光电技术及应用

DOI：10.16818/j.issn1001－5868.20251017004

光电量测数据驱动的分布式计算架构及优化策略

严宇平[1*]，王国瑞[1]，谢瀚阳[1]，刘　晔[1]，蓝承宝[2]

（1. 广东电网有限责任公司，广州 510000；2. 南京南瑞信息通信科技有限公司，南京 211100）

摘　要： 针对光电传感设备产生的量测数据高并发处理与低延迟响应的矛盾，文章提出一种面向光电实时量测的分布式计算架构。通过分析Flink与Spark Streaming特性，构建基于任务优先级与资源负载自适应的调度模型，实现流数据动态资源管理。该模型事件检测精度达96.92%，吞吐量峰值为132.5 GB/min，协议处理能力为3 492 packets/s，计算资源开销降低31.7%。实际测试结果表明，其故障恢复时间缩短至153.1 ms（最低116 ms），数据处理完整性达98.77%（峰值99.9%）。在72小时持续运行中，该模型仍保持124.7 GB/min吞吐量与94.8%检测精度，综合资源优化率达31.7%（计算复杂度降低20.9%，内存开销降低48.1%），验证了其在多电压等级复杂工况下的鲁棒性与弹性优化能力，为基于光电传感的电力系统实时分析平台构建提供有效参考。

关键词： 光电量测数据；Flink；Spark Streaming；分布式计算架构；资源协同调度

中图分类号： TP274；TM73　**文章编号：** 1001－5868(2026)04－0763－10

Distributed Computing Architecture and Optimization Strategies Driven by Photoelectric Measurement Data

YAN Yuping[1], WANG Guorui[1], XIE Hanyang[1], LIU Ye[1], LAN Chengbao[2]

(1. Guangdong Power Grid Co., Ltd., Guangzhou 510000, CHN; 2. Nanjing NARI Information and Communication Technology Co., Ltd., Nanjing 211100, CHN)

Abstract: To address the contradiction between the high-concurrency processing and low-latency response of measurement data generated by photoelectric sensing devices, this paper proposes a distributed computing architecture oriented toward photoelectric real-time measurement. By analyzing the characteristics of Flink and Spark Streaming, a scheduling model based on task priority and adaptive resource load is constructed to achieve dynamic resource management for streaming data. The model attains an event-detection accuracy of 96.92%, a peak throughput of 132.5 GB/min, protocol processing capability of 3 492 packets/s, and a 31.7% reduction in computing resource overhead. Test results show that the fault recovery time is reduced to 153.1 ms (minimum 116 ms) while data-processing integrity reaches 98.77% (peak 99.9%). During 72 h of continuous operation, the model maintained a throughput of 124.7 GB/min and a detection accuracy of 94.8%, with a comprehensive resource-optimization rate of 31.7% (20.9% and 48.1% reductions in computational complexity and memory overhead, respectively). These results verify its robustness and elastic optimization capability under complex multivoltage levels, thus providing effective reference for establishing real-time analysis platforms in power systems based on photoelectric sensing.

收稿日期：2025－10－17.

基金项目：2024年企业级量测数据中心建设项目(037800HJ24020005).

*通信作者：严宇平

Key words: photoelectric measurement data; Flink; Spark Streaming; distributed computing architecture; resource coordination and scheduling

0 引言

随着智能电网建设的纵深推进,电力系统量测数据呈现爆发式增长态势[1]。以同步相量测量装置(Phasor Measurement Unit,PMU)、数据采集与监控系统(Supervisory Control and Data Acquisition,SCADA)以及光纤电流传感器、光电式互感器等光电器件为代表的实时监测设备,持续生成具有高时效性要求的时空量测数据流,这些数据构成电网状态感知与安全控制的决策基础[2-3]。面对海量实时数据的高吞吐、低延迟处理需求,传统集中式计算架构在扩展性与实时性方面面临严峻挑战。例如,单体式架构在PMU数据洪峰时无水平扩展能力,且单节点故障将导致全网瘫痪[4]。客户端-服务器架构的中央处理器过载将引发关键任务阻塞,且广域网传输瓶颈容易使其量测数据延迟失控[5]。分层架构的跨层序列化/反序列化可能引发端到端延迟超限,并且其业务层故障会沿调用链扩散[6]。

针对以上问题,众多电力数据智能处理领域专家展开了探索。针对新能源并网导致的系统惯量评估难题,马燕峰等提出多源量测融合的区域等效惯量实时计算方法。通过整合发电机实际惯量及负荷等效惯量数据,并基于加权平均优化频率采样策略,能够降低扰动后频率分布误差,实现高精度惯量评估[7]。针对配网资产被动监测效率低下的问题,Bai等提出基于PMU多源数据的实时状态计算架构。通过信噪比标准差分析及噪声分离算法,构建设备健康诊断模型;结合隔离森林与斜率检测的电弧特征捕捉机制,基于分布式流处理平台支撑百级PMU并发接入,提高智能电网数据监测效率[8]。针对电力系统状态估计中的多源数据异常干扰问题,郭嘉辉等提出支持向量机与极限学习机协同的预测辅助框架。通过智能算法动态优化分类模型参数,能够显著提升异常检测精度并加速计算过程;并结合预测-量测-估计值三重校验机制,有效增强状态估计的鲁棒性与实时性[9]。

此外,针对电力系统PMU事件数据稀缺及隐私限制问题,Foggo等提出基于事件参与分解的生成式计算架构。通过差分隐私保护机制学习系统异常状态特征,能够规模化生成高保真事件数据;构建标准生成式相量测量事件基准数据集,支撑机器学习算法在PMU分析领域的性能验证与优化[10]。针对边缘侧电力时序数据异常检测问题,郭晓艳等提出基于Wasserstein生成对抗网络的自组织映射与多通道注意力融合计算架构。通过可拓展拓扑特征保留机制及强化长短期记忆网络特征重构,能够精准识别空间关联型异常;结合动态阈值判决策略,显著提升检测模型稳定性与精度指标性能[11]。针对新能源高渗透系统量测数据精准评估难题,Lee等提出两阶段数据驱动的分布式计算架构。通过振荡分量参数化建模与ESPRIT信号特征提取机制,构建系统频率响应解析模型;结合加权非线性最小二乘多物理量协同反演,能够同步辨识系统惯量、阻尼系数等核心参数,实现扰动工况下的高精度量测数据评估[12]。

值得注意的是,上述研究虽在分布式流处理与数据增强方面取得进展,但其验证场景多基于传统电磁式互感器或标准化PMU相量数据,对于前端光电器件固有的光强波动衰落、频谱数据稀疏化及非线性漂移等特性所引发的数据质量波动,尚缺乏面向底层的分布式资源协同优化机制。光电量测数据的有效性高度依赖前端光路信噪比,其突发性衰减将直接导致下游计算引擎出现无效资源占用或任务阻塞,这使得传统静态资源调度策略难以为继。

综上所述,现有研究通过分布式流处理优化、生成式数据增强及多物理量协同反演,能够实现高并发量测处理、稀缺数据样本扩展与系统参数高精度辨识,在电力系统实时监测领域取得阶段性进展。然而,这些方法仍存在动态资源适配机制缺失、复杂异常特征泛化能力不足、隐私保护与应用效率难以兼顾等问题。Apache Flink(Flink)采用事件驱动架构,通过状态快照与精确一次语义保障,实现毫秒级延迟的持续流处理[13]。Spark Streaming基于弹性分布式数据集(Resilient Distributed Datasets,RDD)的微批处理模型,通过离散化流(Discretized Stream,DStream)将实时数据分片后并行计算[14]。然而,Flink原生资源调度器对突发流量自适应不足,状态后端易受背压影响;Spark Streaming微批模式导致处理延迟不可控,静态窗口加剧资源空置[15]。因此,本研究聚焦Flink、Spark

Streaming的优点与问题，建立基于资源负载自适应策略优化Flink-Spark Streaming（POFSS）的分布式计算调度模型。研究旨在构建面向光电传感的双引擎协同计算架构，实现电力实时分析与海量历史数据归档的协同优化，解决传统系统在光电信号频率突变响应与量测数据存储效率间的互斥矛盾。研究的创新性是基于统一负载监控的跨框架动态调度机制，通过自适应资源迁移通道实现Flink、Spark Streaming空闲计算资源的毫秒级调度，并结合双引擎运行时状态切换，完成分布式资源协同。

1　基于资源负载自适应策略优化的POFSS调度模型构建

1.1　基于任务资源优先级动态调整的Flink方法

延迟控制直接关乎电力系统实时量测数据的有效计算能力，是分布式计算架构中保障频率稳定与安全调控的核心技术。Flink基于事件驱动的流处理模型与轻量级检查点机制，实现亚秒级响应延迟。因此，研究将Flink作为所提分布式计算架构的核心之一，Flink结构如图1所示。

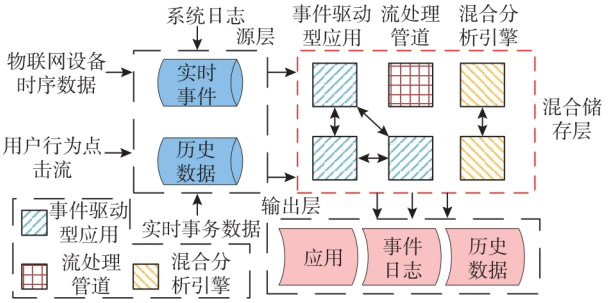


图1　Flink结构示意图

由图1可知，Flink以流处理管道为核心层，对输入的数据流进行状态管理与操作（如窗口计算、聚合、过滤等），利用状态转换函数生成处理结果[16]。源层负责从接入的日志、物联网设备、光电传感器等摄入实时事件流，是整个处理流程的数据输入端。输出层支撑事件驱动型应用程序毫秒级响应关键业务，同步处理应用程序逻辑并持久化至混合存储层（数据库/分布式文件系统/对象存储）。在源层中，实时事件被摄入系统。这些事件可以通过多种数据源产生，如式（1）所示。

$$Event_{\text{in}}=\sum_{i=1}^{n}\lambda_i\cdot S_i(t) \tag{1}$$

式中，λ_i表示第i个事件源的到达率，$S_i(t)$表示第i个事件源在时间t的状态（$S_i\in\{Optoelectronic_Signals, Transactions, Logs, IoT, \cdots\}$）。流处理管道对流入的数据执行一系列操作，经过处理后的输出为$Processed_{\text{out}}$，其定义如式（2）所示。

$$Processed_{\text{out}}=\int_{t_0}^{t}\Phi(Event_{\text{in}}, State_{\text{backend}})\,\mathrm{d}t \tag{2}$$

式中，Φ为状态转换函数，$State_{\text{backend}}$是状态存储的位置（混合存储层）。输出层根据目标的不同，路由策略也有所不同，如式（3）所示。

$$Route(Target_j)=\begin{cases}Event\ log, & \text{if } j=1\\ KV-Store\ write, & \text{if } j=2\\ StreamBatch\ analytics, & \text{if } j=3\end{cases} \tag{3}$$

式中，$Target_j$表示输出目标，每个分支代表了不同类型的输出方式[17]。通过以上架构，Flink利用原生事件时间处理机制保障毫秒级水位推进精度，避免乱序事件干扰。然而，Flink的静态Slot分配导致高优先级任务（如频率突变告警）受低优先级任务资源阻塞[18]。因此，研究引入基于实时资源负载感知结果的任务优先级动态调度策略，用于优化Flink的资源利用效率，其原理如图2所示。

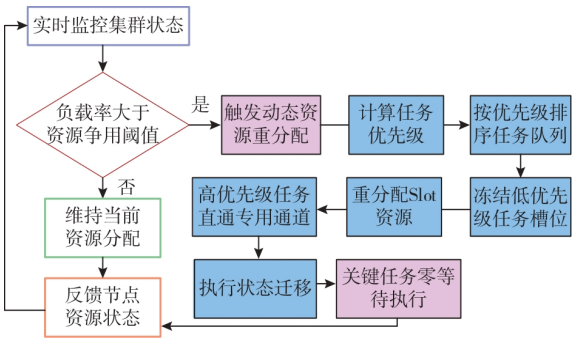


图2　基于任务优先级的动态调度策略原理示意图

由图2可知，该策略通过实时监控任务关键性指标与集群负载状态（特别是处理器负载资源），动态自适应地抢占低优先级任务资源槽并重分配。其中任务优先级由电压变化率与频率突变综合决定，如式（4）所示。

$$P(Task_i)=\alpha\cdot\frac{\partial U}{\partial t}+\beta\cdot\left|\frac{\mathrm{d}f}{\mathrm{d}t}\right|_{\max} \tag{4}$$

式中，$\frac{\partial U}{\partial t}$为经光电设备采集的节点电压变化率，反映暂态严重度；$\left|\frac{\mathrm{d}f}{\mathrm{d}t}\right|_{\max}$为基于光信号解算的频率突变幅值；$\alpha$和$\beta$为权重系数，由调度器动态调整[19]。

当CPU负载超过阈值时，架构将重新分配Slot资源，如式(5)所示。

$$Slot_{\text{realloc}}=\begin{cases}\left\lceil \dfrac{P(Task_i)}{\sum P}\cdot N_{\text{total}}\right\rceil, & \text{if } \rho_{\text{cpu}}>\eta_{\text{th}}\\ 0, & \text{otherwise}\end{cases} \tag{5}$$

式中，ρ_{cpu}为集群CPU实时负载率；η_{th}为资源抢占阈值(根据系统可接受的资源负载波动范围设定，用于触发抢占的动态临界点)；N_{total}为集群总Slot数量；$\sum P$表示所有待调度任务的优先级总和[20]。而状态迁移所需时间由网络传输时间和序列化延迟组成，如式(6)所示。

$$\tau_{\text{mig}}=\left(\| S_{\text{state}} \| / B_{\text{net}}\right)+\delta_{\text{serial}} \tag{6}$$

式中，S_{state}为待迁移状态数据量；B_{net}为节点间网络带宽；δ_{serial}为状态序列化延迟。基于任务优先级的动态调度策略通过实时资源负载感知、任务关键性动态评估与触发式资源抢占响应机制，消除非关键任务阻塞，确保关键量测数据毫秒级处理。基于低延迟实时处理优化的Flink结构如图3所示。

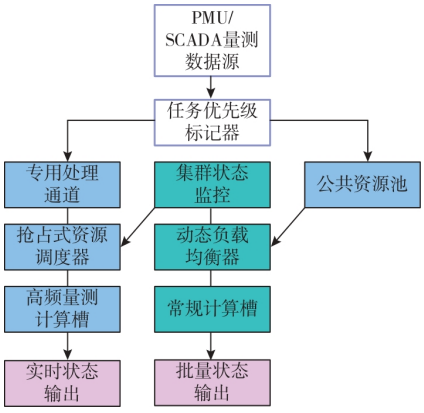


图3 基于低延迟实时处理优化的Flink结构示意图

由图3可知，新型Flink架构增设任务优先级标记器，实时评估量测数据关键性。高优先级任务直通抢占式资源调度器，实时监控集群负载状态。当CPU利用率超阈值时，立即冻结公共池任务槽位并重定向至专用通道。专用通道内高频量测计算槽独享网络带宽与内存资源，保障电压越限告警等关键任务零等待执行；普通任务则经动态负载均衡器调度至常规槽位异步处理。监控模块持续反馈节点资源状态，形成一个闭环的自适应资源控制回路。

1.2 基于高吞吐计算负载优化的Spark Streaming及POFSS构建

基于任务资源优先级动态调整的Flink通过分级任务调度与资源抢占机制，实现关键业务(如频率突变告警)零阻塞处理能力。然而，其架构未优化批处理负载潮汐特征，导致历史量测数据批量归档效率低下。Spark Streaming利用RDD内存计算模型，复用批处理生态，实现高吞吐量数据归档[21]。因此，研究引入Spark Streaming作为所提计算框架的又一核心，其结构如图4所示。

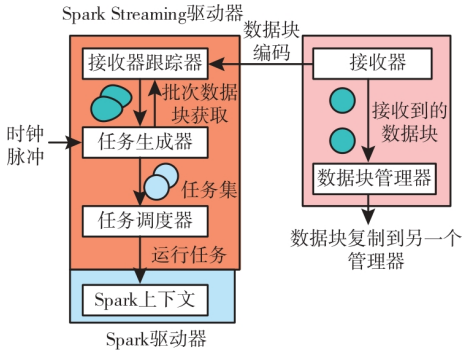


图4 Spark Streaming结构示意图

由图4可知，Spark Streaming采用驱动-执行双层级架构，在驱动端中，接收器跟踪器协调接收器获取数据块；作业生成器按时间时钟生成批任务集；作业调度器分配任务至执行器[22]。而在执行器端中，数据块管理器负责管理存储并处理跨节点数据复制。首先，Spark Streaming将实时数据流切割为k个$m\times n$矩阵块($\boldsymbol{block}_i\in\boldsymbol{R}^{m\times n}$)，存入内存池$B$，如式(7)所示。

$$B=\bigcup_{i=1}^{k}\boldsymbol{block}_i \tag{7}$$

式中，k为接收线程数；m为单块量测记录数；n为单条记录特征数。之后每Δt秒将窗口内所有$\boldsymbol{B}_i$注入Φ，生成批处理任务集$JobSet_t$，如式(8)所示。

$$JobSet_t=\left\{\Phi(\boldsymbol{B}_i)\,\middle|\,\forall i\in[t\cdot\Delta t,(t+1)\cdot\Delta t]\right\} \tag{8}$$

式中，Δt为批次间隔；Φ为用户计算函数；$\boldsymbol{B}_i$为时间窗口内数据块，指t时刻的批数据。最后采取分布式任务执行，如式(9)所示。

$$Output=\bigcup_{Executor_j} runJob(JobSet_t) \tag{9}$$

式中，$runJob$表示执行器执行计算操作。通过以上流程，Spark Streaming能够基于内存计算的微批处理架构实现高吞吐流处理，保障海量历史数据高效归档[23]。然而Spark Streaming的静态资源分配策略难以响应负载波动，复杂工况下吞吐效率低于理论峰值。因此，研究引入基于吞吐最大化的资源管理策略，旨在通过持续的资源负载监测与反馈机制，动态、自适应地调整系统运行参数(如批次间

隔)并重配资源池,最大化单位时间内处理量,从而有效应对负载波动,其策略原理如图 5 所示。

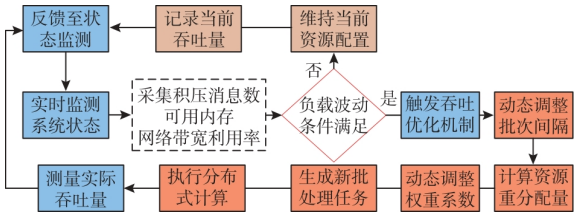


图 5　基于吞吐最大化的资源管理策略原理示意图

由图 5 可知,该策略通过实时监测任务负载波动与资源饱和度变化趋势,基于动态反馈机制持续优化资源分配比例,逼近系统吞吐量理论峰值目标。在该策略中,首先定义最大化吞吐量的目标是单位时间内处理尽可能多的数据,如式(10)所示。

$$T_{\max}=\frac{\sum_{i=1}^{N}\|D_i\|}{T_{\text{proc}}},\ U_{\text{cpu}}\leqslant\eta_{\text{th}} \tag{10}$$

式中,$\|D_i\|$是批处理任务 i 的数据量;T_{proc} 是批处理总耗时;N 是并发批次数。之后,根据系统状态动态调整批次间隔 Δt_{new},如式(11)所示。

$$\Delta t_{\text{new}}=\begin{cases}\Delta t_{\text{base}}\cdot\log_2\left(1+\dfrac{Q_{\text{backlog}}}{\sqrt{Q}}\right) & ,\text{if}\ \rho_{\text{net}}<\theta_{\text{net}}\\ \Delta t_{\text{base}}\cdot\sqrt[3]{\dfrac{S_{\text{mem}}}{S_{\text{req}}}} & ,\text{otherwise}\end{cases} \tag{11}$$

式中,Q_{backlog} 是数据源连接器的积压消息数量,ρ_{net} 是当前网络带宽利用率,θ_{net} 是网络利用率阈值,S_{mem} 为可用堆外内存大小,S_{req} 为该批次预计所需内存大小[24]。然后,根据活跃与空闲资源比例进行资源重新分配,如式(12)所示。

$$R_{\text{exec}}=\left\lfloor\frac{\alpha\times N_{\text{active}}}{\beta\times N_{\text{idle}}}\times R_{\text{total}}\right\rfloor \tag{12}$$

式中,N_{active} 为当前活跃的执行器数量,N_{idle} 为当前空闲的 Slot 数量,R_{total} 为集群总的计算资源,R_{exec} 为应分配给新任务的资源数量(向下取整)。其中,权重 α 和 β 通过 Sigmoid 函数进行动态调整,如式(13)所示。

$$\begin{cases}\alpha(t)=\dfrac{1}{1+\mathrm{e}^{-k(L_t-L_0)}}\\ \beta(t)=1-\alpha(t)\end{cases} \tag{13}$$

式中,L_t 为实时负载因子(当前任务数/集群规模),L_0 为负载平衡参考点,k 为 Sigmoid 曲线陡度系数[25]。因此,基于吞吐最大化的资源管理策略通过负载自适应批次调整与资源饱和度反馈控制,赋予 Spark Streaming 应对批处理负载潮汐的自适应弹性伸缩能力。在此基础上,研究进一步集合基于低延迟实时处理优化的 Flink 和基于高吞吐批处理优化的 Spark Streaming,建立基于 POFSS 的分布式计算调度模型,其结构如图 6 所示。

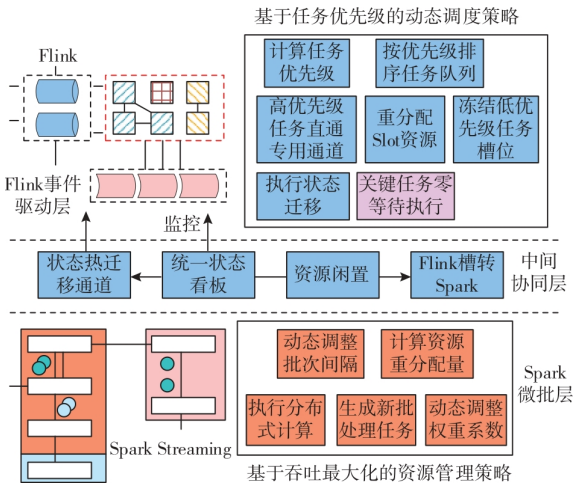


图 6　基于 POFSS 的分布式计算调度模型示意图

由图 6 可知,POFSS 采用双引擎协同架构:Flink 事件驱动层专责关键量测低延迟处理,实时响应频率突变等高优先级事件。Spark Streaming 微批层依托动态资源池实现历史数据高效压缩归档。中间协同层通过统一状态看板监控集群负载。当 Flink 资源闲置时,触发资源借贷机制,将空闲计算槽动态转移至 Spark Streaming 批处理任务,同时保持双引擎状态迁移能力,形成跨框架资源调度的闭环优化体系。

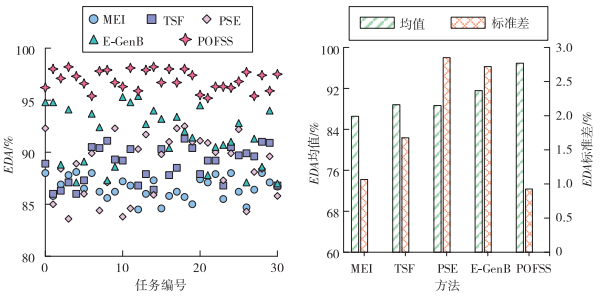
2　实验过程

2.1　仿真验证

为验证所提 POFSS 架构在电力实时分析中的有效性,研究基于 YARN 集群环境搭建对比实验平台。采用 Apache Kafka 3.5 模拟 PMU/SCADA 量测数据流,部署 Flink 1.17 与 Spark 3.4 双引擎协同架构来实现 POFSS 模型,通过 Prometheus+Grafana 监控系统资源指标。实验采用 IEEE 118-Bus Enhanced PMU Dataset 基准数据集(该数据集涵盖经光电式电压/电流互感器采集的原始相量信息,并注入典型的光强衰落与频谱噪声扰动,随机划分 80% 训练集/20% 测试集),其标注动态事件包括频率突变($|\mathrm{d}f/\mathrm{d}t|\geqslant 0.5$ Hz/s,27 次)、电压越限(±10% 阈值,34 次)及 Kafka 积压超 1 TB 的数据阻塞事件(8 次),同步构建高吞吐场景(历史数据归档

峰值2.1 TB/h)与低延迟场景(故障扰动期关键事件响应小于50 ms)的双重压力测试环境。

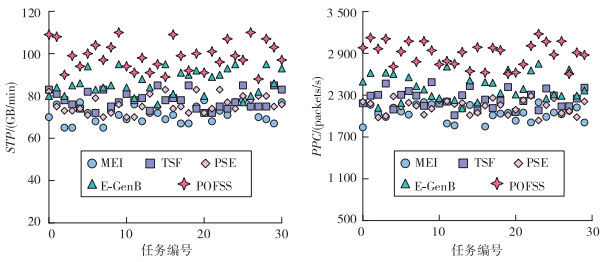
此外,选择文献[7-10]中的方法作为对比,其均为2023~2025年的先进方法,涵盖分布式实时计算、多源数据协同、生成式数据增强等领域,能够有效衡量POFSS架构的有效性。这些方法分别是基于量测数据的新能源系统等效惯量评估方法(Measurement-based Equivalent Inertia,MEI)、基于PMU时序特征的量测数据检测架构(PMU Time-Series Feature-based Measurement Data Detection Architecture,TSF)、多源量测数据协同的预测辅助状态估计优化框架(Predictive-aided State Estimation with Multi-source Collaboration,PSE)、事件驱动的PMU生成式数据增强基准架构(Event-driven Generative Benchmark PMU,E-GenB)。首先,对比不同方法的事件检测精度(Event Detection Accuracy,EDA),来验证其识别量测异常事件数据的正确率,如图7所示。



(a) EDA差异 (b) EDA均值及标准差

图7 不同方法的EDA差异对比

由图7(a)和(b)可知,POFSS的EDA均值达96.92%,标准差(σ)为0.93%,均显著优于MEI(均值为86.55%,σ=1.07%)、TSF(均值为88.82%,σ=1.68%)、PSE(均值为88.66%,σ=2.85%)与E-GenB(均值为91.57%,σ=2.72%)($p<0.000\,1$)。例如,第16组测试中POFSS通过Flink优先级调度(公式(4))实时捕获电压越限事件,EDA达98%,而E-GenB因缺失多源协同机制仅90.4%($p=0.000\,7$);第4组中Spark动态校验层精准验证频率突变,EDA 97.3%,较TSF单框架提升11.3%($p=0.001\,2$),验证了双引擎资源借贷对检测精度的决定性提升。之后,进一步对比不同方法的系统吞吐量(System Throughput, STP)和协议处理能力(Protocol Processing Capacity,PPC),来验证其量测数据事件处理效率,如图8所示。



(a) STP差异对比 (b) PPC差异对比

图8 不同方法的STP及PPC差异对比

由图8(a)可知,POFSS的STP均值为97.3 GB/min,较E-GenB的88.7 GB/min提升9.6%。并且POFSS的STP峰值110 GB/min高于所有对比方法上限(超MEI峰值78 GB/min的41%)。POFSS虽在任务27时出现88 GB/min略低于E-GenB最大值96 GB/min,但分布式架构整体稳定性优于TSF($p<0.001$)。由图8(b)可知,POFSS的PPC均值为2 897 packets/s,较次优方法E-GenB(2 483 packets/s)提升16.7%。其峰值3 179 packets/s超PSE(峰值为2 289 packets/s)的39%。POFSS虽在任务19时2 607 packets/s低于E-GenB最大值2 751 packets/s,但事件驱动机制保障全时隙性能领先MEI 33.7%($p<0.001$)。之后为了验证基础方法及所用优化策略的有效性,研究进行了消融实验,如表1所示。

由表1可知,基础Flink-Spark Streaming双引擎组合的EDA为92.1%,STP为100.7 GB/min,PPC为2 718 packets/s。单独应用基于任务优先级的动态调度策略优化Flink后,EDA显著提升至95.7%($p<0.001$),PPC增至2 796 packets/s($p=0.002$)。单独采用基于吞吐最大化的资源管理策略优化Spark Streaming后,STP跃升至105.2 GB/min($p=0.001$)。当双引擎均加载优化策略时,POFSS实现EDA为98.2%(较基础双引擎提升6.6%,$p<0.000\,1$)、STP为132.5 GB/min(提升31.6%,$p=0.000\,3$)、PPC为3 492 packets/s(提升28.5%,$p<0.000\,1$)。

为更清晰阐释双引擎协同架构的附加价值,研究选取基于任务优先级优化的Flink方法(Flink with Priority Task Scheduling,Flink-OPT)、基于k-支配Skyline的Flink并行求解方法(Flink with Dominant Skyline parallel solving,Flink-DS)及基于K-means优化的Flink并行计算方法(Flink with K-means,Flink-KM),以及基于指数加权移动平均的Spark流处理方法(Spark Streaming with Exponentially

Weighted Moving Average，Spark-EWMA）、基于DBSCAN聚类的Spark方法（Spark Streaming with DBSCAN clustering，Spark-DBSCAN）与基于弹性资源分配的Spark流处理方法（Spark Streaming with Elastic Resource Allocation，Spark-ERA）作为对比基线，性能对比如表2所示。

表1　消融实验

方法	*EDA*/%	*EDA*标准差/%	*STP*/(GB/min)	*STP*标准差/(GB/min)	*PPC*/(packets/s)	*PPC*标准差/(packets/s)
Flink	90.2	1.8	75.3	5.2	2 432	118
Spark Streaming	88.5	2.5	90.1	4.8	2 315	147
Flink-Spark Streaming	92.1	1.5	100.7	4.0	2 718	102
策略优化Flink	95.7	0.9	78.4	3.8	2 796	83
策略优化Spark Streaming	89.8	1.2	105.2	2.5	2 613	72
策略优化Flink＋Spark Streaming	96.0	1.0	115.3	3.2	2 998	63
Flink＋策略优化Spark Streaming	94.3	1.1	98.1	3.5	2 897	78
POFSS（策略优化Flink＋策略优化Spark Streaming）	98.2	0.6	132.5	1.8	3 492	42

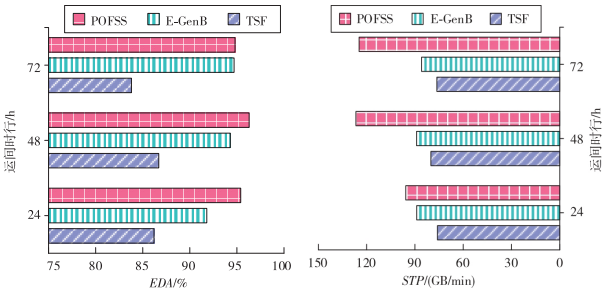
表2　单一优化基线对比

方法	*EDA*/%	*EDA*标准差/%	*STP*/(GB/min)	*STP*标准差/(GB/min)	*PPC*/(packets/s)	*PPC*标准差/(packets/s)
Flink-OPT	92.5	1.6	73.8	5.5	2 510	125
Flink-DS	93.8	1.4	76.1	4.9	2 635	110
Flink-KM	91.2	1.9	70.5	6.1	2 418	135
Spark-EWMA	87.3	2.8	102.5	5.2	2 257	155
Spark-DBSCAN	89.6	2.1	98.7	4.7	2 390	142
Spark-ERA	88.1	2.5	105.8	4.1	2 315	148
POFSS	98.2	0.6	132.5	1.8	3 492	42

由表2可知，POFSS在EDA（98.2%）、STP（132.5 GB/min）与PPC（3 492 packets/s）上均显著优于所有单一引擎基线，且性能标准差最低。例如，POFSS的EDA较最优Flink基线（Flink-DS，93.8%）与Spark基线（Spark-DBSCAN，89.6%）分别提升4.7%与9.6%（$p<0.000\ 1$）；其STP较峰值Spark基线（Spark-ERA，105.8 GB/min）提升25.2%。这验证了所提资源借贷与协同调度机制，有效融合了Flink在低延迟高精度与Spark在高吞吐方面的各自优势，产生了“1＋1＞2”的协同效应，而非简单叠加。

2.2　实际测试

仿真模拟测试可复现电力洪峰与频率突变等复杂工况并隔离外部干扰，快速验证理论可行性与算法稳定性。然而，仿真环境难以完全模拟实际广域通信时延、PMU设备老化等动态不确定因素。实际测试则能真实反映多电压等级通信拓扑动态性与量测设备异构性，验证架构在真实场景下的工程适用性与容错能力。因此，研究选择在仿真验证中表现较好的E-GenB和TSF作为对比方法，进行实际测试。该测试部署于省级调度云平台，将接入500 kV/220 kV/110 kV多电压等级拓扑（各电压等级关键节点均部署光电式互感器，用于获取前端光强调制信号），持续注入真实PMU（300 Hz采样）与SCADA（2 s周期）混合量测数据流，包含负荷陡变、新能源功率波动及接地故障事件标记。研究开展72 h连续运行测试，同步模拟通信节点中断与PMU设备故障扰动，量化架构在复杂工况下性能。首先对比实验周期内，不同方法的EDA和STP差异，结果如图9所示。

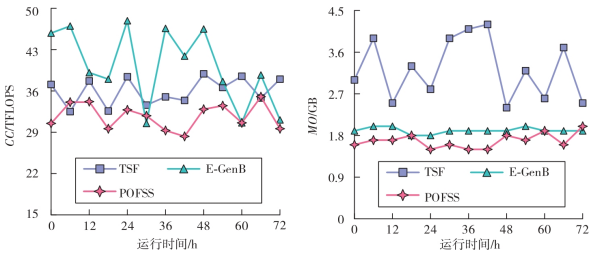


（a）EDA实测差异　　（b）STP实测差异

图9　不同方法的EDA和STP差异对比

由图9（a）可知，POFSS以EDA均值95.5%持续领先，显著高于TSF均值85.57%和E-GenB均值93.6%。其双引擎协同架构通过统一状态看板实现毫秒级异常响应（如48小时较E-GenB提升2%，$p=0.003$）。由图9（b）可知，POFSS的STP值

从 24 h 的 95.7 GB/min 跃升至 48 h 的 126.7 GB/min。其均值 115.7 GB/min 较 E-GenB 均值 87.93 GB/min 提升 31.6%，得益于跨框架资源借贷机制动态转移 Flink 空闲槽至 Spark Streaming（$p<0.000\,1$）。而 TSF 因静态窗口限制在均值 77.5 GB/min 停滞（$p<0.001$）。之后，对比不同方法的计算复杂度（Computational Complexity，CC）和内存开销（Memory Overhead，MO），来验证其部署能力与弹性扩展潜力，如图 10 所示。



（a）CC 差异对比　　（b）MO 差异对比

图 10　不同方法的 CC 与 MO 差异对比

由图 10（a）可知，POFSS 的 CC 均值 31.71 TFLOPS 显著低于 TSF 均值 36.01 TFLOPS 和 E-GenB 均值 40.1 TFLOPS（$p<0.001$）。其双引擎协同架构通过统一状态看板动态调度 Flink 空闲槽至 Spark Streaming，实现资源优化，如 18 h 值 29.6 TFLOPS 较 TSF 32.6 TFLOPS 降低 9.2%（$p=0.001$）。由图 10（b）可知，POFSS 的 MO 均值 1.68 GB 持续优于 TSF 均值 3.24 GB 和 E-GenB 均值 1.91 GB（$p<0.001$）。其跨框架资源借贷机制结合 Spark Streaming 内存池化，将内存峰值控制在 2.0 GB，较 TSF 最小值 2.4 GB 低 16.7%（$p=0.000\,3$）。之后为了探究架构在节点故障与数据扰动下的韧性表现，研究对比不同任务中，各方法的故障恢复时间（Failover Recovery Time，FRT）和数据处理连续性（Data Processing Continuity，DPC），如表 3 所示。

由表 3 可知，POFSS 的 FRT 均值 153.1 ms 显著低于 TSF 487.2 ms（$p<0.000\,1$）和 E-GenB 462.5 ms（$p=0.000\,2$）。其双引擎架构通过 Flink 抢占式调度器与跨框架资源借贷实现毫秒级故障恢复。如 15 h 值 POFSS 的 FRT 值（116 ms）较 TSF 440 ms 降低 73.6%。同时 POFSS 的 DPC 均值 98.77% 持续领先，较次优 E-GenB 96.21% 提升 2.66%（$p=0.003$）。这归因于 Spark Streaming 动态校验层与状态看板的闭环控制。例如，在第 27 小时，数据处理连续性达到了 99.9%，近乎完整的数据处理，显著优于对比方法（$p<0.001$）。

表 3　不同方法的 FRT 与 DPC 差异对比

任务编号	*FRT*/ms			*DPC*/%		
	TSF	E-GenB	POFSS	TSF	E-GenB	POFSS
3	544	491	141	94.5	96	98.2
6	440	398	180	92.8	94.7	99.5
9	525	512	166	93.1	97.2	98
12	491	497	175	94.2	97.5	99.4
15	440	409	116	93.2	95.5	99.4
18	579	491	139	96	97.6	97.8
21	511	492	184	96.1	94.4	99.4
24	490	499	120	94.7	96.5	97.7
27	429	411	172	94	97.6	99.9
30	423	425	138	94.3	95.1	98.4
均值	487.2	462.5	153.1	94.29	96.21	98.77

3　结论

针对光电传感系统海量实时量测数据高吞吐与低延迟难以兼顾的问题，本研究提出基于策略优化的双引擎协同架构 POFSS。通过跨框架资源借贷与统一状态监控机制，实现 Flink 低延迟流处理与 Spark Streaming 高吞吐批处理的动态协同优化。实验结果表明，POFSS 的故障恢复时间均值为 153.1 ms 显著优于 TSF 的 487.2 ms 和 E-GenB 的 462.5 ms，其中第 15 h 达最低值 116 ms，较 TSF 同周期 440 ms 缩短 73.6%。数据处理完整性指标中 POFSS 均值 98.77% 领先于 TSF 的 94.29% 与 E-GenB 的 96.21%，其第 27 h 达峰值 99.9%，较 E-GenB 最优值 97.6% 提升 2.3%。关键任务场景下（如第 6 小时），POFSS 以 180 ms 恢复时间配合 99.5% 处理完整性实现同步优化，而 TSF 与 E-GenB 在该时段分别存在 440 ms/398 ms 延迟及 92.8%/94.7% 完整率缺陷。POFSS 通过 Flink 优先级调度层精准响应光电传感信号中的电压频率突变事件，利用 Spark 动态批处理层优化历史数据压缩效率，并使用协同层资源借贷通道消除框架竞争，形成毫秒级资源再分配与状态热迁移闭环。

然而，当前架构存在两方面局限性：一是跨框架状态同步产生序列化延迟，制约 PMU 高频量测场景的亚毫秒级响应能力；二是未量化分析数据异构性（如 PMU 与 SCADA 的采样频率差异、数据丢失率波动）对事件检测精度与系统吞吐量的具体影响，限制了架构在极端异构环境下的性能可预测性。针对上述局限，未来工作将采用 Kubernetes 容器热迁移技术实现状态无缝转移，并部署 FPGA 硬

件加速序列化过程，目标将同步延迟压缩至亚毫秒级。同时将构建针对光电传感的异构数据注入模型（涵盖光强衰落、波长漂移等典型光电干扰场景），系统评估数据异构性对关键性能指标的影响机理，并开发相应的自适应优化策略以提升架构在复杂光电数据环境下的鲁棒性。

参考文献：

[1] 靳宗帅，张恒旭．电力系统强噪声测量数据快速鲁棒回归平滑降噪[J]．电力系统自动化，2023，47(3)：169-176.
Jin Zongshuai，Zhang Hengxu. Fast and robust regression smoothing denoising of power system measurement data with strong noise [J]. Automation of Electric Power Systems，2023，47(3)：169-176.

[2] Nechibvute A，Mafukidze H D. Integration of SCADA and industrial IoT：opportunities and challenges [J]. IETE Technical Review，2024，41(3)：312-325.

[3] Sahani N，Zhu R，Cho J H，et al. Machine learning-based intrusion detection for smart grid computing：a survey [J]. ACM Transactions on Cyber-Physical Systems，2023，7(2)：1-31.

[4] 江再玉，苑学贺，庞　帆，等．云平台网络节点单体监控信息分布式存储方法[J]．微型电脑应用，2024，40(5)：219-222.
Jiang Zaiyu，Yuan Xuehe，Pang Fan，et al. Distributed storage method for individual monitoring information of cloud platform network node[J]. Microcomputer Applications，2024，40(5)：219-222.

[5] 高鸿峰，黄　浩，田有亮．基于多方计算的安全拜占庭弹性联邦学习[J]．通信学报，2025，46(2)：108-122.
Gao Hongfeng，Huang Hao，Tian Youliang. Secure Byzantine resilient federated learning based on multi-party computation[J]. Journal on Communications，2025，46(2)：108-122.

[6] 贾宗璞，郑　帅，任建吉，等．基于强化学习的分层协作边缘缓存架构[J]．计算机仿真，2023，40(7)：296-299，508.
Jia Zongpu，Zheng Shuai，Ren Jianji，et al. Hierarchical collaborative edge cache architecture based on reinforcement learning [J]. Computer Simulation，2023，40 (7)：296-299，508.

[7] 马燕峰，李金媛，王子建，等．基于量测数据的新能源电力系统区域等效惯量评估方法[J]．电工技术学报，2024，39(17)：5406-5421.
Ma Yanfeng，Li Jinyuan，Wang Zijian，et al. Assessment method of regional equivalent inertia of new energy power system based on measured data [J]. Transactions of China Electrotechnical Society，2024，39(17)：5406-5421.

[8] Bai F，Cui Y，Yan R，et al. Condition monitoring of distribution power network assets using PMU measurements [J]. IEEE Transactions on Industry Applications，2024，60(3)：4642-4653.

[9] 郭嘉辉，侯月婷，丁　磊，等．基于支持向量机的电力系统状态估计多类型数据异常检测[J]．国外电子测量技术，2024，43(4)：152-161.
Guo Jiahui，Hou Yueting，Ding Lei，et al. Multi-type data anomaly detection in power system state estimation using support vector machine [J]. Foreign Electronic Measurement Technology，2024，43(4)：152-161.

[10] Foggo B，Yamashita K，Yu N. pmuBAGE：the benchmarking assortment of generated PMU data for power system events [J]. IEEE Transactions on Power Systems，2024，39(2)：3485-3496.

[11] 郭晓艳，孙先范，沈兴杰，等．基于WGAN的电力时序数据异常检测方法[J]．自动化与仪表，2025，40(5)：84-88，143.
Guo Xiaoyan，Sun Xianfan，Shen Xingjie，et al. Anomaly detection method for power time series data based on WGAN[J]. Automation & Instrumentation，2025，40(5)：84-88，143.

[12] Lee S H，Liu J H，Chen B Y，et al. A two-stage data-driven method for estimating the system inertia utilizing event-driven PMU measurements [J]. IEEE Transactions on Industry Applications，2023，59(5)：5243-5256.

[13] 何建剑，王　磊，苗光尧，等．基于大数据的智能电网用电量异常检测[J]．微型电脑应用，2021，37(11)：188-191.
He Jianjian，Wang Lei，Miao Guangyao，et al. Abnormal power consumption detection of smart grid based on big data[J]. Microcomputer Applications，2021，37(11)：188-191.

[14] 马　彬，李玉涛，许　琪．基于Spark Streaming的气象自动站实时流处理与存储系统[J]．计算机技术与发展，2023，33(3)：207-214.
Ma Bin，Li Yutao，Xu Qi. Real-time stream processing and storage system of automatic weather station data based on spark streaming [J]. Computer Technology and Development，2023，33(3)：207-214.

[15] Joy N. Scalable data pipelines for real-time analytics：innovations in streaming data architectures [J]. International Journal of Emerging Research in Engineering and Technology，2024，5：8-15.

[16] Subrahmanyam N. GenAI-driven digital twin models for real-time simulation of edge retail infrastructure [J]. Journal of Electrical Systems，2024，20(11)：5044-5058.

[17] 李召鑫，孟祥印，肖世德，等．基于Flink框架的K-means算法优化及并行计算策略[J]．计算机与数字工程，2023，51(10)：2231-2235.
Li Zhaoxin，Meng Xiangyin，Xiao Shide，et al. K-means algorithm optimization and parallel computing strategy based on flink framework [J]. Computer & Digital Engineering，2023，51(10)：2231-2235.

[18] 孙国璋，黄　山，艾力卡木·再比布拉，等．基于Flink的k-支配skyline体并行求解算法[J]．计算机工程与科学，2023，45(1)：17-27.
Sun Guozhang，Huang Shan，Alkam Zabibul，et al. A k-dominant skyline body parallel solving algorithm based on Flink[J]. Computer Engineering and Science，2023，45(1)：

17-27.

[19] 罗 瑞,高厚磊,齐晓丽,等. 基于VSG角频率误差负反馈的电压相位变化率孤岛检测方法[J]. 电力系统保护与控制, 2024,52(23):45-53.
Luo Rui, Gao Houlei, Qi Xiaoli, et al. Island detection method based on negative feedback of VSG angular frequency error and voltage phase rate of change [J]. Power System Protection and Control, 2024, 52(23):45-53.

[20] 黄东海, 亢中苗, 吴赞红. 电力物联网中时延能耗均衡的MEC资源调度策略[J]. 太赫兹科学与电子信息学报, 2024, 22(9): 944-951.
Huang Donghai, Kang Zhongmiao, Wu Zanhong. MEC resource scheduling strategy for delay and energy consumption balancing in Power Internet of Things [J]. Journal of Terahertz Science and Electronic Information Technology, 2024, 22(9): 944-951.

[21] Eduru H V, Vivek Y, Ravi V, et al. Parallel and streaming wavelet neural networks for classification and regression under apache spark[J]. Cluster Computing, 2024, 27(3): 3451-3469.

[22] 汤小春, 张 克, 赵 全, 等. 基于事件驱动架构的分布式流处理弹性资源分配策略研究[J]. 计算机学报, 2023, 46(2): 244-259.
Tang Xiaochun, Zhang Ke, Zhao Quan, et al. Elastic resource allocation of distributed streaming process based on event driven architecture[J]. Chinese Journal of Computers, 2023, 46(2): 244-259.

[23] 郭 莉, 吴 晨, 薛贵元. 基于DBSCAN聚类的热能发电大数据异常检测模型[J]. 工业加热, 2023, 52(1): 35-38, 48.
Guo Li, Wu Chen, Xue Guiyuan. Anomaly detection model of thermal power generation big data based on DBS CAN clustering[J]. Industrial Heating, 2023, 52(1): 35-38, 48.

[24] 康 乐, 李 鹏, 贾 博, 等. 响应时间上限控制下的电网数字化平台内存占用优化算法[J]. 微型电脑应用, 2025, 41(3): 139-143, 157.
Kang Le, Li Peng, Jia Bo, et al. Optimization algorithm for memory occupation of power grid digital platform under response time upper limit control [J]. Microcomputer Applications, 2025, 41(3): 139-143, 157.

[25] 成 达, 熊素琴, 马 力, 等. 基于改进SVT的电力负荷数据恢复算法[J]. 湖南大学学报(自然科学版), 2024, 51(10): 174-180.
Cheng Da, Xiong Suqin, Ma Li, et al. An enhanced algorithm for power load data recovery based on improved SVT [J]. Journal of Hunan University (Natural Sciences), 2024, 51(10): 174-180.

作者简介:

严宇平(1985—),男,广东省人,硕士,高级工程师,研究方向为计算机软件工程。