

6-1-2020

Parallel Method for Extracting Pulses from Multi-source Massive Partial Discharge Signals

Liuwang Wang

School of Electrical and Electronic Engineering, North China Electric Power University, Baoding 071003, China;

Yongli Zhu

School of Electrical and Electronic Engineering, North China Electric Power University, Baoding 071003, China;

Yafei Jia

School of Electrical and Electronic Engineering, North China Electric Power University, Baoding 071003, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Parallel Method for Extracting Pulses from Multi-source Massive Partial Discharge Signals

Abstract

Abstract: Aiming at the issue of discharge pulse extraction for multi-source and massive PD signals, a novel parallel method based on Message Passing Interface was proposed. *The proposed method applied a parallel mode called manager-worker-writer. In this method, a manager dynamically assigned task to several workers, and these workers executed tasks in parallel and a writer received results from workers in real time, so data management was separated from task execution. In addition, the manager identified sources of PD signals and sent them to workers as the keys for analyzing different data files and setting algorithm parameters, so multi-source and heterogeneous PD signals can be processed in parallel.* Experimental results show that the proposed parallel method is efficient, feasible and possesses good speed-up ratio. It achieves total extraction accuracy of above 90%, which can satisfy most engineering applications.

Keywords

partial discharge, message passing interface(MPI), parallel computing, big data, pulses extraction, data processing

Recommended Citation

Wang Liuwang, Zhu Yongli, Jia Yafei. Parallel Method for Extracting Pulses from Multi-source Massive Partial Discharge Signals[J]. Journal of System Simulation, 2017, 29(1): 57-66.

一种多源海量局部放电信号脉冲的并行提取方法

王刘旺, 朱永利, 贾亚飞

(华北电力大学电气与电子工程学院, 河北 保定 071003)

摘要: 针对多源、海量局部放电(Partial Discharge, PD)信号的放电脉冲提取问题, 提出了一种基于消息传递接口(Message Passing Interface, MPI)的并行化方法。该方法采用管理者-工人-写者模式, 由管理者动态分配任务, 工人负责业务计算并由写者回收计算结果, 将数据管理和任务执行分离。此外, 由管理者辨别信号来源并发送给工人作为数据文件解析和算法参数设置的依据, 能够解决多源异构信号处理问题。实验结果表明所提方法高效可行, 且具有良好的加速比, 脉冲提取准确率达 90%以上, 满足工程应用需求。

关键词: 局部放电; 消息传递接口; 并行计算; 大数据; 脉冲提取; 数据处理

中图分类号: TP391.9 文献标识码: A 文章编号: 1004-731X (2017) 01-0057-11

DOI: 10.16182/j.issn1004731x.joss.201701009

Parallel Method for Extracting Pulses from Multi-source Massive Partial Discharge Signals

Wang Liuwang, Zhu Yongli, Jia Yafei

(School of Electrical and Electronic Engineering, North China Electric Power University, Baoding 071003, China)

Abstract: Aiming at the issue of discharge pulse extraction for multi-source and massive PD signals, a novel parallel method based on Message Passing Interface was proposed. *The proposed method applied a parallel mode called manager-worker-writer. In this method, a manager dynamically assigned task to several workers, and these workers executed tasks in parallel and a writer received results from workers in real time, so data management was separated from task execution. In addition, the manager identified sources of PD signals and sent them to workers as the keys for analyzing different data files and setting algorithm parameters, so multi-source and heterogeneous PD signals can be processed in parallel.* Experimental results show that the proposed parallel method is efficient, feasible and possesses good speed-up ratio. It achieves total extraction accuracy of above 90%, which can satisfy most engineering applications.

Keywords: partial discharge; message passing interface(MPI); parallel computing; big data; pulses extraction; data processing

引言

随着智能电网建设的不断深入, 输变电设备状

态监测技术逐渐在电力企业中得到推广和应用, 海量监测数据的快速处理面临巨大的技术挑战^[1]。已有的在线监测装置大多对采集数据就地处理后将“熟数据”上传, 这对于设备的状态分析和故障诊断作用十分有限。从在线监测技术的发展趋势而言, 弱化监测装置的存储与处理能力, 直接将采集数据通过高速光纤上传到基础设施完备的监测中心集



收稿日期: 2016-03-04 修回日期: 2016-04-26;
基金项目: 中央高校基本科研业务费专项资金
(2015XS106);

作者简介: 王刘旺(1988-), 男, 安徽安庆, 博士生, 研究方向为输变电设备状态监测大数据分析;
朱永利(1963-), 男, 河北衡水, 博士, 教授, 研究方向为人工智能及其应用, 电力调度自动化。

<http://www.china-simulation.com>

中处理, 更有利于数据管理与软件维护^[2-3]。

局部放电是高压设备绝缘劣化的主要原因和重要表征, 对于设备的绝缘性能评价和故障诊断极为重要, 其在线监测数据随着传感器数量增多与采样频率提高而日益庞大, 大量高压设备产生的不同存储结构的局部放电信号呈现出多源海量异构特征, 如何对其进行快速处理成为亟待解决的问题。然而长久以来, 为了从原始局部放电信号中获取反映放电特性的高价值信息, 国内外学者一直致力于局部放电信号的特征提取研究^[4-7], 但均没有考虑多源海量异构问题。现有方法中, 符合 IEC 60270 标准^[8]的局部放电相位分析(Phase Resolved Partial Discharge, PRPD)方法最为成熟, 且已被广泛应用于实际工程和局放分析仪等产品中。该方法只需要统计放电次数 n 、放电量 q 与所在相位 φ , 物理意义明确, 且不易受噪声影响, 在其基础上可以进行放电趋势分析、绘制放电谱图、模式识别等。PRPD 方法的关键是实现放电参数 φ - q - n 的有效提取, 即放电脉冲的提取, 目前已有相关工作^[9-13]。文献[9]定义了能量与瞬时曲率并通过阈值来确定单个脉冲起点和宽度来实现单个放电脉冲提取。文献[10]根据脉冲振荡特征结合双阈值进行逻辑判断的方法来确定放电脉冲起点和终点。文献[11]依据背景噪声和放电脉冲的时域能量差异来确定脉冲边沿。文献[12]基于数学形态学梯度和峭度指标来定位放电脉冲在工频周期内的位置。文献[13]采用当前 N 阶累积量(CNA)和阈值来建立 PRPD 模式。经笔者分析和编程测试, 文献[10]的方法在脉冲提取准确性上最为优越且易于编码实现。上述方法均不支持多源海量异构局部放电信号的分析需求, 且目前未见文献研究局部放电信号脉冲的并行提取方法。

近几年大数据分析问题在各个领域均得到重要关注, 以 MapReduce 模型为基础的云计算技术成为目前各行业大数据处理的主流方案^[14-18]。为了解决多源异构问题, 大多方案在进行数据处理之前构建统一的数据存储模型^[19-21]来集成、融合多源异构数据, 而数据集成与融合过程仍然避免不了对源

数据进行特定的解析。然而 MapReduce 模型为了追求编程接口简单化, 在隐藏数据划分、任务调度、通信与负载均衡等诸多底层细节的同时也失去了对它们的控制权, 将其用于处理多源异构数据时难度较大。MPI 是一种基于消息传递的高性能并行计算编程接口, 适合于粗粒度的进程级并行计算, 具有可移植性好、功能强大、效率高等优点, 几乎所有的并行计算机厂商都提供对它的支持, 应用十分广泛^[22-24]。MPI 的显式灵活的编程模式使得它在处理多源海量的局部放电信号时具有更大优势, 但目前尚无相关工作, 因此本文在文献[10]工作的基础上, 结合 MPI 并行编程技术, 针对局部放电脉冲提取问题提出一种面向多源海量信号的并行方法, 并通过实验验证了所提并行方法的有效性, 能够解决多源海量局部放电信号脉冲提取的计算瓶颈, 具有良好的工程应用价值。

1 局部放电信号脉冲提取

1.1 双阈值逻辑判断法

局部放电信号的脉冲提取是建立 PRPD 模式的基础, 如图 1 所示, 图中含有 a_e 和 f_g 两段局部放电脉冲, 均是衰减振荡的, 各自都包含了多个半周, 如何将这两次放电脉冲的相位 φ 和幅度 q 正确提取出来是问题关键。针对该问题, 文献[10]提出了一种基于脉冲波形特征进行逻辑判断的方法(本文简称为双阈值逻辑判断法), 该方法试图对波形中所有超过给定幅值 T_A 的半周进行逻辑判断(具体的逻辑判断规则参考文献[10]), 根据前后半周之间的距离关系来确定它们是否属于同一个脉冲, 以此达到提取脉冲的目的。原文只给出了逻辑判断规则, 对方法实现细节并未透露, 因此本节从编程角度给出具体实现步骤和细节, 为后续并行化提供基础。

步骤 1 输入待分析单周期放电信号序列 P_d , 给定幅值 T_A , 异极性阈值 T_1 , 同极性阈值 T_2 。其中 T_A 、 T_1 和 T_2 的值需要根据信号特征合理设置, 监测源相同的信号通常可以采用相同的参数值。

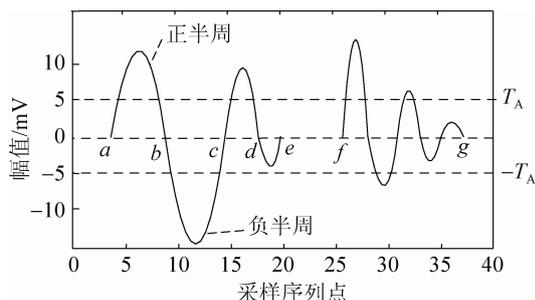


图 1 脉冲提取示意图

Fig.1 Diagram of pulse extraction

步骤 2 全局检测 P_d 中幅值绝对值超过 T_A 的点, 其下标序列记为 I_{above} 。

步骤 3 对 I_{above} 中每一个点搜索前后相邻过零点, 根据过零点是否相同确定超阈值点是否属于同个半周, 依次检测出所有半周, 并确定每个半周的局部峰点。半周的前后过零点和局部峰点下标序列分为记为 S_{start} , E_{end} , P_{peak} 。

步骤 4 对上步中所得所有半周根据逻辑判断规则(详见文献[10])依次按顺序识别是否属于同一个脉冲, 以此来确定 1 个完整放电脉冲的起点和终点。放电脉冲的起点、终点和放电峰点下标序列分别记为 S_{pulse} , E_{pulse} , P_{pulse} 。

步骤 5 根据 P_{pulse} 得到每次放电脉冲的幅值和相位一一对应分别存入序列 Q_{pulse} , I_{pulse} 做下一步分析或写入输出文件。

1.2 运算复杂度分析

基础运算包括加法(ADD)、乘法(MUL)、除法(DIV)和比较(COMP), 以下分别讨论各步骤的运算复杂度。设单周期局部放电信号的采样点数为 n , 步骤 1 和 5 的复杂度可以忽略不计。步骤 2 需要逐点与 T_A 进行比较, 需要进行 n 次 COMP。若步骤 2 中检测到的超阈值点个数为 n_1 , 则步骤 3 中需要对它们进行过零点检测, 而过零点检测时需要前后逐点搜索乘积为非正的点, 因此约需要 $n_1 \times (n_1/n_2)$ 次 MUL 和 COMP, 其中 n_2 为检测到的半周个数, 而 n_1/n_2 为平均每个半周含有的超阈值点个数。步骤 4 对 n_2 个半周按顺序两两进行逻辑判断, 根据逻辑判断规则(详见文献[10])可以估算需要的运算为 $(n_2-1) \times (MUL+2 \times COMP+2 \times ADD+DIV)$ 。

从上述分析可以看出运算复杂度和信号点数 n 、超阈值点个数 n_1 和半周个数 n_2 正相关, 由局部放电信号特点可以推断出同样采样时长下 n 与信号的采样频率正相关, 而 n_1 与放电密度(同时长下的放电次数)和采样频率同时正相关, n_2 与放电密度正相关。因此, 采样频率越高, 放电密度越大, 该方法的运算复杂度越高, 耗时越长。

假设局部放电信号采集系统的采样频率为 f , 工频为 50 Hz 时单个数据源每秒可以采集 50 个周期, 每分钟是 3 000 个周期。信号的采样频率 f 在实际工程和科研实验中均没有固定标准, 但至少在 5 MHz^[25]。对局部放电脉冲波形分辨率要求很高时, 采样频率甚至高达 100 MHz。取 $f=5$ MHz, 则单周期信号有 100 000 个采样点, 每分钟则是 $3\,000 \times 10^5 = 3 \times 10^8$ 个点。这仅仅是单个采集源 1 分钟内的采集量, 而目前高压电力设备配备的局部放电传感器逐渐增多, 例如文献[3]中的局部放电在线监测系统在两回线路的 147 个接头、12 个终端头处共计安装了 159 个局放采集单元。对于一个省网来说, 众多采集源在一定时间内积累的局部放电信号量会很可观, 这对传统的串行处理方法提出了挑战。

1.3 多源异构问题

针对局部放电信号分析, 一方面, 数据来源不同表明监测对象不同, 脉冲提取算法参数设置需要有针对性, 处理结果在汇总时也需要区分开, 这是多源问题; 另一方面, 监测装置规格不同、采样频率不同、数据存储格式不同等都使源数据文件结构不同, 给数据解析造成困难, 这是异构问题。导致数据异构的原因很多, 如计算机体系结构异构、操作系统异构、数据格式异构、数据语义异构等, 本文只考虑数据在存储结构上的异构问题。

多源异构问题同样给算法的并行实现带来难度, 异构的数据在解析处理时所耗的时间不一致, 这对算法并行时的负载平衡影响较大, 如何保证负载平衡是提高并行算法效率的关键之一。另一个关键是尽可能降低并行算法中不同工作节点或进程之间的通信开销(指的是进程或节点之间进行数据

传输时所需的时间), 否则会严重影响并行性能甚至失去并行意义。

2 基于 MPI 的脉冲提取并行方案

2.1 并行性分析

并行计算通常可分为数据并行和任务并行, 数据并行是将数据划分并分发到不同的计算节点进行并发计算, 而任务并行是将较大的任务分解成能够独立运行的子任务来同时进行。

由 1.1 节中的实现步骤可以看出, 每个步骤之间都是前后有序的衔接, 后序步骤的执行依赖于前序步骤, 很难进行任务分解, 因此任务并行行不通。虽然对于单周期的局部放电信号的脉冲提取不能再分解, 但每个周期之间的处理是独立不依赖的, 因此, 可以将信号数据分而治之, 分散到多个计算节点并行执行可以大大降低运行时间。如果计算节点足够多, 理论上可以将每个信号的脉冲提取同时进行, 可以显著减少批处理任务的处理时间。

2.2 并行方案及其实现

2.2.1 管理者-工人-写者模式

一般情况下待分析的数据量事先并不确定, 无法手动实现均等划分, 且均等划分方式在异构集群中很难保证负载平衡。为在程序执行期实现数据动态划分, 自动实现任务与工作进程之间的映射, 本文采用管理者-工人-写者模式, 主程序伪代码如下:

```
代码 1: int main(int argc, char** argv)
1: int rank; int size; //程序序号 rank, 进程总数 size
2: MPI_Init(&argc, &argv); //初始化 MPI
3: MPI_Comm_rank(MPI_COMM_WORLD, &rank);
4: MPI_Comm_size(MPI_COMM_WORLD, &size);
5: string dirname = argv[1]; //获取输入数据路径
6: if (size >= 3){
7:     if (rank == 0)
8:         manager(dirname, size); //0 号进程启动管理者
9:     else if (rank == 1)
10:        writer(dirname, size); //1 号进程启动写者
11:    else
12:        worker(dirname, rank); //其他进程启动工人
13: } //end if(size)
```

```
14: MPI_Finalize(); //释放分配给 MPI 的资源
15: return 0;
```

指定 0 号进程(称为管理者, manager)管理所有数据的划分, 负责将计算任务分配给其他进程(称为工人, worker)。指定 1 号进程(称为写者, writer)来负责从工人那里回收处理结果并做后续处理。

以 1 个工人为例, 管理者-工人-写者模式的工作结构如图 2 所示, 其中实线表示程序执行流程, 虚线表示进程之间的通信事件。主程序开始后, 每个进程都会进入循环, 不满足循环条件时会结束进程。当管理者没有数据分配给工人时, 会发送空任务消息给工人并更新活动的工人数目(图 2 中的变量 workers), 所有工人都结束后管理者才可结束。而工人收到空任务消息时表明已无任务, 自己可以结束, 并发送空结果告知写者。写者不停地从工人处接收结果, 每收到 1 个空结果时就更新活动的工人数目, 当所有工人都终止时表明所有计算结果均已接收, 这时写者结束。事实上, 图 2 给出了一种通用并行框架, 只要将图 2 中放电脉冲提取换成其他处理模块, 同时处理好各进程之间的消息传递事件即可。

2.2.2 多源异构问题解决方法

一般的, 局部放电信号数据文件是以来源标识与采集时间组合作为唯一标识, 这为解决数据多源异构问题提供了依据。为了解决多源异构问题, 需要根据局部放电监测源配置信息编制一张映射表, 表中的每一项为(key, 数据来源, 解析规则, 算法参数), 映射表会在每一个节点上保存副本, 并根据需要同步更新。管理者解析待分析数据文件的元信息得到数据来源, 通过映射表转换成 key 同任务数据一起发给工人。工人根据 key 查询映射表得到正确的解析规则和算法参数对异源异构数据文件做数据解析和脉冲提取处理。其中, 解析规则体现为不同的子程序, 算法参数指的是 1.1 节中的 T_A , T_1 和 T_2 。工人给写者发送计算结果时也要附带该 key, 写者可以对同一来源的计算结果进行后续操作。因此, 能够对多源异构文件进行并行分析。

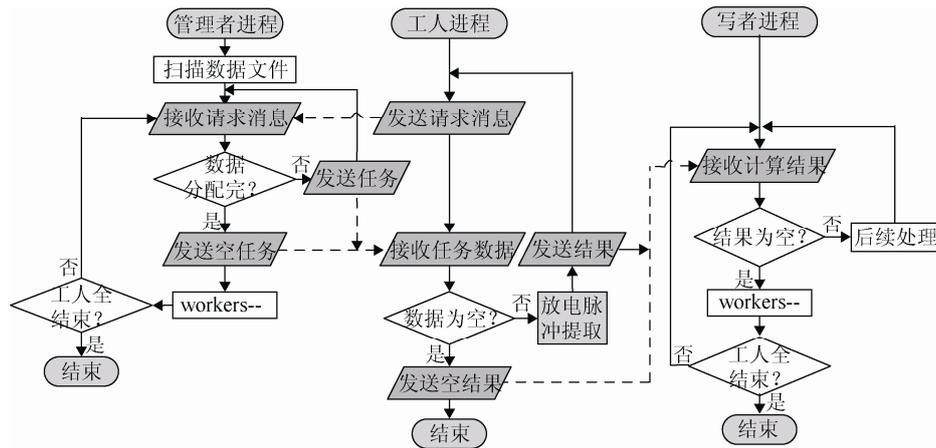


图 2 管理者-工人-写者模式的工作结构

Fig.2 Work structure of manager-worker-writer mode

2.2.3 管理者进程

不同来源的局部放电数据文件的存储格式虽有所不同,但通常都是以工频周期为单位存储局部放电波形采样值,并且 1 个文件内可能存储多个波形。因此,存在两种数据划分方式——以文件为单位和以周期为单位。以文件为单位分配任务可以减少进程之间的通信次数从而减小开销,但容易造成负载不平衡使部分工人空闲而降低并行算法性能。而以周期为单位分配任务会因通信次数多而引入额外的通信开销,但能使程序的并行化粒度更细,有利于负载均衡,尤其适用于异构集群。以周期方式分配任务的管理者进程伪代码如下:

```

代码 2: void manager(string dirname, int size)
1: string* filelist = ScanDir(dirname); //获取数据文件列表
2: while(index < filelist.size){ //还有任务未分配
3:   file = filelist[index]; //待划分文件 file
4:   while(1){ //循环按周期划分文件 file
5:     [key, clcsize] = Analysis(file); //解析文件
6:     pos+=clcsize; //计算当前划分位置的字节偏移量 pos
7:     if(pos>=file.size) {pos=0; index++;break;} //下个文件
8:     MPI_Recv(request,...); //从 worker 处接收任务请求
9:     sprintf(strtask,"%s+%d+%d",file,key,pos); //组成消息
10:    MPI_Send(strtask,..., request[0], ...); //给工人发任务
11:  } //end while(1)
12: } //end while(index)
13: int finished=0;
14: do{ //数据分配完准备退出进程
  
```

```

15: MPI_Recv(request,...); //从 worker 处接收任务请求
16: strtask[0]='\n'; //空消息
17: MPI_Send(strtask, ..., request[0],...); //给工人发空消息
18: finished++; //更新已经结束的 worker 数目
19: }while(finished<(size-2)); //一共有 size-2 个 worker
  
```

2.2.4 工人进程

工人进程在自己空闲时向管理者请求任务,拿到任务后进行脉冲提取处理。工人根据 key 查询映射表选择正确的文件解析程序和脉冲提取算法参数,然后才能对该局部放电信号进行准确的脉冲提取。处理完成后以周期为单位将 key 和结果发送给写者,如此循环直到接收到空任务消息时终止自己,并向写者发送空结果告知,以周期方式分配任务时工人伪代码如下:

```

代码 3: void worker(string dirname, int rank)
1: while(1){ //循环请求任务直到没任务可做
2:   request[0]=rank; //请求任务消息,rank 用来标记工人
3:   MPI_Send(request, ..., 0, ...); //给 0 进程发送任务请求
4:   MPI_Recv(strtask, ...); //接收任务消息
5:   if(strtask[0]!='\n'){ //非空消息
6:     [file,key,pos]=Parse(strtask); //分解任务指令
7:     data=GetData(file,dirname,pos,key); //按规则解析数据
8:     result = AnalysisData(data,key); //进行脉冲提取
9:     MPI_Send(result,...,1, key, ...); //给 1 进程发送结果
10:  } else{ //收到空消息,准备退出进程
11:    result[0]=1.0; //空消息标记
12:    MPI_Send(result, ..., 1, key,...); //给 1 进程发空结果
  
```

```

13: break; //工人 rank 进程退出
14: } //end if
15: } //end while(1)

```

2.2.5 写者进程

写者从程序启动时就在等待从工人处接收以周期为单位的处理结果, 接收到结果之后根据来源 key 和不同需求做后续处理, 例如绘制放电指纹图、 φ - q - n 放电谱图、写入文档等, 接收完所有结果之后就会退出进程, 伪代码如下:

```

代码 4: void writer(string dirname, int size)
1: int finished=0;
2: do { //循环接收结果直到所有工人终止
3: MPI_Recv (result, ..., key, ...); //接收任务消息
4: if(result[0]!=1.0)
5: PostProcess(result,key); //根据 key 进行后续处理
6: else
7: finished++; //更新已经结束的 worker 数目
8: } while(finished<(size-2)); //一共有 size-2 个 worker

```

2.2.6 算法性能分析

所提并行算法的执行时间 T 包括进程之间的通信开销 T_{com} 和数据处理时间 T_{pro} 两部分, 事实上, 由于所提并行算法呈现出流水线特性, T_{com} 与 T_{pro} 有重叠, 但为了便于分析, 忽略此点。

假设待分析数据文件为 f 个, 每个文件包含 c 个周期的放电信号, 工人有 w 个, 进程间单位时间内可传输 v 字节。首先分析 T_{com} , 由图 2 可以看出, 主要存在以下 3 种通信事件: 1) 工人向管理者请求任务, 发送自己的进程号即可, 采用 1 个短整型(short, 2 byte)。2) 管理者向工人发送任务, 为了减少通信量, 可将数据放在每个工人节点本地, 管理者只发送任务数据的元信息, 按文件划分只发送文件名和数据来源 key, 按周期划分还需要加上字节偏移量。文件名采用字符数组(char[]), 占用字节视文件名长度 l 而定, 数据来源 key 采用 1 个短整型(short, 2 byte), 字节偏移量采用 1 个整型(int, 4 byte)。3) 工人向写者发送处理结果, 这里通信量最大, 需要发送数据来源 key 和本周期信

号提取到的放电脉冲的幅值和相位, 若每周期内平均放电 q 次, 幅值和相位均采用单精度浮点数(float, 4 byte), 则需要 $(8q+2)$ byte。因此可以推导出两种数据划分方式的通信开销分别为

$$T_{\text{file_com}} = [f(4+l)+fc(8q+2)]/v \quad (1)$$

$$T_{\text{cycle_com}} = [fc(8+l)+fc(8q+2)]/v \quad (2)$$

再看数据处理部分 T_{pro} , 可细分为串行处理部分(t_s)和并行处理部分, 两种数据划分方式只在并行部分存在差异, 假设单周期局部放电信号的脉冲提取所需时间平均为 t_0 , 则两种数据划分方式的数据处理耗时分别为($\lceil x \rceil$ 表示对 x 向上取整)

$$T_{\text{file_pro}} = t_s + \lceil f/w \rceil ct_0 \quad (3)$$

$$T_{\text{cycle_pro}} = t_s + \lceil fc/w \rceil t_0 \quad (4)$$

因此, 总时间分别为

$$T_{\text{file}} = t_s + [fc(8q+2)]/v + [f(4+l)]/v + \lceil f/w \rceil ct_0 \quad (5)$$

$$T_{\text{cycle}} = t_s + [fc(8q+2)]/v + [fc(8+l)]/v + \lceil fc/w \rceil t_0 \quad (6)$$

经过比较可以看出, 两者的差异在于式(5)、(6)中第 2 个加号后面的部分, 当数据量 fc 固定时, c 越大, f 越小, 则 T_{file} 中 $[f(4+l)]/v$ 项会越小, 但由于划分数据时不能拆分文件, 最坏情况下 f/w 余数为 1 时, 最后一轮多出的 ct_0 也会越大。而对于 T_{cycle} 来说, 虽然通信开销显然比 T_{file} 大, 但由于 $\lceil fc/w \rceil t_0 \leq \lceil f/w \rceil ct_0$, 且最差情况下最后一轮也只会多出 t_0 , 负载更平衡。例如, $f=60$, $c=100$, $w=16$, 则 $\lceil f/w \rceil ct_0 = 400t_0$, $\lceil fc/w \rceil t_0 = 375t_0$ 。理论上, 某数据块的通信开销必须小于工人处理该数据块的时间, 并行计算才有意义, 经实验测定, 对于高采样频率的局部放电信号, t_0 大于所需的通信开销, 因此, 以周期划分数据更合理。

3 实验与分析

3.1 实验数据

为了验证本文所提方法的有效性, 在实验室内采用不同的放电模型来产生局部放电信号进行分析。实验模型接线图如图 3 所示, 采集仪器采用 TWPD-2F 局部放电综合分析仪, 采样频率为 20~80 MHz 不等, 采集频带为 40~300 kHz。

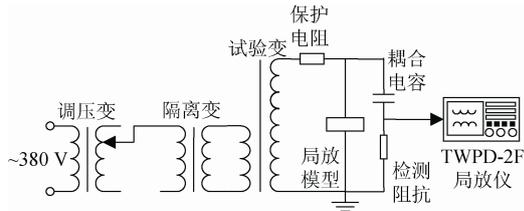


图 3 实验接线图
Fig.3 Wiring diagram for experiment

3.2 并行计算环境

所搭建的 MPI 并行计算平台由 4 台 ThinkServer RD460 和 1 台 Dell R210II 组成, 服务器间用千兆交换机互连, 在服务器上安装 Xenserver 并虚拟出 9 个节点, 每个节点均为双核心, 详细配置参数如表 1 所示。由于每个 MPI 进程需要 1 个处理器核心, 因此本平台最多可拥有 16 个工人。并行算法和串行算法均采用 C++ 语言编写, 后者运行在其中单个虚拟机①上。

表 1 MPI 并行计算平台配置参数

Tab.1 Configuration parameters of MPI parallel computing platform

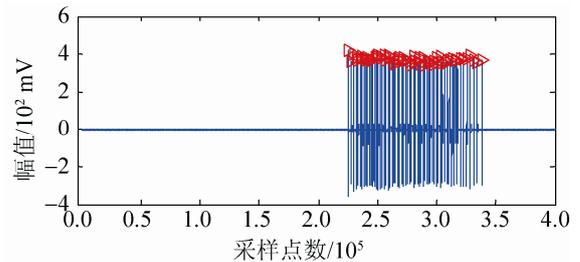
配置项	配置信息
虚拟机①×8	处理器 E5-2609 v2 2.50 GHz; 内存 7 GB; 硬盘 300 GB
虚拟机②×1	处理器 i3-2120 3.30 GHz; 内存 3 GB; 硬盘 250 GB
交换机	型号 guidway S1724G, 传输速率 10/100/1000 Mbps
操作系统	Ubuntu 12.04 LTS
软件	Xenserver-6.5; mpich-3.1.4; g++ 4.6.3

3.3 算法准确性分析

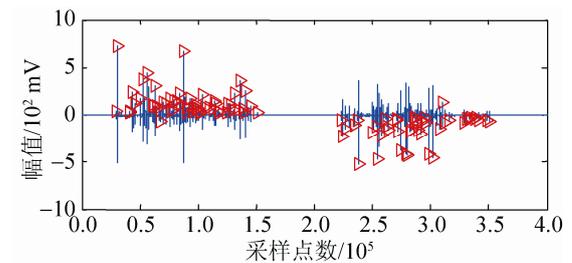
为了检验算法对于局部放电脉冲提取的有效性, 对不同放电模型试验所得的局部放电信号进行脉冲提取, 阈值参数取默认值 $T_1=1.5$, $T_2=3$, 结果如图 4 所示。由结果可知, 算法对不同放电模型产生的放电均能有效提取出放电脉冲。

为了进一步从宏观上检验算法的准确性, 从实验数据中随机取 10 个周期的信号, 应用算法对其进行检测并统计结果如表 2 所示, 其中的漏检和误检结果均是同人工观察结果比较所得。观察表 2

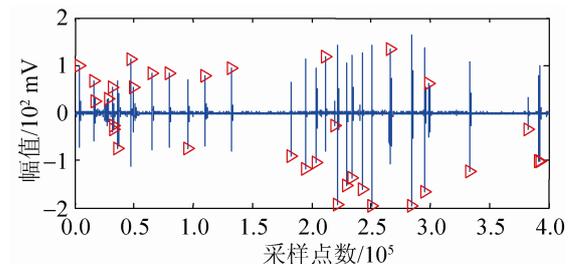
可知, 算法对电晕放电的检测结果非常准确, 对其他 3 种放电也很少出现误检, 但漏检的脉冲较多, 尤其是板对板模型的信号。观察图 4 可以发现, 电晕放电信号的脉冲只分布于半周, 且脉冲之间的间隔大致相等, 非常契合该算法的思想, 因此准确率很高。而其他 3 种放电分布杂乱, 且有较多脉冲挨得很近以至漏检, 但误检和漏检率均在可接收范围内。



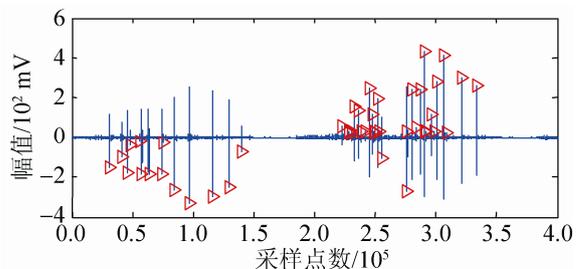
(a) 电晕模型



(b) 板对板模型



(c) 锥对板模型



(d) 多尖对板模型

图 4 脉冲提取结果

Fig.4 Results of pulses extraction

表 2 脉冲提取统计结果

放电模型	结果	漏检	误检	准确率/%
电晕	620	0	0	100.0
板对板	752	103	5	87.88
锥对板	341	6	4	98.25
多尖对板	527	27	2	95.11
总计	2 240	136	11	94.25

3.4 计算性能比较

脉冲提取算法的输入是单个工频周期的局部放电信号,因此本文以周期作为计量单位对串行算法(LSerial)和 MPI 并行算法(以文件划分 LFile 和以周期划分 LCycle)进行实验。分别采用不同算法对不同数据规模的局部放电信号进行脉冲提取,它们的采样频率均为 20 MHz 且放电密度相当,每 100 个周期信号存为 1 个数据文件,实验结果如图 5 所示(图例中的数字 16 和 8 表示采用的工人数目)。

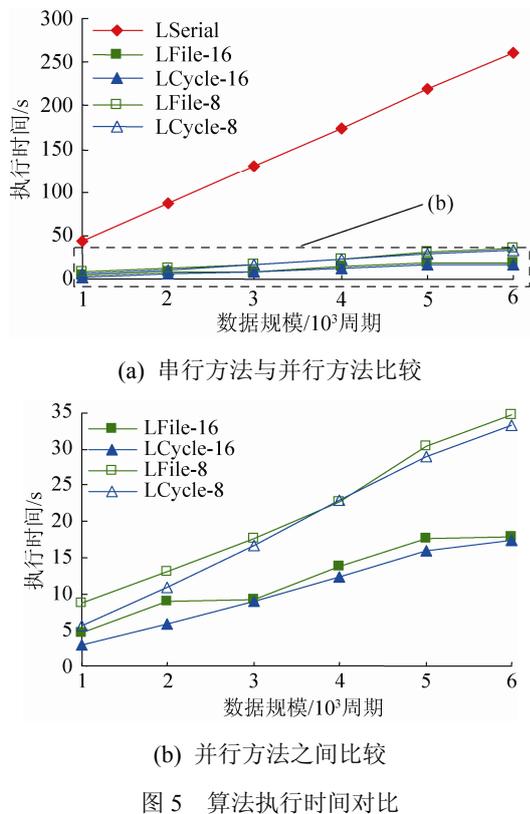


图 5 Comparison for execution time of different algorithms

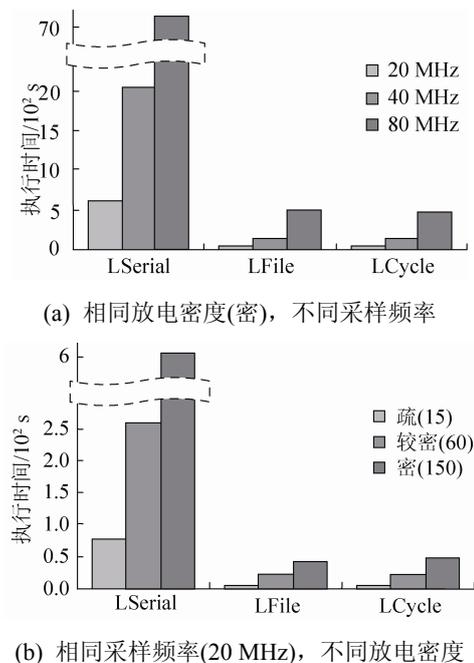
由图 5(a)可以看出,并行算法较串行算法具有显著的效率提升,且执行时间随数据规模增大而增

长得较缓,工人数目越多则执行时间越短。实验中数据规模仅为 6 000 个周期,但实际应用中数据规模会远远大于此,并行算法将更能发挥优势。

图 5(b)则表明 LCycle 优于 LFile,这是由于当文件数目不被工人数整除时,以文件划分数据时在最后一轮会使部分工人空闲,造成计算资源浪费。例如 16 个工人处理 2 000 个周期(文件数目为 20)信号时,需要 2 轮并行完成,但第 2 轮仅有 4 个工人在工作。当以周期划分时,16 个工人正好能整除 2 000 个周期,需要 125 轮并行完成,充分利用了计算资源,由图 5(b)可以看出两者的差异。即便不能恰好整除,由于并行粒度更细,负载更容易平衡,这在工人计算能力不等时尤其有利。

3.5 采样频率和放电密度的影响

由 1.2 节中复杂度分析可知,该算法的性能与采样频率和放电密度息息相关。为了探讨采样频率和放电密度对并行算法执行效率的影响,将总周期数为 6 000 的不同采样频率(固定放电密度)和不同放电密度(固定采样频率)下的信号作为处理对象进行测试,16 个工人投入计算,结果见图 6(图例中“疏(15)”表示单周期内放电次数大致为 15 次)。

图 6 采样频率和放电密度的影响
图 6 Effect of sampling frequency and discharge density

由图 6 可以看出, 采样频率和放电密度均对该算法的执行时间有较大影响, 验证了 1.2 节中的复杂度分析。无论是高采样频率还是高放电密度, 串行算法都难以胜任, 而两种并行算法均有效提升运算效率。一般而言, 在现场环境中, 放电密度很难达到实验室中的水平, 放电大多较为稀疏, 而对信号的采样频率要求则可能更高, 在如此高采样频率下会更凸显本文所提并行算法的实用价值。

3.6 加速比分析

为验证并行算法的加速比, 分别采用 3 000 和 6 000 个周期的信号对两种并行算法进行实验, 改变采用的工人数目运行并行算法, 统计执行时间并由式(7)计算加速比。式(7)中 T_s 是并行算法在单个工人下的执行时间, T_w 是并行算法在 w 个工人时的执行时间, 所得加速比如图 7 所示。

$$S_{\text{speedup}} = T_s / T_w \quad (7)$$

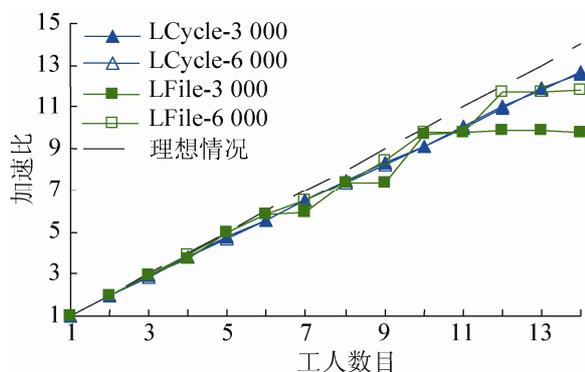


图 7 加速比比较

Fig.7 Comparison for acceleration ratio

由图 7 可知, LCycle 的加速比优于 LFile, 前者更接近于理想加速比, 且呈线性增长。而后者在某些情况下却保持不增长, 这是由以文件为单位划分任务的粗粒度并行造成的。图 7 中 LCycle 在两种数据规模下的加速比曲线几乎重叠, 这表明数据规模对 LCycle 加速比的影响较小。而对于 LFile 而言, 数据规模对其加速比的影响取决于文件数目能否被工人数整除。

4 结论

针对在线监测中局部放电信号脉冲提取存在的多源、海量问题, 本文在现有工作基础上提出了一种基于 MPI 编程模型的并行方法, 并搭建了并行计算平台, 验证该方法的使用效果, 得到以下结论:

1) 采用的脉冲提取算法对不同类型的局部放电信号均有效, 脉冲提取总准确率达到 90% 以上, 满足工程应用需求。

2) 基于管理者、工人、写者 3 种角色分工协作的并行框架, 通过辨别数据来源能够解决多源异构局部放电信号的处理问题。

3) 所提并行方法在处理多源海量局部放电信号时具有明显优势, 以 1 个工频周期作为任务粒度动态划分数据能够充分保证负载平衡, 得到更好的加速比。

参考文献:

- [1] 宋亚奇, 周国亮, 朱永利. 智能电网大数据处理技术现状与挑战 [J]. 电网技术, 2013, 37(4): 927-935. (Song Yaqi, Zhou Guoliang, Zhu Yongli. Present Status and Challenges of Big Data Processing in Smart Grid [J]. Power System Technology, 2013, 37(4): 927-935.)
- [2] Williams J W, Aggour K S, Interrante J, et al. Bridging High Velocity and High Volume Industrial Big Data through Distributed In-Memory Storage & Analytics [C]//Big Data (Big Data), 2014 IEEE International Conference on. Washington, DC, USA: IEEE, 2014: 932-941.
- [3] 姜芸, 周毓捷. 分布式局部放电在线监测技术在上海 500kV 交联聚乙烯电力电缆线路中的应用 [J]. 高电压技术, 2015, 41(4): 1249-1256. (Jiang Yun, Zhou Yunjie. Application of Distributed Partial Discharge On-line Monitoring Technology in Shanghai 500 kV XLPE Power Cable Line [J]. High Voltage Engineering, 2015, 41(4): 1249-1256(in Chinese).)
- [4] Majidi M, Oskuoee M. Improving Pattern Recognition Accuracy of Partial Discharges by New Data Preprocessing Methods [J]. Electric Power Systems Research (S0378-7796), 2015, 119: 100-110.
- [5] Raymond W J K, Illias H A, Abu Bakar A H, et al. Partial Discharge Classifications: Review of Recent Progress [J]. Measurement (S0263-2241), 2015, 68: 164-181.

- [6] 张晓星, 舒娜, 徐晓刚, 等. 基于三维谱图混沌特征的 GIS 局部放电识别 [J]. 电工技术学报, 2015, 30(1): 249-254. (Zhang Xiaoxing, Shu Na, Xu Xiaogang, et al. GIS Partial Discharge Recognition Based on Chaos Features of the Three-Dimensional Spectra [J]. Transactions of China Electrotechnical Society, 2015, 30(1): 249-254(in Chinese).)
- [7] Wei G, Tang J, Zhang X, et al. Gray Intensity Image Feature Extraction of Partial Discharge in High-Voltage Cross-Linked Polyethylene Power Cable Joint [J]. International Transactions on Electrical Energy Systems (S2050-7038), 2014, 24(2): 215-226.
- [8] IEC. High-Voltage Test Techniques - Partial Discharge Measurements [S]. Commission IET, 2000.
- [9] Nagesh V, Gururaj B I. Automatic Detection and Elimination of Periodic Pulse Shaped Interferences in Partial Discharge Measurements [J]. IEE Proceedings of Science, Measurement and Technology (S1350-2344), 1994, 141(5): 335-342.
- [10] 张士宝, 董旭柱, 林渡, 等. 局部放电监测中现场干扰的分析与抑制 [J]. 清华大学学报(自然科学版), 1997, 37(8): 109-112. (Zhang Shibao, Dong Xuzhu, Lin Du, et al. Analysis and rejection of interference in on-line partial discharge monitoring [J]. Journal of Tsinghua University (Science and Technology), 1997, 37(8): 109-112.)
- [11] 万元, 李朝晖, 薛松, 等. 水轮发电机局部放电在线监测中的脉冲识别方法 [J]. 高电压技术, 2009, 35(9): 2169-2175. (Wan Yuan, Li Zhaohui, Xue Song, et al. Pulses Recognition Method in Partial Discharge Online Monitoring of Hydro-generators [J]. High Voltage Engineering, 2009, 35(9): 2169-2175(in Chinese).)
- [12] Ashtiani M B, Shahrtash S M. Partial Discharge Pulse Localization in Excessive Noisy Data Window [J]. IEEE Transactions on Dielectrics and Electrical Insulation (S1070-9878), 2015, 22(1): 428-435.
- [13] Shahsavarian T, Shahrtash S M. Online Partial Discharge Signal Conditioning for Phi-Q-N Representation Under Noisy Condition in Cable Systems [J]. IET Science Measurement & Technology (S1751-8822), 2015, 9(1): 20-27.
- [14] 杨来, 史忠植, 梁帆, 等. 基于 Hadoop 云平台的并行数据挖掘方法 [J]. 系统仿真学报, 2013, 25(5): 936-944. (Yang Lai, Shi Zhongzhi, Liang Fan, et al. Parallel Approach in Data Mining Based on Hadoop Cloud Platform [J]. Journal of System Simulation (S1004-731X), 2013, 25(5): 936-944(in Chinese).)
- [15] 宫夏屹, 李伯虎, 柴旭东, 等. 大数据平台技术综述 [J]. 系统仿真学报, 2014, 26(3): 489-496. (Gong Xiayi, Li Bohu, Chai Xudong, et al. Survey on Big Data Platform Technology [J]. Journal of System Simulation (S1004-731X), 2014, 26(3): 489-496(in Chinese).)
- [16] 李贵兵, 金炜东, 蒋鹏, 等. 面向大规模监测数据的高铁故障诊断技术研究 [J]. 系统仿真学报, 2014, 26(10): 2458-2464. (Li Guibing, Jin Weidong, Jiang Peng, et al. Research on Fault Diagnosis Technology of High-speed Rail for Large Scale Monitoring Data [J]. Journal of System Simulation (S1004-731X), 2014, 26(10): 2458-2464(in Chinese).)
- [17] 宋亚奇, 周国亮, 朱永利, 等. 云平台下并行总体经验模态分解局部放电信号去噪方法 [J]. 电工技术学报, 2015, 30(18): 213-222. (Song Yaqi, Zhou Guoliang, Zhu Yongli, et al. Research on parallel ensemble empirical mode decomposition denoising method for partial discharge signals based on cloud platform [J]. Transactions of China Electrotechnical Society, 2015, 30(18): 213-222(in Chinese).)
- [18] 宋亚奇, 周国亮, 朱永利, 等. 云平台下输变电设备状态监测大数据存储优化与并行处理 [J]. 中国电机工程学报, 2015, 35(2): 255-267. (Song Yaqi, Zhou Guoliang, Zhu Yongli, et al. Storage Optimization and Parallel Processing of Condition Monitoring Big Data of Transmission and Transforming Equipment Based on Cloud Platform [J]. Proceedings of the CSEE, 2015, 35(2): 255-267(in Chinese).)
- [19] 赵春晖, 吴志力, 姜欣, 等. 跨平台电网规划数据融合与存储模式 [J]. 电力建设, 2015, 36(3): 119-122. (Zhao Chunhui, Wu Zhili, Jiang Xin, et al. Cross-Platform Data Fusion and Storage Pattern of Power Grid Planning [J]. Electric Power Construction, 2015, 36(3): 119-122(in Chinese).)
- [20] 徐田华, 杨连报, 胡红利, 等. 高速铁路信号系统异构数据融合和智能维护决策 [J]. 西安交通大学学报, 2015, 49(1): 72-78. (Xu Tianhua, Yang Lianbao, Hu Hongli, et al. Heterogeneous Data Fusion and Intelligent Maintenance Decision for High Speed Railway Signaling Systems [J]. Journal of Xi'an Jiaotong University, 2015, 49(1): 72-78(in Chinese).)
- [21] 吴焯, 陈萃, 熊伟, 等. 面向高效检索的多源地理空间数据关联模型 [J]. 计算机学报, 2014, 37(9): 1999-2010. (Wu Ye, Chen Luo, Xiong Wei, et al. Multi-Source Geospatial Data Correlation Model for Efficient Retrieval [J]. Chinese Journal of Computers, 2014, 37(9): 1999-2010(in Chinese).)
- [22] Liang F, Lu X. Accelerating Iterative Big Data Computing through MPI [J]. Journal of Computer Science and Technology (S1000-9000), 2015, 30(2): 283-294.

(下转第 75 页)

<http://www.china-simulation.com>