降低云计算数据中心的计算负载,进而达到降低能耗的目的.

当前,线性增长的集中式云计算能力已无法匹配爆炸式增长的海量边缘数据,基于云计算模型的单一计算资源已不能满足大数据处理的实时性、安全性和低能耗等需求,在现有以云计算模型为核心的集中式大数据处理基础上,亟待需要以边缘计算模型为核心,面向海量边缘数据的边缘式大数据处理技术,二者相辅相成,应用于云中心和边缘端大数据处理,解决万物互联时代云计算服务不足的问题.

1.1.2 万物互联的兴起

传感器、智能手机、可穿戴设备以及智能家电等设备将成为万物互联的一部分,并产生海量数据,而现有云计算的带宽和计算资源还不能高效处理这些数据.因此,在网络边缘端处理源数据,筛选出有效的信息并发送到云端将成为一种新的计算模型,其有效降低云中心的网络带宽和计算负载,边缘计算模型受到学术界和产业界的广泛关注.

图 1 所示为传统云计算模型.源数据由生产者发送至云端,终端用户、智能手机、个人电脑等数据消费者向云中心发送使用请求.图 1 中,蓝色实线表示数据生产者发送源数据到云中心,红色实线表示数据消费者向云中心发送使用请求,红色虚线表示云中心将结果反馈给数据消费者.云计算利用大量计算资源来处理数据,但万物互联环境下,传统云计

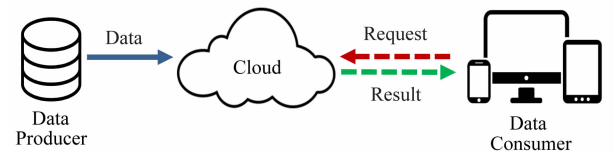


Fig. 1 Cloud computing paradigm
图 1 传统云计算模型

算模型不能有效满足万物互联应用的需求,其原因主要有:①直接将边缘设备端海量数据发送到云端,造成网络带宽负载和计算资源浪费;②传统云计算模型的隐私保护问题将成为万物互联架构中云计算模型的障碍;③万物互联架构中大多数边缘设备节点的能源是有限的,并且 GSM, WiFi 等无线传输模块的能耗较大.

针对于此,利用边缘设备已具有的计算能力,将应用服务程序的全部或部分计算任务从云中心迁移到边缘设备端执行,降低能源消耗.

1.1.3 从数据消费者到生产者

在云计算模型中,边缘终端设备通常作为数据消费者(如用智能手机观看在线视频),如今智能手机也可生产数据,从数据消费者到生产者角色的转变要求边缘设备具有更强的计算能力,如人们通过 Facebook、Twitter、微信等分享照片及视频. YouTube 用户每分钟上传长达 72 h 的视频内容; Twitter 用户每分钟近 30 万次的访问量; Instagram 用户每分钟上传近 22 万张新照片^[34]; 微信朋友圈和腾讯 QQ 空间每天上传的图片高达 10 亿张; 腾讯视频每天播放量达 20 亿次. 这些图片和视频数据量较大,上传至云计算中心过程会占用大量带宽资源. 为此,在源数据上传至云中心之前,可在边缘设备执行预处理,以减少传输的数据量,降低传输带宽的负载. 此外,若在边缘设备处理个人身体健康数据等隐私数据,用户隐私会得到更好地保护.

1.2 边缘计算定义

边缘计算是指在网络边缘执行计算的一种新型计算模型,边缘计算中边缘的下行数据表示云服务,上行数据表示万物互联服务,而边缘计算的边缘是

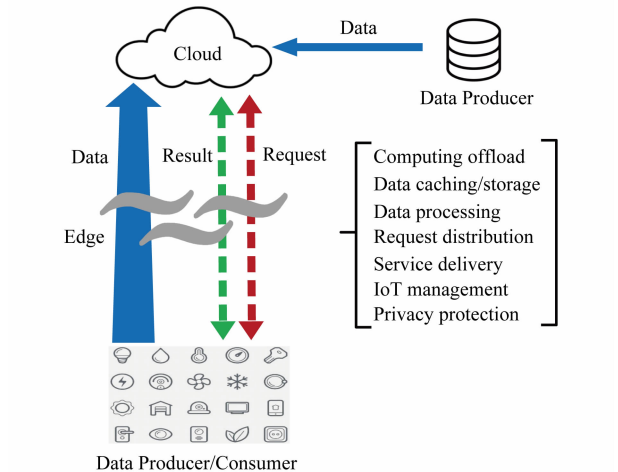


Fig. 2 Edge computing paradigm
图 2 边缘计算模型

指从数据源到云计算中心路径之间的任意计算和网络资源.图2表示基于双向计算流的边缘计算模型.云计算中心不仅从数据库收集数据,也从传感器和智能手机等边缘设备收集数据.这些设备兼顾数据生产者和消费者.因此,终端设备和云中心之间的请求传输是双向的.网络边缘设备不仅从云中心请求内容及服务,而且还可以执行部分计算任务,包括数据存储、处理、缓存、设备管理、隐私保护等.因此,需要更好地设计边缘设备硬件平台及其软件关键技术,以满足边缘计算模型中可靠性、数据安全性的需求.

1.3 边缘计算优势

边缘计算模型将原有云计算中心的部分或全部计算任务迁移到数据源的附近执行.根据大数据的3V特点,即数据量(volume)、时效性(velocity)、多样性(variety),通过对比云计算模型为代表的集中式大数据处理(如图3所示)和以边缘计算模型为代表的边缘式大数据处理(如图4所示)时代不同数据特征来阐述边缘计算模型的优势.

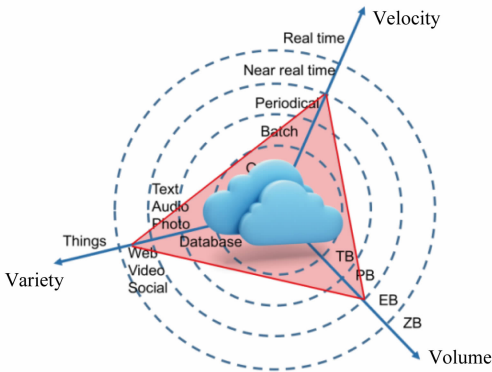


Fig. 3 Centralized big data processing platform v1.0

图3 集中式大数据处理

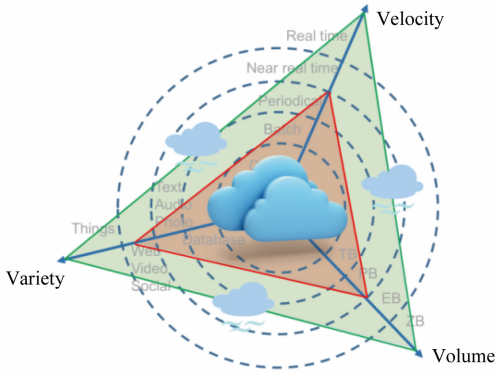


Fig. 4 Edge computing-based data processing platform v1.0

图4 边缘式大数据处理

集中式大数据处理时代,数据的类型主要以文本、音视频、图片以及结构化数据库等为主,数据量

维持在PB级别,云计算模型下的数据处理对实时性要求不高.万物互联背景下的边缘式大数据处理时代,数据类型变得更加负责多样,其中万物互联设备的感知数据急剧增加,原有作为数据消费者的用户终端已变成了具有可产生数据的生产者终端,并且边缘式大数据处理时代,数据处理的实时性要求较高,此外,该时期的数据量已超过ZB级.针对此,边缘式大数据处理时代,由于数据量的增加以及对实时性的需求,需将原有云中心的计算任务部分迁移到网络边缘设备(如图4的边缘云)上,以提高数据传输性能,保证处理的实时性,同时降低云计算中心的计算负载.

为此,边缘式大数据处理时代的数据特征催生了边缘计算模型.然而,边缘计算模型与云计算模型并不是非此即彼的关系,而是相辅相成的关系,边缘式大数据处理时代是边缘计算模型与云计算模型的相结合的时代,二者的有机结合将为万物互联时代的信息处理提供较为完美的软硬件支撑平台.

2 边缘计算典型应用实例

应用是检验新技术是否有价值的最直接、最有效的方式.同样,也适用于边缘计算模型,边缘计算是否有价值取决于基于边缘计算的关键应用场景,只有通过应用才能发现边缘计算的发展中所遇到的各种挑战和机遇.因此,下面我们给出基于边缘计算模型的6种实际应用案例.通过这些案例可以展望边缘计算在万物互联背景下的研究机遇和应用前景.

2.1 云计算任务迁移

云计算中,大多数计算任务在云计算中心执行,这会导致响应延时较长,损害用户体验.根据用户设备的环境可确定数据分配和传输方法,EAWP(edge accelerated Web platform)模型^[35]改善了传统云计算模式下较长响应时间的问题.许多研究团队^[36-40]已经开始研究解决云迁移在移动云环境中的能耗问题.边缘计算中,边缘端设备借助其一定的计算资源实现从云中心迁移部分或全部任务到边缘端执行.

在线购物应用中,消费者可能频繁地操作购物车,默认条件下,用户购物车状态的改变先在云中心完成,用户设备上购物车内产品视图再更新.这个操作时间取决于网络带宽和云中心负载状况.由于移动网络的低带宽,移动端购物车的更新延时较长.目前,使用移动客户端网购变得流行,因此缩短响应延时,

改善用户体验的需求日益增加. 如果购物车内产品视图的更新操作从云中心迁移到边缘节点, 这样会降低用户请求的响应延时. 购物车数据可被缓存在边缘节点, 相关的操作可在边缘节点上执行. 当用户的请求到达边缘节点时, 新的购物车视图立即推送到用户设备. 边缘节点与云中心的数据同步可在后台进行.

2.2 边缘计算视频监控

城市安全视频监控系统主要应对因万物互联的广泛应用而引起的新型犯罪及社会管理等公共安全问题. 传统视频监控系统前端摄像头内置计算能力较低, 而现有智能视频监控系统的智能处理能力不足. 为此, 我们以云计算和万物互联技术为基础, 融合边缘计算模型和视频监控技术, 构建基于边缘计算的新型视频监控应用的软硬件服务平台, 以提高视频监控系统前端摄像头的智能处理能力, 进而实现重大刑事案件和恐怖袭击活动预警系统和处置机制, 提高视频监控系统的防范刑事犯罪和恐怖袭击的能力.

针对海量视频数据, 云计算中心服务器计算能

力有限等问题: 1) 我们构建了一种基于边缘计算的视频图像预处理技术 (preprocessing). 通过对视频图像进行预处理, 去除视频图像冗余信息, 使得部分或全部视频分析迁移到边缘处, 由此降低对云中心的计算、存储和网络带宽需求, 提高视频分析的速度, 此外, 预处理使用的算法采用软件优化、硬件加速等方法, 提高视频图像分析的效率. 2) 为了降低上传的视频数据, 基于边缘预处理功能, 我们构建基于行为感知的视频监控数据弹性存储机制 (elastic storage). 边缘计算软硬件框架为视频监控系统提供具有预处理功能的平台, 实时提取和分析视频中的行为特征, 实现监控场景行为感知的数据处理机制; 根据行为特征决策功能, 实时调整视频数据, 既减少无效视频的存储, 降低存储空间, 又最大化存储“事中”证据类视频数据, 增强证据信息的可信性, 提高视频数据存储空间利用率.

图 5 表示基于边缘计算的视频监控系统框图, 其中具有边缘计算功能的模块作为协处理单元, 简称边缘计算硬件单元 (hardware unit) 与原有视频监控系统的摄像头终端系统进行系统融合.

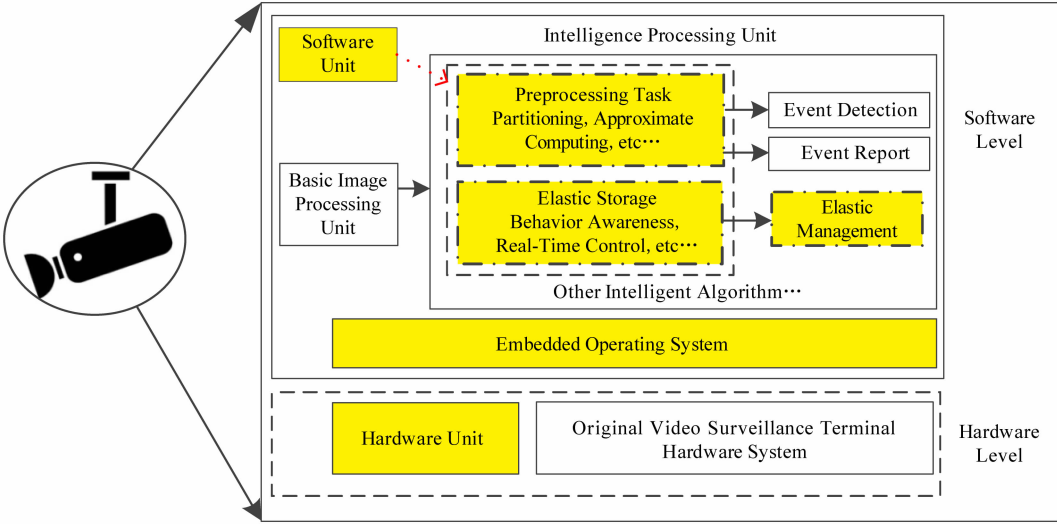


Fig. 5 Edge computing-based video surveillance system
图 5 边缘计算视频监控系统框图

现有的视频监控系统在记录视频数据之后, 采用直接或简单视频处理后传输到云计算中心. 而随着视频数据呈现海量的特征, 公共安全领域的应用要求视频监控系统能够提供实时、高效的视频数据处理. 针对此, 如图 4 所示利用边缘计算模型将具有计算能力的硬件单元集成到原有的视频监控系统硬件平台上, 配以相应的软件支撑技术, 实现具有边缘计算能力的新型视频监控系统. 在边缘计算模型中,

计算通常发生在数据源的附近^[41], 即在视频数据采集的边缘端进行视频数据的处理.

为此, 一方面, 基于智能算法的预处理功能模块, 在保证数据可靠性的前提下, 利用模糊计算模型, 对实时采集的视频数据执行部分或全部计算任务, 这能够为实时性要求较高的应用请求提供及时的应答服务, 而且还降低云计算中心计算和带宽的负载; 另一方面, 我们还设计了具有可伸缩的弹性存

储功能模块,利用智能算法感知监控场景内行为的变化来选择性存储视频数据,实现最小空间存储最大价值的数据(如犯罪行为证据等).最后,在兼容现有智能处理的功能基础上,增加了“事中”事件监测和“事中”事件报告的功能,及时有效地向用户发送响应信息.

2.3 智能家居

家居生活随着万物互联应用的普及变得越来越智能和便利,如智能照明控制系统、智能电视、智能机器人等.然而,在智能设备中,仅通过一种 WiFi 模块连接到云计算中心的做法,远远不能满足智能家居的需求.智能家居环境中,除了联网设备外,廉价的无线传感器和控制器应部署到房间、管道、地板和墙壁等,出于数据传输负载和数据隐私的考虑,这些敏感数据的处理应在家庭范围内完成.

传统的云计算模型已不能完全适用于智能家居类应用,而边缘计算模型是组建智能家居系统的最优平台.在家庭内部的边缘网关上运行边缘操作系统(edge operation system, EdgeOS),如图 6 所示.利用该操作系统,在家庭内部较易连接和管理智能家居设备,并在本地处理这些设备所产生的数据,降低数据传输带宽的负载,同时基于 EdgeOS 的应用服务程序可向用户提供更好的资源管理和分配.

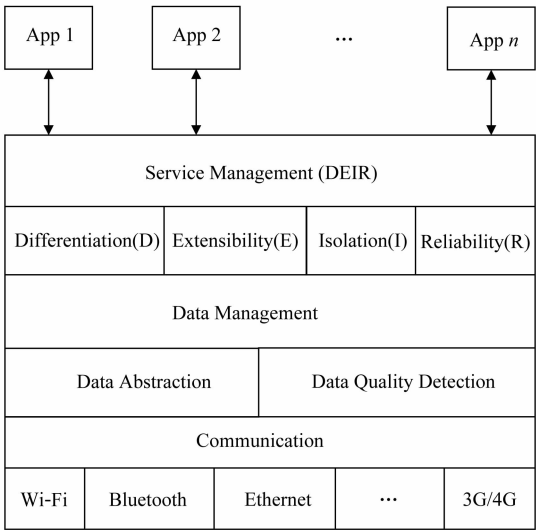


Fig. 6 Structure of EdgeOS in the smart home environment

图 6 智能家居的边缘操作系统结构

图 6 表示 EdgeOS 在智能家居中的一种变体.EdgeOS 需要从移动设备中收集数据,多种设备利用不同的通信协议进行通讯,如 WiFi、蓝牙、局域网以及蜂窝网络等.不同来源的数据在数据抽象层进

行融合和处理(详见 3.3 节).数据抽象层之上是服务管理层,该层需满足服务差异性(differentiation)、可扩展性(extensibility)、隔离性(isolation)及可靠性(reliability)(详见 3.4 节)等需求.此外,命名规则在每层内因功能不同而有所差异(详见 3.2 节).

2.4 智慧城市

边缘计算模型可从智能家居灵活地扩展到社区甚至城市的规模.根据边缘计算模型中将计算最大程度迁移到数据源附近的原则,用户需求在计算模型上层产生并且在边缘处理.边缘计算可作为智慧城市中一种较理想的平台,主要取决于以下 3 个方面:

1) 大数据量.据思科全球云指数预测^[9],到 2019 年,一个百万人口的城市每天将产生 180 PB 的数据,其主要来自于公共安全、健康数据、公共设施以及交通运输等领域.用云计算模型处理这些海量数据是不现实的,因为云计算模型会引起较重传输带宽负载和较长传输延时.在网络边缘设备进行数据处理的边缘计算模型将是一种高效的解决方案.

2) 低延时.万物互联环境下,大多数应用具有低延时的需求(比如健康急救和公共安全),边缘计算模型可以降低数据传输时间,简化网络结构.此外,与云计算模型相比,边缘网络对决策和诊断信息的收集将更加高效.

3) 位置识别.如运输和设施管理等基于地理位置的应用,对于位置识别技术,边缘计算模型优于云计算模型.在边缘计算模型中,基于地理位置的数据可进行实时处理和收集,而不必传送到云计算中心.

2.5 智能交通

智能交通是解决城市居民面临的出行问题,如恶劣的交通现状、拥堵的路面条件、贫乏的停车场地、窘迫的公共交通能力等.智能交通控制系统实时分析由监控摄像头和传感器收集的数据,并自动做出决策.这些传感器模块用于判断目标物体的距离和速度等.随着交通数据量的增加,用户对交通信息的实时性需求也在提高,若传输这些数据到云计算中心,将造成带宽浪费和延时等待,也不能优化基于位置识别的服务.在边缘服务器上运行智能交通控制系统来实时分析数据,根据路面的实况,利用智能交通信号灯减轻路面车辆拥堵状况或改变行车路线.同样,智能停车系统可收集用户周围环境的信息,在网络边缘分析用户附近的可用资源,并给出指示.

无人驾驶汽车(如特斯拉、谷歌汽车)是车辆智能化的一种表现形式,其主要依靠车内以计算机系统为主的智能驾车仪,通过车载传感系统感知路面

环境,自动规划行车路线并控制车辆到达预定目标来实现无人驾驶.它能针对实时交通情况做出合理决策,并辅助甚至替代驾驶员驾驶车辆的能力,从而减小驾驶员的劳动强度,使车辆行驶过程变得更安全.传感器数据上传到云计算中心会增加实时处理的难度,因此,在数据源(汽车上)执行边缘计算可加速处理,增强路面环境决策的实时性.此外,无人机(如大疆)本身的电源有限,如果数据传输到云中心,会消耗较大的电能,同时实时性也较弱.如空中或地面监测的应用中,无人机对森林火灾、倒塌的建筑物以及田地等监测所产生的大量数据以高清视频的形式存在,很难实现无线网络的实时传输以及接收中心的命令.在灾难环境混乱的情况下,这些问题就会更加凸显,边缘计算较好地解决了这些问题,边缘端处理无人机感知的数据,降低数据传输的电能损耗,保证实时性.

此外,对于多飞行器之间的协调控制,边缘计算模型除了能够实现飞行器本身所采集数据的实时处理,同时与其他飞行器实时共享这些信息,这样降低了原有云计算模型下经数据中心中转的时间,并且减少了因数据传输所消耗的电能.

2.6 协同边缘

云计算中,由于隐私和数据传输成本,数据拥有者很少与他人分享数据.边缘可以是物理上具有数据处理能力的一种微型数据中心,连接云端和边缘端.协同边缘是连接多个数据拥有者的边缘,这些数据拥有者在地理上是分布的,但具有各自的物理位置和网络结构^[42].类似于点对点的边缘连接方式,在数据拥有者之间提供数据的共享.

如图 7 所示的互联网医疗涉及到分布式地理数据处理,需多企业间合作和共享数据.为了消除共享障碍,协同边缘融合了由虚拟共享数据视图所创建的分布式地区数据.利用预先定义的服务接口,终端

用户可虚拟共享数据,而服务应用程序向终端用户提供所需服务.这些服务由协同边缘的参与者提供,计算任务仅在参与者内部执行,对终端用户透明,确保数据的隐私性和完整性.

我们以流感病情为例阐述协同边缘的优势,互联网医疗中,医院总结、分享流感疫情的信息(如平均花费、临床特征及感染人数信息等).医院治疗流感病人后更新其电子病历,病人根据药方从药房买药,若病人未按照医嘱进行治疗,导致重返医院治疗,医院须为病人的二次治疗负责,由此引起医疗责任纠纷,因为医院没有证据证明病人未按照药方来治疗.利用协同边缘,药房可以将该病人的购买记录推送到医院,这有助于解决医疗责任纠纷.此外,利用协同边缘,药房检索由医院提供的流感人数,根据现有库存来存放药品,以便获得最大利润.药房利用制药公司提供的数据,向物流公司推送一个关于运输价格的询问请求.根据检索到的信息,药房制定总成本最优方案和药物采购计划.制药公司可在收到药房的流感药品订单信息之后,重新制定药品的生产计划,调整库存.疾病控制中心在大范围区域内监控流感人群的变化趋势,可据此在有关区域内发布流感预警,采取措施阻止流感的扩散.

基于保险单规定,保险公司必须报销流感病人部分医疗消费.保险公司可以分析流感爆发期间感染人数,这与治愈流感所花费的成本作为调整下一年保单价格的重要依据.而且,如果患者愿意分享,保险公司可根据患者电子病历提供个性化的医疗政策.

可见,从减少操作成本和提高利润的角度,通过该案例,大多数参与者(药店、药厂等)可以利用协同边缘来获益.个人病例信息作为源数据,医院担任源数据收集的角色,对于社会医疗健康而言,医院可以提前做好资源的分配以此来提高服务效率.

3 挑 战

第 2 节描述了边缘计算的 6 种潜在应用.为了实现边缘计算的设想,我们认为计算机系统、网络以及应用服务程序的研究和开发人员需进行紧密的合作和交流.本节总结在边缘计算研究中可能遇到的迫切要解决的 7 个关键问题,并结合我们的研究成果,提出一些解决思路和值得进一步思考的研究方向,主要包括:可编程性、命名规则、数据抽象、服务管理、数据隐私保护及安全、理论基础以及商业模式.

3.1 可编程性

云计算模型中,用户编写应用程序并将其部署

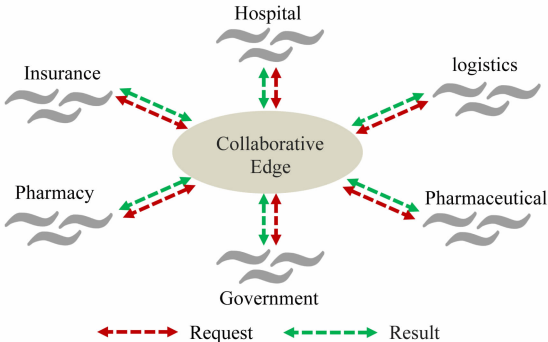


Fig. 7 Collaborative edge example: connected health
图 7 协同边缘案例:连接医疗

到云端. 云服务提供商维护云计算服务器,用户对程序的运行完全不知或知之较少,这是云计算模型下应用程序开发的一个优点,即基础设施对用户透明. 用户程序通常在目标平台上编写和编译,在云服务器上运行. 边缘计算模型中,部分或全部的计算任务从云端迁移到边缘节点,而边缘节点大多是异构平台,每个节点上的运行时环境可能有所差异,因此,在边缘计算模型下部署用户应用程序时,程序员将遇到较大的困难. 而现有传统编程方式 MapReduce, Spark 等均不适合,需研究基于边缘计算的编程方式.

为了实现边缘计算的可编程性,我们提出了一种计算流的概念. 计算流是指沿着数据传输路径,在数据上执行的一系列计算/功能. 计算/功能可以是某个应用程序的全部或部分函数,其发生在允许应用执行计算的数据传输路径上. 该计算流属于软件定义计算流的范畴,主要应用于源数据的设备端、边缘节点以及云计算环境中,以实现高效分布式数据处理.

烟花模型(firework)^[43]是我们提出的一种基于边缘计算的编程模型,其主要包括烟花模型管理器(firework manager)和烟花模型节点(firework

node)两部分,如图 8 所示. 万物互联时代,数据生产和消费被迁移到边缘设备上,增加了大数据分布式共享和处理的需求. 为此,我们提出了烟花模型,其可实现大数据分布式共享和处理,并使私有数据可以在数据利益相关者的设备上处理(如图 7 所示不同烟花模型节点). 烟花模型通过创建虚拟的共享数据视图,融合了地理上分布的数据源,而数据利益相关者(烟花模型节点)为终端用户提供一组预定义的功能接口以使用户访问.

烟花模型的接口形式是一组数据集(datasets)和功能(functions),如图 8 烟花模型节点,并且功能与数据集绑定. 烟花模型使数据处理更接近数据生产者,避免了从网络边缘设备到云中心的长距离数据传输,降低响应延迟. 在烟花模型中,所有数据利益相关者(烟花模型节点)都需要注册其各自的数据集及相应功能,以便抽象成一种数据视图. 已注册的数据视图对同一个烟花模型中所有参与者均是可见的,任何参与者可以将多个数据视图进行组合,以实现特定情境下的数据分析. 烟花模型管理器将组合数据视图的服务请求分解成若干子任务,并发送给每一个参与者,每个子任务将在其本地设备上执行相应的计算任务.

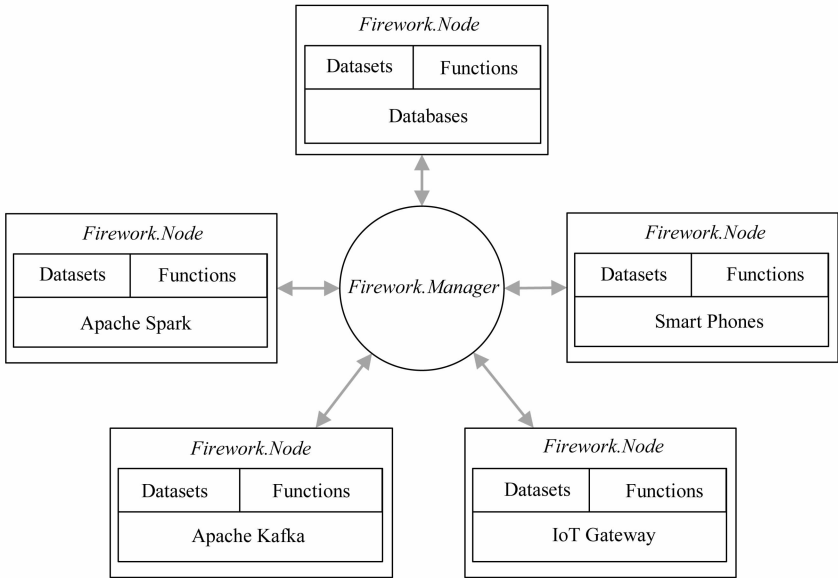


Fig. 8 A high level overview of Firework

图 8 烟花模型

烟花模型扩展了数据的可视化边界,为协同边缘环境下的分布式数据处理提出了一种新的编程模式. 烟花模型中每个参与者可以在本地设备上实现数据处理,实现云计算和边缘计算资源的融合.

此外,需要注意的是边缘计算模型中协同问题

(如同步、数据/状态的迁移等)是可编程性方面亟待解决的问题之一.

3.2 命名规则

边缘计算模型中一个重要假设是边缘设备的数目巨大. 边缘节点平台上运行多种应用程序,每个应

用程序提供特定功能的服务. 与计算机系统的命名规则类似, 边缘计算的命名规则对编程、寻址、识别和数据通信具有非常重要的作用, 而当前暂无较为高效的命名规则. 为了实现系统中异构设备间的通信, 边缘计算研究者需要学习多种网络和通信协议. 此外, 边缘计算的命名规则需要满足移动设备、高度动态的网络拓扑结构、隐私安全等需求.

传统的命名机制如 DNS, URI 满足大多数的网络结构, 但却不能灵活地为动态边缘网络提供服务, 原因在于大多数的边缘设备具有高度移动性和有限资源, 而对于该类边缘设备而言, 基于 IP 的命名规则, 因复杂性和开销太大而难以应用到边缘计算中.

已有研究中, 命名数据网络(named data networking, NDN)^[44] 和移动优先(MobilityFirst)^[45] 等新的命名机制可满足边缘计算的需求. NDN 不仅提供了以数据为中心的网络分层结构命名规则和友好的服务管理, 还保障了边缘计算具有可扩展性. 为了适合如蓝牙或 ZigBee 等通信协议, NDN 需要额外的代理. 由于很难将设备硬件信息与服务提供商隔离, 因此 NDN 存在一定的安全隐患. 移动优先技术将命名和网络地址分开, 以更好地支持移动性. 如果边缘服务器上的程序具有较高的移动性, 那么移动优先技术的应用就会提高边缘服务器的效率. 但是, 在移动优先技术中, 命名规则需要一种全局唯一标识符, 而在网络边缘环境(家居环境)下, 固定信息聚合服务程序不需要这种标识符. 移动优先技术的另一个缺点是全局唯一标识符(globally unique identifier, GUID)不够人性化, 使服务程序较难管理.

在范围较小且固定的边缘环境下(如家居环境), 利用边缘操作系统给每个设备分配网络地址可能是一种解决方案. 系统内部的每个设备拥有唯一一个可以描述其位置(where)、角色(who)以及数据描述(what)的命名. 如图 9 所示, 边缘操作系统将分配标识符(identifier)和网络地址(address)给边缘设备, 每个设备的命名是唯一的且用于服务管理、设备识别和部件替换. 这个命名规则使用户和服务提供者较易管理服务, 同时阻止服务提供商得到硬件信息, 保护隐私数据. 设备的命名可正确匹配到其标识符和网络地址, 标识符用于边缘操作系统的设备管理, 而网络地址如 IP 地址或 MAC 地址则用于支持如蓝牙、ZigBee 或 WiFi 等多种网络通信协议. 当应用到城市级别的系统时, 命名规则仍是一个待解决的问题.

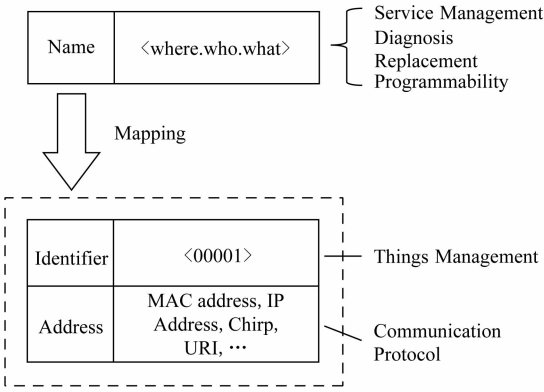


Fig. 9 Naming mechanism in EdgeOS

图 9 边缘操作系统的命名机制

3.3 数据抽象

边缘操作系统的应用程序通过服务管理层 API 来使用数据或提供服务. 相比云计算, 边缘计算中数据抽象更具挑战性. 智能家居中, 作为数据生产者的智能设备均向边缘操作系统发送数据, 但部署在家庭周围的设备较少, 而大多数网络边缘设备, 均会定时地向网关发送感知数据. 如温度传感器每分钟向网关发送温度数据, 但使用频率较低. 基于此, 我们提出减少边缘计算中人为参与, 而由边缘节点处理, 用一种主动的方式将结果与用户交互. 该情况下, 网关层实现对数据的预处理(如去噪、事件检测、隐私保护等), 之后数据被发送至系统上层, 作为应用服务所需源数据, 该过程将遇到 3 种挑战:

1) 不同设备所传输数据格式的多样性. 如图 10 所示, 基于数据隐私和安全的考虑, 源数据对网关运行的任务是透明的, 该类任务应从整合的数据表中提取其处理所需的信息. 我们提出含有编号(ID)、名字(Name)、时间(Time)和数据(Data)的表结构, 以便边缘设备数据可以存入该表, 而这会隐藏感知数据的细节影响数据使用.

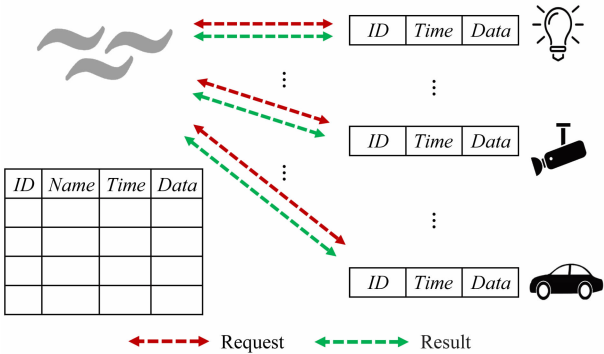


Fig. 10 Data abstraction issue for edge computing

图 10 边缘计算的数据抽象

2) 数据抽象程度的不确定性. 如果数据抽象过滤较多的源数据, 将导致一些应用或服务程序因无法获得足够信息而运行失败; 反之, 若保留大量源数据, 数据存储和管理将是系统开发者所面临的另一种挑战. 此外, 边缘设备发送的数据具有不可靠性, 如何从不可靠信息源中抽象出有用的信息仍是一个技术挑战.

3) 数据抽象的适用性. 边缘设备端收集数据并提供给应用程序使用, 完成特定的服务, 应用程序应具有读、写设备的权限, 为用户的特定需求提供服务. 数据抽象层将数据的表示和操作结合, 并为连接到边缘操作系统的设备提供一种公共的交互接口. 由于边缘设备的异构性, 数据表示及操作也有所不同, 这将成为通用数据抽象的障碍.

3.4 服务管理

边缘计算的服务管理方面, 任意一种可靠系统均具有 4 种特征: 即差异性 (differentiation)、可扩展性 (extensibility)、隔离性 (isolation) 及可靠性 (reliability) (简称 DEIR 模型)^[17].

1) 差异性. 随着万物互联应用的快速发展, 我们期望在网络的边缘部署多种服务, 而这些服务的优先级不同. 如心力衰竭检测健康相关的服务高于其他娱乐服务的优先级.

2) 可扩展性. 网络边缘的可扩展性在未来研究中将是一个巨大的挑战. 当某个设备损耗后, 新的设备是否可以继续之前的服务等. 针对这些问题, 需要设计一种灵活且可扩展的边缘操作系统来实现服务层的管理.

3) 隔离性. 隔离性是网路边缘的重要问题之一. 手机操作系统中应用程序崩溃通常会导致系统的崩溃或重启. 现有分布式操作系统采用不同的同步机制来管理共享资源, 常见方法有加锁或令牌环机制. 这种问题在边缘操作系统中将更加复杂, 由于存在多个应用共享同一种数据. 如智能家居中照明的控制, 如果应用程序崩溃但系统能正常运行, 用户仍能对照明进行控制. 我们发现引入/取消部署框架是未来解决这个挑战的一种研究方向. 隔离性的另一个挑战是如何隔离第三方程序与用户的隐私数据, 如活动跟踪类应用程序不具有访问电量的权限. 为此, 边缘操作系统中服务管理层应增加符合应用场景的访问控制机制.

4) 可靠性. 可靠性是边缘计算模型挑战之一, 从服务、系统和数据的角度给予阐述:

① 服务角度. 在实际场景下, 有时较难确定服务失败的具体原因, 如空调停止工作, 原因可能是电源线断、压缩机故障或温控器的电量用完等. 当节点断开连接时, 系统的服务很难维持, 但在节点出现故障后, 可采取方法来降低终止服务的风险. 如边缘操作系统告知用户哪一部件出现问题. 我们发现, 采用无线传感网或者如 PROFINET^[46] 工业网络可作为可靠性的一种有效解决方案.

② 系统角度. 边缘操作系统能够较好地维护整个系统的网络拓扑结构是非常重要的, 系统中每个组件能够发送状态/诊断信息到边缘操作系统. 基于这种特点, 用户可以方便在系统层部署如故障检测、设备替换和数据质量检测等服务.

③ 数据角度. 可靠性主要取决于数据感知和通信. 边缘设备故障原因不尽相同, Shi 等人^[47] 研究发现, 低电量等不可靠情况下, 边缘设备会发送精度较低的数据. 大数据背景下数据收集方面, DaCosta^[48] 提出多种新型通信协议. 这些协议可较好地适应大规模传感器节点和高度动态网络, 但相比蓝牙或者 WiFi, 其连接的可靠性较差. 在传感器数据通信不可靠的情况下, 如何利用多维参考数据源和历史数据, 提供可靠的服务仍是一个难题.

3.5 数据隐私保护及安全

数据隐私保护及安全是边缘计算提供的一种重要服务. 如在家庭内部署万物互联系统, 大量的隐私信息会被传感器捕获, 如何在隐私保护下提供服务将是一种挑战. 我们发现在数据源附近进行计算是保护隐私和数据安全的一种有效方法. 在边缘计算中, 数据隐私保护及安全的研究将面临以下挑战:

1) 社会对隐私和安全的意识. 本文以 WiFi 网络安全为例. 调查表明^[49], 在 4 亿多使用无线连接的家庭中, 49% 的 WiFi 网络是不安全的, 80% 家庭仍然使用默认密码设置他们的路由器. 对于公共的 WiFi 热点, 89% 的热点是不安全的. 如果用户没有保护好个人隐私数据, 很容易被他人利用网络摄像头、健康监测仪等设备, 窥探个人的隐私数据.

2) 边缘设备兼顾数据收集者和所有者. 如手机收集的数据将在服务提供商处存储和分析, 而保留边缘数据且让用户拥有这些数据是一种较好保护隐私数据的方案. 网络边缘设备所收集的数据应存储在边缘, 并且用户应有权限限制服务提供商使用这些数据. 为保护用户隐私, 应在边缘设备中删除高度隐私的数据.

3) 数据隐私和安全有效工具的缺乏. 网络边缘设备资源有限, 现有数据安全的方法并不能完全适用于边缘计算. 而且, 网络边缘高度动态的环境也会使网络更易受到攻击. 为加强对隐私数据的保护, 研究人员对隐私保护平台进行了研究, 如 Deborah 团队开发的 Open mHealth 平台^[50], 以实现健康数据的标准化处理和存储, 但未来的研究仍需要开发更多的工具来处理边缘计算的数据.

3.6 理论基础

CAP 理论是分布式系统理论的基础, 该理论是分布式系统、特别是分布式存储领域中被使用最多的理论. Brewer 在 Inktomi^[51] 期间研发搜索引擎、分布式 Web 缓存时得出关于数据一致性 (consistency)、服务可用性 (availability)、分区容错性 (partition-tolerance) 的猜想, 并在 2000 年 PODC 会议上提出^[52-53], 该猜想在提出 2 年后被证明成立^[54] 并作为 CAP 定理.

因此, 基于边缘计算模型的计算机系统研究中, 边缘计算理论基础将是学界和产业界进行边缘计算研究所面临的关键性挑战之一. 边缘计算是一种综合性很强的科学研究, 横跨计算、数据通信、存储、能耗优化等多个领域. 一方面, 边缘计算理论可以基于一种多目标优化的理论为基础, 实现计算、数据通信以及能耗的综合最优; 另一方面, 可以分别在计算、通信、能耗等不同维度建立边缘计算相关理论基础, 如计算维度上建立计算任务的负载均衡理论, 指导云中心与边缘端的任务分配, 实现云中心和边缘端计算能力的最大使用效率; 同时根据计算负载均衡和分布式系统理论, 边缘端与云端的数据通信, 最优化网络传输带宽; 研究分布式多维边缘端设备能耗理论模型 (如利用多维度李亚普诺夫理论), 建立多边缘端的能耗效率模型, 优化边缘端设备能耗, 提高有限能量资源的利用率, 此外, 也可以根据类似李亚普诺夫可靠性理论, 建立基于多边缘设备的边缘计算可靠性理论.

边缘计算的理论基础当前并不成熟, 需要综合计算、数据通信、存储及能耗优化等多学科已有比较完善的理论基础, 提出综合性或多维度的边缘计算理论, 这是目前我们在开展边缘计算研究中首要解决的关键性问题. 合理的边缘计算理论基础对学界和产业界未来更好地开展基于边缘计算模型的应用服务研究和开发工作具有极为重要的指导意义.

3.7 商业模式

云计算的商业模型比较简单, 用户通过自己的

需求向相关服务提供商进行购买, 具体而言, 云计算所提供的云服务是基于互联网相关服务的增加、使用和交付模式, 通常涉及通过互联网来提供动态易扩展且虚拟化的资源. 云计算需求客户通过网络以按需、易扩展的方式获得所需服务. 这种服务可以是 IT 基础设施和软件资源以及互联网相关的其他资源或服务. 云计算的计算能力也可作为一种服务或商品, 通过互联网进行流通.

边缘计算横跨信息技术 (IT)、通讯技术 (CT) 等多个领域, 涉及软硬件平台、网络联接、数据聚合、芯片、传感、行业应用等多个产业链角色. 边缘计算的商业模型更多的将会不仅是以服务为驱动, 用户请求相应的服务, 而更多的将以数据为驱动, 如我们在边缘计算的烟花模型中提到, 每个用户需求向数据拥有者 (利益相关者) 提供数据请求, 然后云中心或是边缘端数据拥有者将处理的结果反馈给用户, 由原来的中心-用户的单边商业模式转变为用户-中心、用户-用户的多边商业模式.

边缘计算的商业模型取决于参与该模型的多个利益相关者, 如何结合现有的云计算商业模型, 发展边缘计算的多边商业模型也是边缘计算所面临的重要问题之一.

4 相关研究工作

随着大数据时代的发展, 为了解决云计算中心计算负载和数据传输带宽的问题, 研究者也提出多种关于计算任务从云计算中心迁移到网络的边缘的技术, 其中主要典型模型包括: 分布式数据库模型、P2P 模型、CDN 模型、移动边缘计算模型、雾计算模型以及海云计算. 本节我们将分别阐述不同模型的相关研究工作.

4.1 分布式数据库模型

分布式数据库系统是数据库技术和网络技术两者结合的结果. 大数据时代, 数据种类和数量的增长使分布式数据库成为数据存储和处理的核心技术. 分布式数据库部署在自组织网络服务器或分散在互联网、企业网或外部网以及其他自组织网络的独立计算机上^[55]. 数据存储在多台计算机上, 分布式数据库操作不局限于单台机器, 而允许在多台机器上执行事务交易, 以此来提高数据库访问的性能^[56].

分布式数据库已成为大数据处理的核心技术. 按照数据库的结构, 分布式数据库包括同构^[57] 和异构系统^[58]. 前者数据库实例的运行环境具有相同的

软件和硬件,同构分布式数据库具有单一的访问接口;后者的运行环境中硬件、操作系统和数据库管理系统以及数据模型等均有所不同。按照处理数据类型,分布式数据库主要包括 SQL(关系型)、NoSQL(非关系型)、基于可扩展标记语言(XML)以及 NewSQL 分布式数据库。其中, NoSQL 和 NewSQL 分布式数据库使用最为广泛。NoSQL 分布式数据库^[59]主要为满足大数据环境下海量数据对数据库高并发、高效存储访问、高可靠性和高扩展性的需求,主要分为键值存储类、列存储数据库、文档型数据库、图形数据库等。NewSQL 分布式数据库^[60-62]是一种具有实时性、复杂分析、快速查询等特征的,面向大数据环境下海量数据存储的关系型分布式数据库,主要包括 Google Spanner, Clustrix, VoltDB 等。SQL 分布式数据库是针对表式结构的关系型分布式数据库,典型代表有微软分布式数据库^[63]和 Oracle 分布式数据库^[64]。基于 XML 的分布式数据库主要存储以 XML 为格式的数据,本质上是一种面向文档的类似于 NoSQL 的分布式数据库^[65]。

相比于边缘计算模型,分布式数据库提供了大数据环境下的数据存储,较少关注其所在设备端的异构计算和存储能力,主要用以实现数据的分布式存储和共享。分布式数据库技术所需的空间较大且数据的隐私性较低,对基于多数据库的分布式事务处理而言,数据的一致性技术是分布式数据库均要面临的重要挑战^[66]。边缘计算模型中数据位于边缘设备端,具有较高的隐私性、可靠性和可用性。万物互联时代,“终端架构具有异构性并需支持多种应用服务”将成为边缘计算模型应对大数据处理的基本思路。

4.2 P2P(peer-to-peer computing)

P2P 计算^[67]不仅与边缘计算紧密相关,而且还是较早将计算迁移到网络边缘的一种文件传输技术。P2P 的术语于 2000 年首次被提出并用于实现文件共享系统,此后,其逐渐发展成为分布式系统的重要子领域,其中分散化、最大化可扩展性、容忍较高层节点流失以及恶意行为防止已经成为 P2P 主要的研究主题,该领域的主要成就包括:1) 分布式 Hash 表,其后来演变为云计算模型中 key-value 分布式存储一般范式;2) 广义 Gossip 协议,其已被广泛地用于非简单信息扩散的复杂任务处理类应用中,如数据融合和拓扑管理;3) 多媒体流技术,其表现形式有视频点播、实时视频、个人通信等。但是, P2P 多数被用于非法文件共享和相关诉讼的广泛媒

体报道。结果,实际基于 P2P 模式的一些商业技术未得到承认。

边缘计算模式源于 P2P,但拓展了一些 P2P 中原来没有的新技术和新手段,将 P2P 的概念扩展到网络边缘设备,实现 P2P 计算和云计算的融合。

4.3 内容分发网络(content distribution networks, CDN)

内容分发网络是基于互联网的缓存网络^[68],通过在网络边缘部署缓存服务器来降低远程站点的数据下载延时,加速内容交付。2004 年,研究人员实现了一种新的体系结构模型^[69],提出 ACDN 原型,作为在传统 CDN 的一种改善,帮助内容提供商免于预测预先配置的资源 and 决定资源的位置。ACDN 允许应用部署在任意一台服务器上,通过设计一些新的算法,根据需要迁移和复制应用到服务器。

CDN 自 1998 年 Akamai 公司^[70]提出至今,得到了学术界和工业界的高度关注而快速发展。如亚马逊^[71], Akamai 等公司拥有了比较成熟的 CDN 技术, Akamai 公司利用 CDN 技术研发的中国 CDN^[72],为我国用户交付期望获得的性能和体验,同时也降低了提供商的组织运营压力。近年来,我国学术界研究 CDN 优化技术,如清华大学团队设计和实现的边缘视频 CDN^[73],其提出利用数据驱动的方法来组织边缘内容热点,基于请求预测的服务器峰值转移的复制策略,实现把内容从服务器复制到边缘热点上为用户提供服务。产业界也涌现出许多 CDN 服务公司,如 chinacache(蓝汛)^[74]、网宿^[75]等。

内容分发网络的缓存服务器与边缘计算模型中边缘服务器类似,均位于网络的边缘。但是,边缘计算模型的“边缘”不限制在边缘节点,还包括网络边缘的摄像头、智能手机、网关、可穿戴的计算设备和传感器等设备。

4.4 移动边缘计算(mobile edge computing, MEC)

万物互联的发展实现了网络中多类型设备(如智能手机、平板、无线传感器及可穿戴的健康设备等)的互联,而大多数网络边缘设备的能量和计算资源有限,这使万物互联的设计变得尤为困难。移动边缘计算^[76]是在接近移动用户的无线电接入网范围内,提供信息技术服务和云计算能力的一种新的网络结构,并已成为一种标准化、规范化的技术。2014 年 ETSI 提出对移动边缘计算术语的标准化^[77],并指出移动边缘计算提供了一种新的生态系统和价值链。利用移动边缘计算,可将密集型移动计算任务迁移到附近的网络边缘服务器^[78-79]。由于移动边缘

计算位于无线接入网内并接近移动用户,因此可以实现较低延时、较高带宽来提高服务质量和用户体验.移动边缘计算同时也是发展 5G 的一项关键技术^[80],有助于从延时、可编程性、扩展性等方面满足 5G 的高标准要求.移动边缘计算通过在网络边缘部署服务和缓存,中心网络不仅可以减少拥塞,还能高效地响应用户请求.

任务迁移是移动计算技术难点之一,已有的优化算法主要包括 LODCO 算法^[81]、分布式计算迁移^[82]、EPCO 和 LPCO 算法^[83]以及 Actor 模型^[84]等.移动边缘计算已被应用到车联网、物联网网关、辅助计算、智能视频加速、移动大数据分析等多种场景^[85-86].

移动边缘计算模型强调在云计算中心与边缘设备之间建立边缘服务器,在边缘服务器上完成终端数据的计算任务,但移动边缘终端设备基本认为不具有计算能力.相比而言,边缘计算模型中终端设备上具有较强的计算能力,因此,移动边缘计算是一种边缘计算服务器,作为边缘计算模型的一部分.

4.5 雾计算(fog computing)

思科于 2012 年提出雾计算^[42],并将雾计算定义为迁移云计算中心任务到网络边缘设备执行的一种高度虚拟化的计算平台.雾计算在终端设备和传统云计算中心之间提供计算、存储和网络服务,是对云计算的补充. Vaquero 等人对雾计算进行了较全面的定义^[87],雾计算^[88]通过在云与移动设备之间引入中间层,扩展了基于云的网络结构,而中间层实质是由部署在网络边缘的雾服务器组成的“雾层”^[89].雾计算避免云计算中心和移动用户之间多次通信.通过雾服务器,可以显著减少主干链路的带宽负载和能耗,在移动用户量巨大时,可以访问雾服务器中缓存的内容、请求一些特定的服务^[90].此外,雾服务器可以与云计算中心互连,并使用云计算中心强大的计算能力和丰富的应用和服务.

边缘计算和雾计算^[91]概念具有很大的相似性,在很多场合表示同一个意思.如果要仔细区分二者,我们认为边缘计算除了关心基础设施,也关注边缘设备,包括 Things 的管理等,而雾计算则更多是指基础设施.

4.6 海云计算

万物互联背景下,待处理数据量达到 ZB 级,信息系统的感知、传输、存储和处理的能力需提高 3 个数量级,同时整体能耗要保持在 2010 年的水准,也就是说,需实现 1000 倍每瓦性能的优化.针对这一挑战,中国科学院于 2012 年启动了 10 年战略优先

研究倡议,称之为下一代信息与通信技术倡议(next generation information and communication technology initiative, NICT).倡议的主旨是要开展“海云计算系统项目”的研究,其核心是通过“云计算”系统与“海计算”^[92]系统的协同和集成,增强传统云计算能力,其中,“海”端指由人类本身、物理世界的设备和子系统组成的终端(客户端).“海云计算系统项目”目标是实现面向 ZB 级数据处理的能效要比现有技术提高 1000 倍,研究内容主要包括从整体系统结构层、数据中心级服务器及存储系统层、处理器芯片级等角度提出系统级解决方案.

与边缘计算相比而言,海云计算关注“海”的终端设备,而边缘计算是从“海”到“云”之间的任意中间计算资源和网络资源,海云计算是边缘计算的一个非常好的子集实例.

5 总 结

边缘式大数据处理时代下,云计算模型已无法有效解决云中心负载、传输带宽、数据隐私保护等问题.万物互联应用服务快速发展催生了边缘计算,其是万物互联背景下边缘式大数据处理时代的软硬件关键支撑平台.边缘计算模型中数据处理模式可以保证较短的响应时间和较高的可靠性,同时越来越多的应用服务将从云计算中心迁移到网络边缘设备端.此外,如果大部分数据能在边缘设备上被处理而不用上传到云计算中心,这样就大大节省传输带宽和设备端电能的消耗.

本文提出了边缘计算的定义及其基本原理,列举边缘计算的 6 个应用案例.介绍协同边缘,因为边缘可以实现终端用户与云计算中心物理层或逻辑层的连接.边缘计算模型仍支持传统的云计算模型,而且还可连接远程计算资源,实现数据的共享和协作.提出边缘计算面临的挑战,包括可编程性、命名规则、数据抽象、服务管理、数据的隐私保护与安全、理论基础及商业模式.我们希望通过本文引起产业界和学术界对边缘计算的关注.然而,要实现边缘计算模型这一愿景,除了计算机系统、通信、网络和应用程序等技术领域的研究人员参加之外,还需要能够从这些技术领域受益的其他机构加入,如环境和公共卫生、执法、消防、以及公用事业服务等.

在过去的几年里,边缘计算愿景实现的过程已经开始,如 2015 年 10 月,雾计算的支持者组成开放

雾联盟^①。该联盟旨在通过汇集公司、高校科研机构、研究者个人等资源,加快雾计算技术的部署,促进雾计算生态系统的快速形成。美国联邦政府包括国家科学基金会、美国国家标准局,在2016年都分别把边缘计算列入了项目申请指南。此外,边缘计算领域的相关国际会议已经开始兴起,如2016年10月在美国华盛顿特区举办的第一届IEEE/ACM边缘计算会议(IEEE/ACM Symposium on Edge Computing, SEC)、将于2017年10月举办的边缘计算和雾计算世界联合大会等。2016年11月30日,中国成立了产学研结合的边缘计算产业联盟^②,旨在搭建边缘计算产业合作平台,推动运行技术(operation technology, OT)和信息与通信技术(information and communications technologies, ICT)产业开放协作,引领边缘计算产业蓬勃发展,深化行业数字化转型。

我们相信,按照这种发展趋势继续进行下去,万物互联背景下边缘计算模型的愿景定在不久的将来成为现实。

致谢 作者感谢《计算机研究与发展》杂志主编徐志伟研究员的邀请和早期的讨论。另外,我们也要感谢李尤慧子博士为本文所做的前期内容整理和修正工作,感谢安徽大学的陈国栋、梁旭、刘永辉同学为本文所做的资料收集和校验工作!

参 考 文 献

- [1] Armbrust M, Fox A, Griffith R, et al. A view of cloud computing [J]. Communications of the ACM, 2010, 53(4): 50-58
- [2] Ghemawat S, Gobioff H, Leung S T. The Google file system [J]. ACM SIGOPS Operating Systems Review, 2003, 37(5): 29-43
- [3] Dean J, Ghemawat S. MapReduce: Simplified data processing on large clusters [J]. Communications of the ACM, 2008, 51(1): 107-113
- [4] Shvachko K, Kuang H, Radia S, et al. The Hadoop distributed file system [C] //Proc of the 26th IEEE Symp on Mass Storage Systems and Technologies (MSST). Piscataway, NJ: IEEE, 2010: 1-10
- [5] Zaharia M, Chowdhury M, Franklin M J, et al. Spark: Cluster computing with working sets [C] //Proc of HotCloud 2010. Berkeley, CA: USENIX Association, 2010: 10
- [6] Ashton K. That ‘Internet of things’ thing [J]. RFiD Journal, 2009, 22(7): 97-114
- [7] Sundmaeker H, Guillemin P, Friess P, et al. Vision and challenges for realizing the Internet of things [OL]. [2016-12-03]. http://www.robvankranenburg.com/sites/default/files/Rob%20van%20Kranenburg/Clusterbook%202009_0.pdf
- [8] Gubbi J, Buyya R, Marusic S, et al. Internet of things (IoT): A vision, architectural elements, and future directions [J]. Future Generation Computer Systems, 2013, 29(7): 1645-1660
- [9] David E C. The once and future Internet of everything [OL]. [2016-12-03]. http://sites.nationalacademies.org/cs/groups/cstbsite/documents/webpage/cstb_160416.pdf
- [10] Networking C V. Cisco Global Cloud Index: Forecast and Methodology, 2015—2020 [OL]. [2016-12-03]. <http://www.cisco.com/c/dam/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.pdf>
- [11] Evans D. The Internet of things how the next evolution of the Internet is changing everything [OL]. [2016-12-03]. https://www.researchgate.net/publication/306122904_The_Internet_of_Things_How_the_Next_Evolution_of_the_Internet_is_Changing_Everything
- [12] Armbrust M, Fox A, Griffith R, et al. A view of cloud computing [J]. Communications of the ACM, 2010, 53(4): 50-58
- [13] European Telecommunications Standards Institute (ETSI). Mobile-edge computing [OL]. [2016-12-03]. https://portal.etsi.org/Portals/0/TBpages/MEC/Docs/Mobile-edge_Computing_-_Introductory_Technical_White_Paper_V1%2018-09-14.pdf
- [14] Dastjerdi A V, Gupta H, Calheiros R N, et al. Fog computing: Principals, architectures, and applications [OL]. [2016-12-03]. <https://arxiv.org/ftp/arxiv/papers/1601/1601.02752.pdf>
- [15] Gai Keke, Qiu Meikang, Zhao Hui, et al. Dynamic energy-aware cloudlet-based mobile cloud computing model for green computing [J]. Journal of Network & Computer Applications, 2015, 59(C): 46-54
- [16] Xu Zhiwei. Cloud-sea computing systems: Towards thousand-fold improvement in performance per watt for the coming zettabyte era [J]. Journal of Computer Science and Technology, 2014, 29(2): 177-181
- [17] Shi Weisong, Cao Jie, Zhang Quan, et al. Edge computing: Vision and challenges [J]. IEEE Internet of Things Journal, 2016, 3(5): 637-646
- [18] Buyya R, Yeo C S, Venugopal S, et al. Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility [J]. Future Generation Computer Systems, 2009, 25(6): 599-616

^① <http://www.openfogconsortium.org>

^② <http://www.econsortium.net/>

- [19] Turner V, Gantz J F, Reinsel D, et al. The digital universe of opportunities; Rich data and the increasing value of the Internet of things [OL]. [2016-12-03]. <https://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014.pdf>
- [20] Finnegan M. Boeing 787s to create half a terabyte of data per flight, says virgin atlantic. [OL]. [2016-12-03]. <http://www.computerworlduk.com/data/boeing-787s-create-half-terabyte-of-data-per-flight-says-virgin-atlantic-3433595/>
- [21] Datafloq. Self-Driving cars will create 2 petabytes of data, what are the big data opportunities for the car industry? [OL]. [2016-12-03]. <https://datafloq.com/read/self-driving-carscreate-2-petabytes-data-annually/172>
- [22] LeClaire J. IHS predicts 54 million self-driving cars by 2035 [OL]. [2016-12-03]. http://www.newsfactor.com/story.xhtml?story_id=013001CTNJ8D
- [23] Lopez G P, Montresor A, Epema D, et al. Edge-centric computing; Vision and challenges [J]. ACM Sigcomm Computer Communication Review, 2015, 45(5): 37-42
- [24] Cortés R, Bonnaire X, Marin O, et al. Stream processing of healthcare sensor data; Studying user traces to identify challenges from a big data perspective [J]. Procedia Computer Science, 2015, 52(1): 1004-1009
- [25] Greenberg A, Hamilton J, Maltz D A, et al. The cost of a cloud; Research problems in data center networks [J]. ACM Sigcomm Computer Communication Review, 2009, 39(1): 68-73
- [26] Beloglazov A, Abawajy J, Buyya R. Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing [J]. Future Generation Computer Systems, 2012, 28(5): 755-768
- [27] Sverdlik Y. Here's how much energy all US data centers consume [OL]. [2016-12-03]. <http://www.datacenterknowledge.com/archives/2016/06/27/heres-how-much-energy-all-us-data-centers-consume/>
- [28] Worthington S. Chinese data centers use enough electricity for two countries.[OL]. [2016-12-03]. <https://datacenternews.asia/story/chinese-data-centers-use-enough-electricity-two-countries/>
- [29] Varghese B, Wang N, Barbhuiya S, et al. Challenges and opportunities in edge computing [C] //Proc of IEEE Int Conf on Smart Cloud (SmartCloud). Piscataway, NJ: IEEE, 2016; 20-26
- [30] Cole D. Data center energy efficiency-looking beyond pue [OL]. [2016-12-03]. http://www.nolimitssoftware.com/docs/DataCenterEnergyEfficiency_LookingBeyond.pdf
- [31] Sharma M, Arunachalam K, Sharma D. Analyzing the data center efficiency by using PUE to make data centers more energy efficient by reducing the electrical consumption and exploring new strategies [J]. Procedia Computer Science, 2015, 48: 142-148
- [32] Lim S H, Sharma B, Tak B C, et al. A dynamic energy management scheme for multi-tier data centers [C] //Proc of the IEEE Int Symp on Performance Analysis of Systems and Software (ISPASS). Piscataway, NJ: IEEE, 2011; 257-266
- [33] Gao Yongqiang, Guan Haibing, Qi Zhengwei, et al. Service level agreement based energy-efficient resource management in cloud data centers [J]. Computers & Electrical Engineering, 2014, 40(5): 1621-1633
- [34] Domo Inc. Data never sleeps 2. 0 [OL]. [2016-12-08]. <https://www.domo.com/blog/2014/04/datanever-sleeps-2-0/>
- [35] Nippon Telegraph and Telephone Corporation. Announcing the "Edge computing" concept and the "Edge accelerated Web platform" prototype to improve response time of cloud applications [OL]. [2016-12-03]. <http://www.ntt.co.jp/news2014/1401e/140123a.html>
- [36] Chun B G, Ihm S, Maniatis P, et al. Clonecloud: Elastic execution between mobile device and cloud [C] //Proc of the 6th Conf on Computer Systems. New York: ACM, 2011; 301-314
- [37] Rudenko A, Reiher P, Popek G J, et al. Saving portable computer battery power through remote process execution [J]. ACM SIGMOBILE Mobile Computing and Communications Review, 1998, 2(1): 19-26
- [38] Hunt G C, Scott M L. The coign automatic distributed partitioning system [C] //Proc of the 3rd Symp on Operating Systems Design and Implementation. Berkeley, CA: USENIX Association, 1999; 187-200
- [39] Kumar K, Lu Y H. Cloud computing for mobile users; Can offloading computation save energy? [J]. Computer, 2010, 43(4): 51-56
- [40] Kosta S, Aucinas A, Hui P, et al. Thinkair: Dynamic resource allocation and parallel execution in the cloud for mobile code offloading [C] //Proc of the 31st ACM INFOCOM. Piscataway, NJ: IEEE, 2012; 945-953
- [41] Shi Weisong, Dustdar S. The promise of edge computing [J]. Computer, 2016, 49(5): 78-81
- [42] Bonomi F, Milito R, Zhu Jiang, et al. Fog computing and its role in the Internet of things [C] //Proc of the 1st Edition of the MCC Workshop on Mobile Cloud Computing. New York: ACM, 2012; 13-16
- [43] Zhang Quan, Zhang Xiaohong, Zhang Qingyang, et al. Firework: Big data sharing and processing in collaborative edge environment [C] //Proc of IEEE/ACM Symp on Edge Computing (SEC2016). Piscataway, NJ: IEEE, 2016; 81-82
- [44] Zhang L, Estrin D, Burke J, et al. Named data networking (NDN) project [J]. Transportation Research Record Journal of the Transportation Research Board, 2010, 1892(1): 227-234
- [45] Raychaudhuri D, Nagaraja K, Venkataramani A. MobilityFirst: A robust and trustworthy mobility-centric architecture for the future Internet [J]. ACM Sigmobile Mobile Computing & Communications Review, 2012, 16(3): 2-13
- [46] Feld J. PROFINET-scalable factory communication for all applications [C] //Proc of the 6th IEEE Int Workshop on Factory Communication Systems. Piscataway, NJ: IEEE, 2004; 33-38

- [47] Cao Jie, Ren Lingmei, Shi Weisong, et al. A framework for component selection in collaborative sensing application development [C] //Proc of 2014 Int Conf on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom). Piscataway, NJ: IEEE, 2014: 104-113
- [48] DaCosta F. Rethinking the Internet of Things: A scalable approach to connecting everything [M]. Berkeley: Apress, 2013
- [49] Perla C B. WiFi network security statistics/graph [OL]. [2016-12-03]. <http://graphs.net/wifi-stats.html/>
- [50] Open mHealth. Open mHealth platform [OL]. [2016-12-07]. <http://www.openmhealth.org/>
- [51] Brewer E. Inktomi's wild ride—A personal view of the Internet bubble [OL]. [2016-12-07]. <https://curiosity.com/videos/inktomis-wild-ride-a-personal-view-of-the-internet-bubble-computer-history-museum/>
- [52] Fox A, Brewer E A. Harvest, yield, and scalable tolerant systems [C] //Proc of the 7th Workshop on Hot Topics in Operating Systems. Piscataway, NJ: IEEE, 1999: 174-178
- [53] Brewer E A. Towards robust distributed systems [OL]. [2016-12-03]. http://awoc.wolski.fi/dlib/big-data/Brewer_podc_keynote_2000.pdf
- [54] Gilbert S, Lynch N. Brewer's conjecture and the feasibility of consistent, available, partition-tolerant Web services [J]. ACM Sigact News, 2002, 33(2): 51-59
- [55] O'Brien J A, Marakas G M. Management Information Systems [M]. New York: McGraw-Hill Irwin, 2008: 185-189
- [56] The Institute for Telecommunication Sciences (ITS). Distributed database [OL]. [2016-12-07]. http://www.its.blrdoc.gov/fs-1037/dir-012/_1750.htm
- [57] Liu Ling, Özsu M T. Homogeneous Distributed Database Systems [M]. Berlin: Springer, 2009
- [58] Sujansky W. Heterogeneous database integration in biomedicine [J]. Journal of Biomedical Informatics, 2001, 34(4): 285-298
- [59] Grolinger K, Higashino W A, Tiwari A, et al. Data management in cloud environments: NoSQL and NewSQL data stores [J]. Journal of Cloud Computing, 2013, 2(1): 1-24
- [60] Stonebraker M. NewSQL: An alternative to NoSQL and old SQL for new OLTP apps [OL]. [2016-12-07]. <http://cacm.acm.org/blogs/blog-cacm/109710-new-sql-an-alternative-to-nosql-and-old-sql-for-new-oltp-apps/fulltext>
- [61] Highscalability. Google spanner's most surprising revelation: NoSQL is out and NewSQL is in [OL]. [2016-12-07]. <http://highscalability.com/blog/2012/9/24/google-spanners-most-surprising-revelation-nosql-is-out-and.html>
- [62] Nirmala S, Venkatesh P. NewSQL—The new way to handle big data [OL]. [2016-12-07]. <http://opensourceforu.com/2012/01/newsql-handle-big-data/>
- [63] Microsoft Corporation. TechNet Glossary [OL]. [2016-12-07]. <https://technet.microsoft.com/en-us/library/cc966484.aspx>
- [64] Ashdown L, Kyte T, Engsig B, et al. Oracle® database concepts 11g release 2 (11.2) [OL]. [2016-12-07]. http://docs.oracle.com/cd/E36909_01/server.1111/e25789.pdf
- [65] Jagadish H V, Al-Khalifa S, Chapman A, et al. TIMBER: A native XML database [J]. The VLDB Journal, 2002, 11(4): 274-291
- [66] Fernandez I. No! to SQL and No! to NoSQL [OL]. [2016-12-07]. <https://iggyfernandez.wordpress.com/2013/07/28/no-to-sql-and-no-to-nosql/>
- [67] Milojicic D S, Kalogeraki V, Lukose R, et al. Peer-to-peer computing [OL]. [2016-12-03]. <https://www.cs.kau.se/cs/education/courses/dvad02/p2/seminar4/Papers/HPL-2002-57R1.pdf>
- [68] Peng Gang. CDN: Content distribution network [OL]. [2016-12-03]. <https://arxiv.org/pdf/cs/0411069.pdf>
- [69] Rabinovich M, Xiao Zhen, Aggarwal A. Computing on the edge: A platform for replicating internet applications [C] //Proc of the 8th Int Conf on Web Content Caching and Distribution. Berlin: Springer, 2004: 57-77
- [70] Akamai Technologies. CDN Learning Center [OL]. [2016-12-03]. <https://www.akamai.com/us/en/cdn/>
- [71] Amazon AWS. Content Delivery Networks (CDN) [OL]. [2016-12-03]. https://www.amazonaws.cn/en/content-delivery/?nc1=h_ls
- [72] Akamai Technologies. China CDN [OL]. [2016-12-03]. <https://www.akamai.com/us/en/solutions/intelligent-platform/china-content-delivery-network.jsp>
- [73] Hu Wen, Wang Zhi, Ma Ming, et al. Edge video CDN: A WiFi content hotspot solution [J]. Journal of Computer Science and Technology, 2016, 31(6): 1072-1086
- [74] ChinaCache. Webluker CDN [OL]. [2016-12-03]. <http://www.chinacache.com/>
- [75] Chinacenter. CDN Acceleration Technologies [OL]. [2016-12-03]. <http://chinanetcenter.com/>
- [76] European Telecommunication Standards Institute (ETSI). Executive Briefing-Mobile Edge Computing (MEC) Initiative [OL]. [2016-12-03]. <https://portal.etsi.org/Portals/0/TBpages/MEC/Docs/MEC%20Executive%20Brief%20v1%2028-09-14.pdf>
- [77] European Telecommunication Standards Institute (ETSI). Mobile Edge Computing [OL]. [2016-12-03]. <http://www.etsi.org/technologies-clusters/technologies/mobile-edge-computing>
- [78] Zhao Tianchu, Zhou Sheng, Guo Xueying, et al. A cooperative scheduling scheme of local cloud and Internet cloud for delay-aware mobile cloud computing [C] //Proc of 2015 IEEE GLOBECOM Workshops. Piscataway, NJ: IEEE, 2015: 1-6

[79] Sarukkai R R, Mendhekar A. Method and apparatus for accessing targeted, personalized voice/audio Web content through wireless devices; US Patent 6728731 [P]. 2004; 4-27

[80] Hu Yunchao, Patel M, Sabella D, et al. Mobile edge computing; A key technology towards 5G [OL]. 2015 [2016-12-06]. http://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp11_mec_a_key_technology_towards_5g.pdf

[81] Mao Yuyi, Zhang Jun, Letaief K B. Dynamic computation offloading for mobile-edge computing with energy harvesting devices [J]. IEEE Journal of Solid-State Circuits, 2016, 51 (3): 712-723

[82] Chen Xu, Jiao Lei, Li Wenzhong, et al. Efficient multi-user computation offloading for mobile-edge cloud computing [J]. IEEE/ACM Trans on Networking, 2016, 24(5): 2795-2808

[83] Wang Yanting, Sheng Min, Wang Xijun, et al. Mobile-edge computing; Partial computation offloading using dynamic voltage scaling [J]. IEEE Trans on Communications, 2016, 64(10): 4268-4282

[84] Haubenwaller A M, Vandikas K. Computations on the edge in the Internet of things [J]. Procedia Computer Science, 2015, 52(C): 29-34

[85] Ahmed A, Ahmed E. A survey on mobile edge computing [C] //Proc of the 10th IEEE Int Conf on Intelligent Systems and Control. Piscataway, NJ: IEEE, 2016; 1-8

[86] European Telecommunication Standards Institute (ETSI). Mobile-Edge Computing (MEC): Service Scenarios [OL]. [2016-12-03]. http://www.etsi.org/deliver/etsi_gs/MEC-IEG/001_099/004/01.01.01_60/gs_MEC-IEG004v010101p.pdf

[87] Vaquero L M, Rodero-Merino L. Finding your way in the fog; Towards a comprehensive definition of fog computing [J]. ACM SIGCOMM Computer Communication Review, 2014, 44(5): 27-32

[88] Luan T H, Gao Longxiang, Li Zhi, et al. Fog computing: Focusing on mobile users at the edge [OL]. [2016-12-07]. <https://arxiv.org/pdf/1502.01815.pdf>

[89] Klas G I. Fog computing and mobile edge cloud gain momentum open fog consortium, ETSI MEC and cloudlets [J/OL]. [2016-12-03]. http://yucianga.info/wp-content/uploads/2015/11/15_11_22_Fog_computing_and_mobile_edge_cloud_gain_momentum_Open_Fog_Consortium-ETSI-MEC-Cloudlets_v1_1.pdf

[90] Greyhound. Greyhound launches “BLUE?”—An exclusive WiFi enabled onboard entertainment system [OL]. [2016-12-03]. <http://www.prnewswire.com/news-releases/greyhound-launches-blue-an-exclusive-wi-fi-enabled-onboard-entertainment-system-214583791.html>

[91] OpenFog Consortium Architecture Working Group. OpenFog architecture overview [OL]. [2016-12-03]. <http://www.openfogconsortium.org/wp-content/uploads/OpenFog-Architecture-Overview-WP-2-2016.pdf>

[92] Jiang Mianheng. Urbanization meets informatization; A great opportunity for China’s development [J]. Informatization Construction, 2010 (6): 8-9 (in Chinese)
(江绵恒. 城市化与信息化——中国发展的时代机遇[J]. 信息化建设, 2010 (6): 8-9)



Shi Weisong, born in 1974. Professor, PhD supervisor. IEEE Fellow, Charles H. Gershenson Distinguished Faculty Fellow, ACM Distinguished Scientist. His main research interests include edge computing, computer systems, and energy-efficiency.



Sun Hui, born in 1983. PhD, assistant professor. His main research interests include computer system, edge computing, performance evaluation, non-volatile memory-based storage systems, file systems, and I/O architectures.



Cao Jie, born in 1989. PhD candidate at Wayne State University, USA. His main research interests include edge computing, smart home, and wireless health.



Zhang Quan, born in 1988. PhD candidate at Wayne State University. His main research interests include edge computing, big data system, stream processing, and energy-efficiency.



Liu Wei, born in 1993. Mater candidate in Anhui University. His main research interests include computer systems, key-value storage system.