基于宽带电力线通信的大数据分析在用电领域的应用

如何有效采集电力系统在各环节的数据，并使用大数据技术加以分析和应用，是当前物联网、云计算和大数据背景下电力行业信息化、智能电网的重要研究课题。重点论述了用电领域大数据的采集传输和分析应用。首先论述了用电信息采集关键通信技术的挑战和发展；随后分析了大数据存储与计算技术；最后通过一个具体用电量分析应用，测试对比了用电大数据的性能。

齐火箭　徐海宾　王嘉宏　杨红欣/国网冀北电力有限公司张家口供电公司

齐火箭/高级工程师

关键词/Keywords
电力线载波通信·
分布式存储·
集群计算·
大数据·
用电分析·

电力用户用电信息采集系统经过多年的建设和运营，已经基本满足用户表底的抄读和缴费等基本业务。随着物联网、云计算和大数据等信息技术的日趋成熟，如何让集抄系统采集更多用户侧数据、并利用大数据技术挖掘出更多价值和信息以支撑电网的经营和决策，成为下个阶段的研究热点。在此过程中，本地通信信道的带宽和实时性、主站的计算性能和海量用电数据的可靠存储，均是技术上的关键点。

国家电网在近4年统一集采2.3亿支电表，其中65%左右为窄带电力线载波。窄带载波频带主要分布在3～500 kHz，存在很多应用上的问题。首先，窄带一般以单载波为主，带宽较窄，易与现场噪声的主要能量频段重合，同时对时变噪声、脉冲干扰的抵抗能力差，导致通信性能不稳定；其次，窄带数据传输速率低，在几十Kbit/s以下，仅能支持简单的抄表业务；再次，窄带的路由变化自适应性差，导致长时间运行抄通率下降。为解决窄带方案的上述问题，基于宽带电力线载波的解决方案应运而生。宽带设备一般工作于1.8～250 MHz，传输速率从几Mbit/s到几Gbit/s。当前主流的宽带标准主要包括HomePlug AV/GP、HD－PLC、IEEE 1901和ITU－T G.hn等。

国内电力线信道特征

电力线并不是专用的通信线路，其作为数据通信的传输介质，具有噪声大、干扰强和衰减剧烈等特点。通过分析电力线信道实测数据及经典信道模型，获得国内电力线信道的特征，是宽带通信方案设计的基础。

首先，电力线的输入阻抗直接影响了信号耦合的效率，是低压电力线传输特性的重要参数。理想情况下，电力线可以视为一条各项特性均匀分布的传输线，其输入阻抗应随频率的增大而减小。然而由于实际电力线环境中，大量负载的随机接入及用电设备的起停，导致电力线的输入阻抗体现为随时间和频率变化。进一步地，输入阻抗随频率和负载的变化会影响输出功率。研究表明，在2 MHz以下频段，电力线的输入阻抗较小，发射功率效率低下，不利于信号传输；同时，电力线的输入阻抗具有明显的频率选择性。

其次，电力线信道具有较强的衰减，并且衰减强度表现出时变性。信道衰减主要包括耦合衰减和线路衰减。耦合衰减由发射端和接收端电力线阻抗不匹配导致，大小约为10 dB。线路衰减来自多径传播、线路损耗及延时，配电变压器的衰减损耗可到几十至上百dB。根

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

据Zimmermann等人提出的通用多径传输模型为

$$H(f)=\sum_{i=1}^{N}\left|g_i(f)\right|\mathrm{e}^{\varphi_{gi}(f)}\mathrm{e}^{-(\alpha_0+\alpha_1 f^k)d_i}\mathrm{e}^{-\mathrm{j}2\pi f\tau_i} \quad (1)$$

式中，i为路径标号；N为总路径数目；$\left|g_i(f)\right|\mathrm{e}^{\varphi_{gi}(f)}$为第$i$路的加权因子；$\alpha_0$、$\alpha_1$为衰减参数；$k$为衰减因子的指数；$d_i$为第$i$路的路径长度；$\tau_i$为第$i$路的路径延迟。

如图1所示，电力线信道随频率和距离的增加而衰减，在20 MHz频率范围内的信道传输特性很好地符合理论模型。

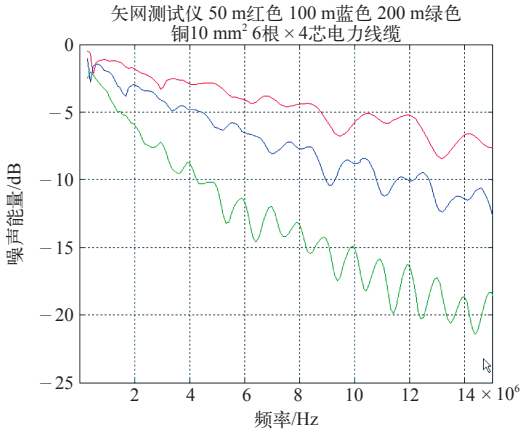

图1　信道衰减随频率变化情况

最后，电力线通信环境存在各种时变噪声。电力线信道的噪声主要包括有色背景噪声、窄带噪声、与工频异步的周期性脉冲噪声、与工频同步的周期性脉冲噪声和异步脉冲噪声。

白炽灯噪声脉冲周期为10 ms，属于与工频同步的周期性脉冲噪声。在频域上，其能量主要集中在1 MHz以下，噪声能量随频率减小而增大，500 kHz附近的噪声能量比2 MHz以上频率处的噪声能量高15 dB。

电动车充电噪声的时频特性与白炽灯相似，属于与工频同步的周期性脉冲噪声。但其在频域上体现为周期性梳状谐波，频率间隔约为50 kHz，影响范围为几十kHz到十几MHz。

小区变频水泵噪声属于与工频异步的周期性脉冲噪声，周期为30～80 μs，1 MHz以下的噪声能量比2 MHz以上的噪声能量高20 dB。

电梯运行的起停是随机的，因此其噪声属于异步脉冲噪声。500 kHz以下及1～2.5 MHz频率范围内噪声能量大。

总体上，电力线信道噪声在2 MHz以下频率范围内能量集中，然而某些用电设备也会引起更高频率范围内的周期性谐波以及高能量。

宽带电力线通信方案

通过研究电力线信道模型，表明2 MHz以下电力线阻抗小，信道噪声能量大，不适合信号传输。因此，物理层采用2～12 MHz进行数据通信，可适应国内电力线信道特征。图2为OFDM技术抵抗干扰和衰落示意图。

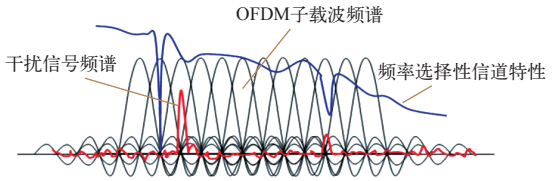

图2　OFDM技术抵抗干扰和衰落示意图

宽带电力线载波方案采用正交频分复用（OFDM）技术，在保证频谱利用率的基础上，具有更强的对抗脉冲噪声与信道快衰落的能力。OFDM技术将多个子载波上的数据并行传输，提高符号持续时间，同时加入循环前缀结构，抗码间干扰能力增强。

为对抗电力线信道中的强噪声和强干扰，宽带电力线载波方案提出更为优化的物理层帧结构，如图3所示。帧结构中的Preamble采用13个1024长度的OFDM符号并提高采样率，使以

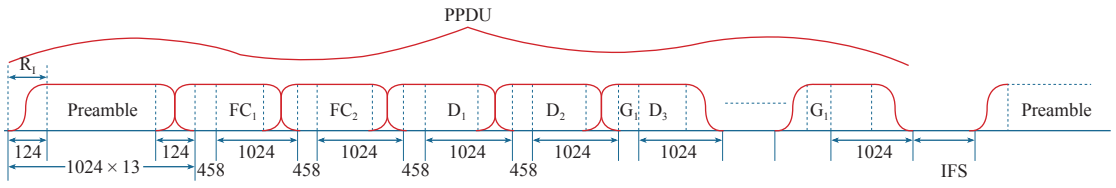

图3　宽带电力线通信物理层帧结构

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

Preamble为基础的信道估计、定时同步和信噪比估计等算法获得充分的数据信息，提升通信性能。帧控制信息存储于FC中，对于负载数据的译码具有重要作用，因此FC进行多次分集拷贝，有效提高抵抗电力线脉冲干扰的能力。

分集拷贝技术在频域和时域上均可以实现。为了抵抗频域上脉冲干扰，对子载波进行拷贝，避免子载波信息被完全破坏。同样地，在时域上对OFDM符号进行周期拷贝发送，可以有效抵抗时域随机脉冲干扰。对于频域上体现为周期谐波的脉冲噪声，采用合理的子载波分配方法，将数据置于脉冲间隔内，使数据传输不受该脉冲的影响。

根据信道情况，FC可以配置为不同调制方式，如BPSK或QPSK，采用低阶调制对抗恶劣信道环境。对于某一频段固定存在的窄带干扰，物理层自适应地在2～12 MHz频段范围内选择合适的频段进行传输。

综上所述，宽带电力线通信方案采用多种物理层策略，可以保证数据的可靠传输。在用电信息采集系统中，宽带方案数据传输速率最高可达2 Mbit/s，完全满足电力信息采集系统的业务需求。

大数据技术

大数据技术要面对的基本问题是海量数据如何可靠存储和高效计算。Appach的Hadoop是用Java语言实现的开源软件框架，是一个开发和运行处理大规模数据的软件平台，实现在集群中对海量数据进行分布式计算。Hadoop框架核心是HDFS和MapReduce。

1. 大数据存储技术

图4为HDFS结构图。HDFS属于分布式文件系统的一种，数据存储在集群中多个节点上，同时具有备份功能，当集群中某些节点过载或者出错失效时，可查找和使用备份节点数据，增强数据的安全性。分布式文件系统可存储达到PB级别的数据内容，这是传统关系型数据库所不能达到的。

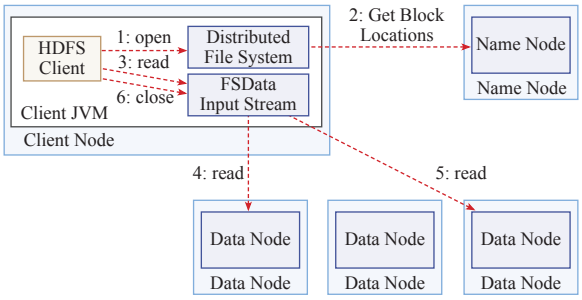

图4 HDFS结构图

HDFS具有高可靠性、高扩展性和自动的负载均衡能力。主要是由NameNode和若干DataNode组成。HDFS将文件的Block分配信息存放在NameNode服务器之上，各文件Block分布地存放在DataNode服务器上。当整个系统需要扩充容量时，只需要增加DataNode的数量，系统会自动地实时将新的服务器匹配进整体阵列之中。之后，文件的分布算法会将数据块搬迁到新的DataNode之中，不需任何人工维护或干预。HDFS假设系统故障（服务器、网络和存储等）是常态，而不是异常。故障检测和自动快速恢复是HDFS设计的一个核心目标。

2. 大数据计算技术

MapReduce提供了一个开源集群的计算环境，将计算过程的并发、分布式和容错处理等细节隐藏起来，使用户仅关注自己业务逻辑的实现，大大简化了大数据挖掘的实现复杂度。图5为MapReduce处理过程。

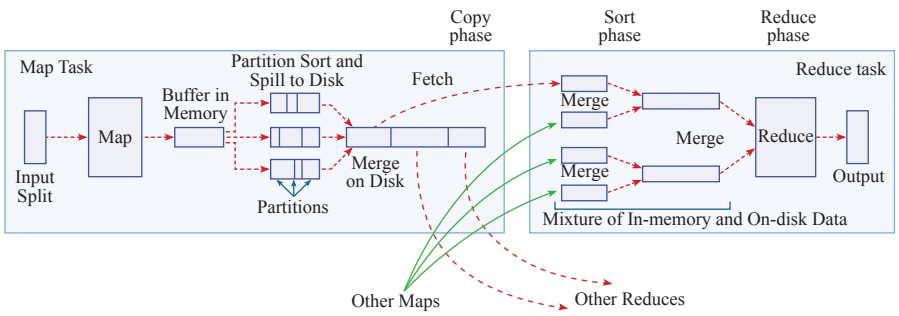

图5 MapReduce处理过程

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

用电大数据分析应用

1. 电力线载波性能测试对比

智能电表可以采集上百种数据项，包括各种电压、电流等实时量，正向有功、反向有功等冻结量，以及开表盖、掉电等事件量。用电末端采集数据的丰富性，凸显了本地信道可靠性和速率的瓶颈。通信可靠性方面，从目前的窄带电力线载波台区看，抽样平均成功率在80%左右，导致一些数据缺失。宽带电力线载波采集成功率可达99.9%以上，能保证基础数据的完整性。通信速率方面，窄带电力线载波1个数据项抄读用时平均在4 s左右，宽带电力线载波1个数据项抄读用时平均小于0.4 s。如果充分利用宽带的带宽，使用并发技术，每个数据项抄读平均用时小于0.1 s，速度非常快，且随着并发数的增加，速度会成倍增加。通过在不同规模台区实际测试对比，性能数据如表1所示。

表 1　宽窄带电力线载波抄表性能对比

	窄带电力线载波			宽带电力线载波			宽带电力线载波并发		
	实测数据	1 h采集	0.5 h采集	实测数据	1 h采集	0.5 h采集	实测数据	1 h采集	0.5 h采集
1户台区 1个数据项	4 s	满足	满足	<0.4 s	满足	满足	<0.1 s	满足	满足
100户台区 3个数据项	18 min 10 s	满足	满足	1 min 23 s	满足	满足	9 s	满足	满足
300户台区 3个数据项	40 min 37 s	满足	不满足	4 min 36 s	满足	满足	51 s	满足	满足
500户台区 3个数据项	86 min 21 s	不满足	不满足	7 min 47 s	满足	满足	1 min 24 s	满足	满足
800户台区 3个数据项	169 min	不满足	不满足	13 min 20 s	满足	满足	2 min 19 s	满足	满足
1 000户台区 3个数据项	207 min 0 s	不满足	不满足	16 min	满足	满足	3 min 46 s	满足	满足

从数据生产及消耗角度分析，数据的消耗速率必须大于生产速率，否则将发生数据拥塞，进而丢失，使数据的分析产生错误。从实际的测试数据对比看，台区规模到达300左右时，各类0.5 h采集数据产生的速率已经超过窄带的抄读速率。而宽带载波、特别是加上并发技术，即便是在1 000规模台区时，数据的传输速率也远远胜任数据采集量的需要。

2. 用电大数据的系统架构

用电大数据系统可在原用电信息采集系统之上，做增量式扩充和改造。如图6所示，在采集与控制层，即传感器网络，将本地信道更换为宽带电线载波，可满足大量采集数据的传输需求。在数据管理层，扩展分布式文件系统，用于存储大量台区现场的实时数据。在业务逻辑层，扩展Hadoop分布式计算环境，各种用电大数据分析业务的核心计算逻辑部署于此，原有主站的采集任务管理、费控管理和事件上报等业务逻辑不变。在业务呈现层，新增基于Android、iOS和微信的移动终端，便于现场的运维管理，便于与电力用户的双向互动。

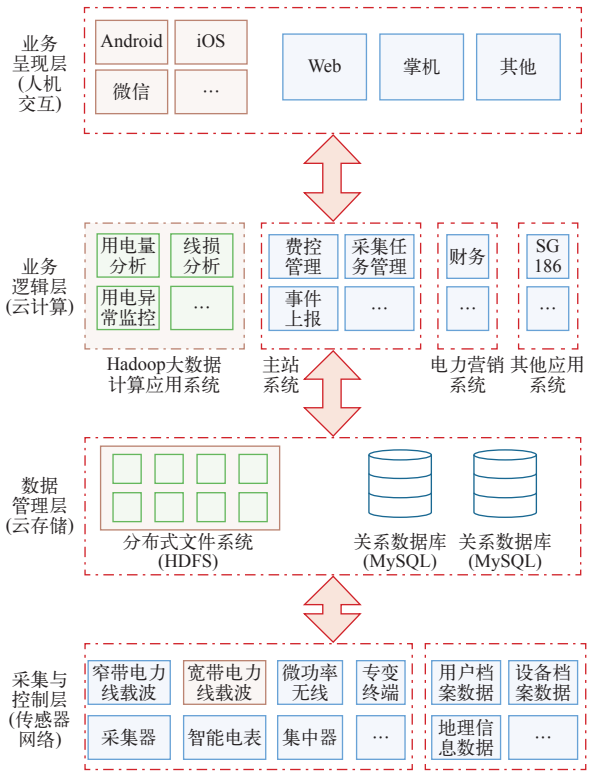

图6　电力用户用电信息采集系统大数据扩展架构

3. 用电大数据分析应用

宽带电力线载波技术目前已经成功在各省网

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

公司有一定规模试点和应用，参照大数据系统架构，搭建了试验性质的HDFS和MapReduce集群，并开发了基于Android的客户端，实现了用户用电量大数据分析。

以浙江省为例，140万台采集终端，覆盖约3 400万用户，每次产生数据30 GB左右，系统2006年建成以来，累积数据已达40 TB左右。此类用电大数据系统在国际上已经有成功的应用典范。美国Opower是一家家庭能源数据分析公司，通过与电网公司合作，获得用户消费能源的原始数据，使用基于Hadoop的大数据分析系统，向用户提供详细账单，用户则会更有意识地去节约用电。下面简述类似系统的实现。

定义用电大数据在HDFS中的存储管理方式。以省网公司、年份和区县名称组成文件的存储路径：省网名称代号\年份\地市名称代号\区县名称代号\台区代号，如：浙江省\2014\3350_嘉兴市\3355_海宁市\123456.data。每个台区以年为单位对应一个文件，存储台区所有用户全年用电量数据。123456.data的文件格式如下，以文本形式存储，每条记录用户全年日用电量。表2为某台区用户全年用电数据。

利用MapReduce的编程模型，计算实现某台区某时间段内的用电量走势曲线图，如图7所示。

表2 某台区用户全年用电数据

表号	20140101	20140102	…	20140630	20140701	…	20141231
100000001234	0045	0057	…	0089	0097	…	0051
100000005678	0020	0023	…	0040	0047	…	0030
…	…	…	…	…	…	…	…
1000000006789	0056	0055	…	0090	0102	…	0060

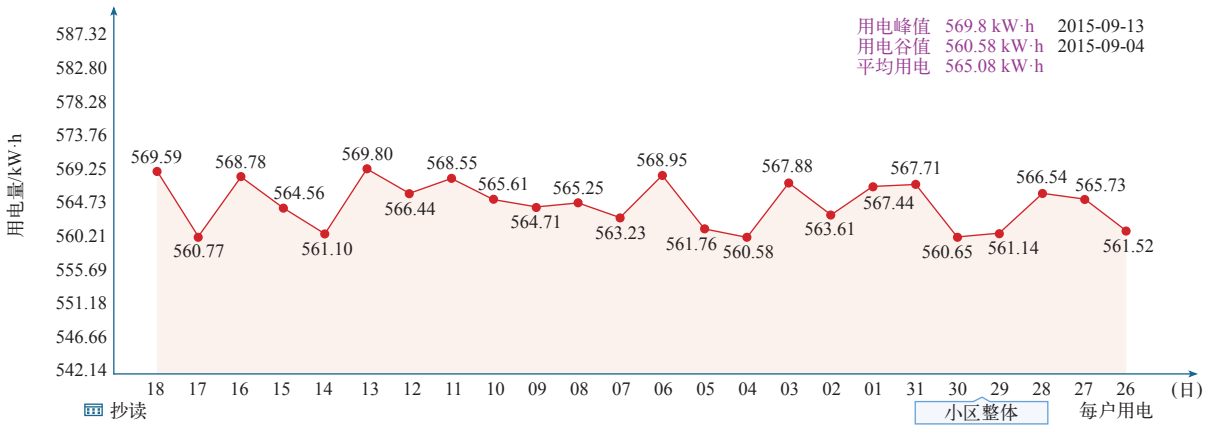

图7 某台区日用电曲线

为了测试对比计算性能，加大了文件的存储条目，与MySQL对比查询性能。测试环境单台PC配置为8核2.4 GHz CPU、4 G内存、512 G硬盘，集群为10台同配置PC组成。从表3所示的测试结果可以看出，240万条前单机系统具有优势，这是因为分布系统自身的调动也需要额外开销。随着数据量持续增大，集群的优势越发明显。

表3 查询时间对比

数据条目数	MySQL/s	Hadoop/s
10万	4.7	11.4
30万	13.8	23.7
60万	18.2	24.3
120万	22.5	28.8
240万	62	59.4
500万	136.1	88.9
1 000万	245.6	161
2 000万	512.5	286.7

结束语

用电信息采集系统发展演进的两个重要特点是本地通信宽带载波化和主站系统分布式化。这

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

是采集系统在大数据应用发展的必然要求。宽带电力线载波，在抗干扰和速率上的优势，为海量用电数据的采集传输提供了成熟的CT支撑。主站系统的分布式扩展，为海量用电数据的可靠存储和高性能计算，提供了成熟的IT支撑。基于此，各种用电大数据分析应用正在逐步落地，并呈现其特有价值。

参考文献

[1] Tome White．Hadoop 权威指南 [M]．2版．北京：清华大学出版社，2011．

[2] 刘树仁，宋亚奇，朱永利，等．基于Hadoop的智能电网状态监测数据储存研究 [J]．计算机科学，2013，40 (1)：81-84．

[3] 陆嘉恒．Hadoop 实战 [M]．2版．北京：机械工业出版社，2013．

[4] 郑启龙，房明，汪胜，等．基于MapReduce模型的并行科学计算 [J]．微电子学与计算机，2009，28 (6)：13-17．

[5] Han J W，Kamber M，Pei J．Data mining：concepts and techniques [M]．SanFrancisco：Morgan Kaufmann—Publishers，2011．

[6] 张文亮，刘壮志，王明俊．智能电网的研究进展及发展趋势 [J]．电网技术，2009，33 (13)：1-11．

[7] 李国杰．大数据研究的科学价值 [J]．中国计算机学会通信，2012 (9)：8-15．

[8] Xavier Carcelle．Power line communication in practice [M]．Boston：Artech House Publisher，2011．

（收稿日期：2015－09－15）EA

(上接第151页)

系统自动化，2013，37 (4)：1-6．

[6] 赵峰，孙宏斌，谭嫣．综合考虑多种电网安全主题的关键断面自动发现方法 [J]．电网技术，2014，38 (5)：1169-1174．

[7] 赵峰，孙宏斌，张伯明．基于电气分区的输电断面及其自动发现 [J]．电力系统自动化，2011，35 (5)：42-46．

[8] 周德才，张保会，姚峰，等．基于图论的输电断面快速搜索 [J]．中国电机工程学报，2006，26 (12)：34-41．

[9] 张保会，姚峰，周德才，等．输电断面安全性保护及其关键技术研究 [J]．中国电机工程学报，2006，26 (21)：1-7．

[10] 姚峰，张保会，周德才．输电断面有功安全性保护及其快速算法 [J]．中国电机工程学报，2006，26 (13)：31-36．

[11] 程临燕，张保会，郝治国，等．基于线路功率组成的关键输电断面快速搜索 [J]．中国电机工程学报，2010，30 (10)：50-56．

[12] 任建文，李刚，王增平，等．基于背离路径的输电断面搜索新算法 [J]．电网技术，2012，36 (4)：121-127．

[13] 王增平，李刚，任建文，等．基于前K最短路径的输电断面搜索新算法 [J]．电工技术学报，2012，27 (4)：193-201．

[14] 魏萍，倪以信，吴复立，等．基于图论的输电线路功率组成和发电机与负荷间功率输送关系的快速分析 [J]．中国电机工程学报，2000，20 (6)：21-25．

[15] 杨文辉，毕天姝，薛安成．基于图论的潮流转移路径的快速搜索 [J]．电网技术，2012，36 (4)：84-88．

（收稿日期：2015－06－08）EA

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net