On Backup Battery Data in Base Stations of Mobile Networks: Measurement, Analysis, and Optimization

Xiaoyi Fan
School of Computing Science
Simon Fraser University
Burnaby, BC, Canada
xiaoyif@sfu.ca

Feng Wang
Department of Computer and Information Science
The University of Mississippi
University, MS, USA
fwang@cs.olemiss.edu

Jiangchuan Liu
School of Computing Science
Simon Fraser University
Burnaby, BC, Canada
jcliu@cs.sfu.ca

ABSTRACT

Base stations have been massively deployed nowadays to afford the explosive demand to infrastructure-based mobile networking services, including both cellular networks and commercial WiFi access points. To maintain high service availability, backup battery groups are usually installed on base stations and serve as the only power source during power outages, which can be prevalent in rural areas or during severe weather conditions such as hurricanes or snow storms. Therefore, being able to understand and predict the battery group working condition is of immense technical and commercial importance as the first step towards a cost-effective battery maintenance on minimizing service interruptions.

In this paper, we conduct a systematical analysis on a real world dataset collected from the battery groups installed on the base stations of China Mobile, with totally 1,550,032,984 records from July 28th, 2014 to February 17th, 2016. We find that the working condition degradation of a battery group may be accelerated under various situations and can cause premature failures on batteries in the group, which can hardly be captured by nowadays maintenance procedure and easily lead to a power-outage-triggered service interruption to a base station. To this end, we propose BatPro, a battery profiling framework, to precisely extract the features that cause the working condition degradation of the battery group. We formulate the prediction models for both battery voltage and lifetime and develop a series of solutions to yield accurate outputs. By real world trace-driven evaluations, we demonstrate that our BatPro approach can precisely predict the battery voltage and lifetime with the RMS error less than 0.01 v.

CCS Concepts

•Information systems → Data stream mining; •Hardware → Batteries; •Applied computing → *Enterprise applications;* **•Computer systems organization →** *Reliability;*

CIKM'16 , October 24 - 28, 2016, Indianapolis, IN, USA

ISBN 978-1-4503-4073-1/16/10. . . $15.00

DOI: http://dx.doi.org/10.1145/2983323.2983734

Keywords

Backup power system; battery aging profiling; remaining lifetime prediction; multi-instance multi-label learning

1. INTRODUCTION

The global mobile service markets have drastically expanded due to recent advances in the wireless telecommunication technologies, particularly the 3G/4G as well as the emerging 5G, which also render the mobile network bandwidth becomes higher than ever. Meanwhile, a large and growing number of wireless access points are being deployed by network operators, e.g., ShawOpen[1] and TELUS Wi-Fi hotspots[2]. To afford such great changes, more and more base stations have been strategically constructed and deployed in the mobile network infrastructure to satisfy the service coverage and quality (e.g., bandwidth) requirements. The base stations, which may range from large towers to small towers for local coverage, are then aggregated into the carrier network through the backhaul infrastructure by fiber lines or microwave links. However, this renders that the network downtime will cascade to all the dependent base stations if any part experiences a power outage [1], which although is arguably rare in urban areas but may often happen during bad weathers such as hurricane, tornado or snow storm, especially in rural areas.

Besides connecting to the utility grids, each base station is also equipped with a backup battery group to improve the service availability. When a power outage happens in the utility grid, to avoid any service interruption, the battery group discharges to support the communication equipment until the emergency generator is delivered to provide enough power supply. This makes the battery group and its working conditions play a critical role during the power outage.

However, while the scale of base stations is fast expanding for service coverage, the practice is that only a limited number of engineers can be dispatched for regular examination on the backup power systems, and the duration between two consecutive checks can be as long as over three months. During this period, if the battery working condition deteriorates, a power outage can easily keep interrupting the service, especially in rural areas or during bad weather conditions, where the emergency repairing can take days or even weeks.

Therefore, being able to predict the working condition of the battery group is of immense technical and commercial

[1]https://www.shaw.ca/wifi/
[2]https://wififinder.telus.com/

importance for system maintenance, as it is the first step towards a cost-effective battery maintenance on minimizing service interruptions. Yet a successful prediction requires not only the knowledge of the ageing processes that leads to a loss of performance within a battery, but also a good understanding of the stress status that may induce and accelerate the aging process, as well as the relationships between the stress status and the ageing process.

To this end, we collaborated with China Mobile, the mobile phone operator with the largest number of subscribers (currently 806 million users), and collected 46,913 equipment data with totally 1,550,032,984 rows from July 28th, 2014 to February 17th, 2016 including 105 categories of status, which are further processed and analyzed by our cloud computing platform with Hadoop 1.2.1 and Hive 1.2.1. We find that the working condition degradation of a battery group may be accelerated under various situations and can cause premature failures on batteries in the group, which can hardly be captured by nowadays maintenance procedure and easily lead to a power-outage-triggered service interruption to a base station.

To address this issue, we propose BatPro, a battery profiling framework, to precisely predict base station battery group working conditions by extracting the features that cause the working condition degradation. In particular, we decompose the voltage in time series into the aging and fluctuation terms. Based on the voltage aging term and status in logs, we analyze the battery aging trend to profile its evolution including the slope and the variance that indicates the working conditions of the battery. We formulate the battery aging problem as a multi-labels multi-instances problem with the objective to minimize the root-mean-square (RMS) error between the predictive value and the real-world value, so as to ensure an accurate battery voltage profiling. Based on the framework, we further formulate the prediction models for both battery voltage and lifetime and propose a series of solutions to yield accurate outputs. By real world trace-driven evaluations, we demonstrate that our BatPro approach can precisely predict the battery voltage and lifetime with the RMS error less than 0.01 v.

The rest of the paper is organized as follows. Section 2 provides our detailed observations on the battery properties. Section 3 presents BatPro framework to efficiently solve the problem. Section 4 discusses the performance evaluation results on our BatPro approach. We provide a literature review in Section 5 and we conclude this paper in Section 6.

2. OVERVIEW AND DATA ANALYSIS

In this section, we describe the background and motivation of our work, and then discuss the collected dataset from China Mobile with our observations to better understand the battery working condition and its deterioration process.

2.1 Backup Battery in Base Stations

We illustrate a generic backup power system in the base stations of mobile networks. The equipment in base stations is supported by the utility grid, where the battery group is installed as the backup power. In case that the utility grid interrupts, the battery discharges to support the communication switching equipment during the period of the power outage. As Fig. 1(a) shows, there are two battery groups in the mobile network base station and each battery group contains 24 cell batteries. In Fig. 1(b), the monitoring system

(a) Two battery groups (b) The monitoring systems

Figure 1: The monitoring systems

connects to each cell of the battery group and periodically records the voltage and status in both normal and abnormal situations. When the monitoring system reports the alert status, e.g. the power outage, the emergency repairing service is scheduled depending on the accident severity. Since few base stations have the diesel generators permanently installed on site, the engineers have to drive the emergency diesel generator to provide the power, which can take time to arrive at the site. The power outage can occur frequently and severely in the rural areas and developing countries due to the unstable utility grid. To make it even worse, the construction of infrastructure often makes that the base stations are difficult to reach, e.g. slippery rock trails in the mountains, where the workers have to manually carry the heavy generators to the site.

As a result, the mobile network carriers may have to trade off between the cost and the quality of service, so that they even abandon the base stations in the tough surroundings until the utility grid is restored. On the other hand, considering the labor costs for the mobile telecom carriers, the periodical maintenance for the battery group is generally of long intervals, which further exaggerates the possibilities of battery accidents during the outage of the utility grid. If the repairing engineers cannot arrive at the site before the battery group is exhausted, the availability of the base stations cannot be guaranteed. Thus prediction for the lifetime of battery group is meaningful for the service availability, which is helpful for maintenance engineers to solve the potential issues in advance during the periodical maintenance.

Although the Li-ion and NiCd batteries demonstrate the latest development in battery technology due to their smaller size, lower weight, and better storage efficiency, the major drawbacks of these types of batteries are the high cost. On the other hand, Lead-acid batteries in Fig. 1(a) have large capacities and thus have been widely used for storage in backup power supplies in base stations. The aging mechanism of Li-ion batteries attracts many efforts [21], where the frequent activities of Li-ion batteries produce lots of logs and provide possibilities to measure the battery working conditions. Yet the lead-acid batteries in base stations normally keep in the float-charging status, where float-charging status represents that a battery maintains the capacity by compensating for self-discharge after being fully charged. The monitoring system collects the float voltage from the float-charging batteries twice per day, which makes the dataset considerately sparse. Therefore extracting the features from such a sparse data source and predicting the working conditions of lead acid batteries pose many challenges and here we take the first attempt to tackle these issues.

historybattery			
equipmentid	recordtime	floatvalue	signalseverity
12405850	2015-01-28 9:37 AM	1.38341	0
12405850	2015-01-28 3:36 PM	2.22239	255
12405850	2015-01-28 4:51 PM	0.97471	0
historystatus			
equipmentid	starttime	endtime	meanings
12405850	2015-01-28 9:31 AM	2015-01-28 10:06 AM	voltage low
12405850	2015-01-28 12:01 PM	2015-01-28 12:09 PM	voltage low
12405850	2015-01-28 4:49 PM	2015-01-28 6:24 PM	voltage low

Figure 2: A piece of a real log file

2.2 Battery Data and Status

The log data that we have collected is from July 28th, 2014 to February 17th, 2016. In our dataset, we have identified 105 categories of status codes and obtained 46,913 equipment data with totally 531 tables and 1,550,032,984 rows. As Fig. 2 shows, the voltage data is recorded in the table named historybattery with an equipment id, record time, float value of voltage and signal severity, while a status in the table named historystatus consists of an equipment id, start timestamp, end timestamp and message text describing the status.

2.2.1 Voltage Readings

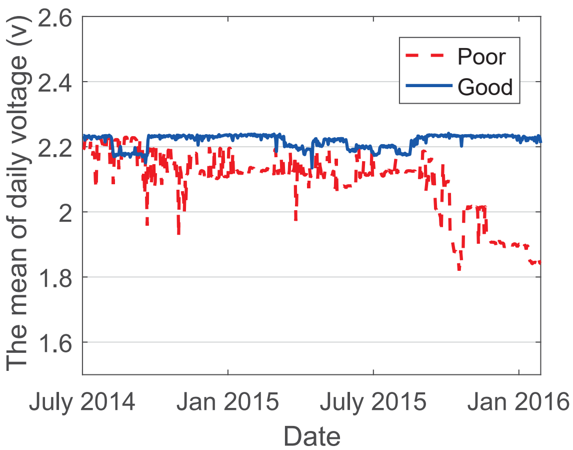


Figure 3: Mean voltage versus battery status

The voltage of each cell battery is the most important feature that we have measured, as it reflects the power output pattern of the battery. In general, we have observed two representative categories of cell batteries, where we manually choose 1578 batteries as the newly-installed group and put 1459 batteries into the nearly-dead group depending on the repair records. The rated voltage of cell is around 2.23v and the rated voltage of battery group is 53.5v, where 24 cell batteries are connected in serial as one battery group. Based on this, we further analyze the typical status of the voltage patterns inside the two representative cell battery categories. Fig. 3 shows the significant differences in mean voltage between the newly-installed and nearly-dead batteries. The blue solid line plots the mean voltage of newly-installed batteries, which judders between 2.14v and 2.24v. The red dotted line shows the decay trend on the mean voltage of the nearly-dead batteries. There is a clear downward trend close to the failure date, where the battery power frequently falls down and becomes quickly exhausted, causing many issues and alerts in the mobile network base station.

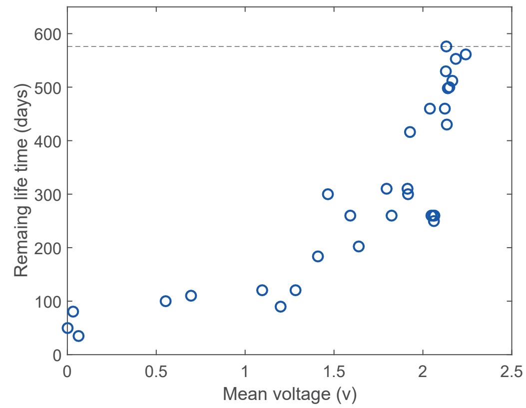


Figure 4: Correlation between the remaining life and mean voltage

Fig. 4 further plots how the mean voltage and the length of remaining lifetime correlate with each other, which indicates that the mean voltage has strong correlations with the battery life. Fig. 5 shows the results on the voltage variances, where the blue solid line represents the newly-installed battery can output a steady power and the variance of the voltage keeps very close to zero. The red dotted line illustrates that the variance of the nearly-dead batteries increases much faster than the newly-installed batteries.

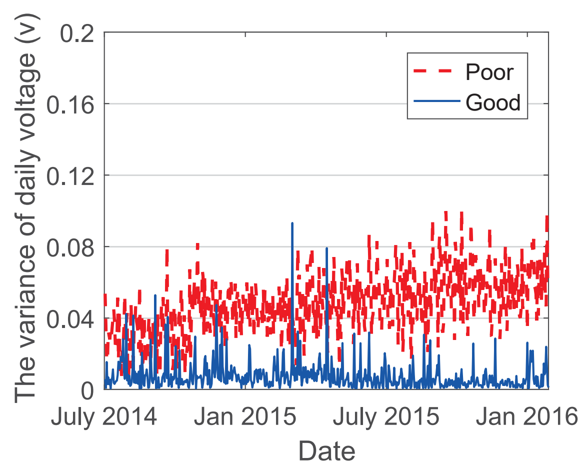


Figure 5: Voltage variances versus battery status

Fig. 6 illustrates that the voltage variance has the correlation with the length of remaining lifetime, indicating that the variance of the output voltage from the batteries over time also reflects the aging trend of battery quality degradation. These observations motivate us to predict battery working conditions based on the battery historical voltages.

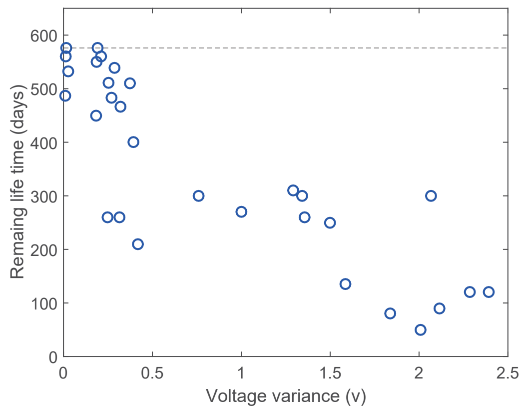


Figure 6: Correlation between the remaining life and voltage variance

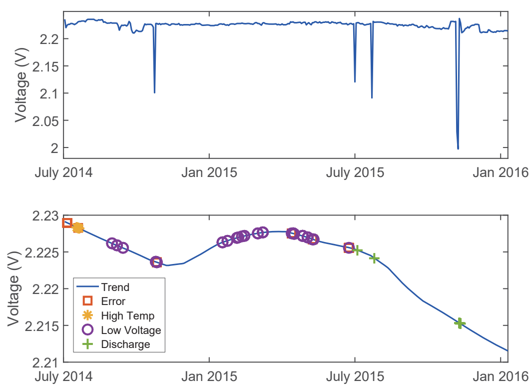


Figure 7: Example of time series showing battery aging phenomenon with status

2.2.2 *Battery Status*

We perform an analysis on the battery status in the logs to explore their potential relationships with the working conditions of the batteries. Fig. 7 shows an example of an eighteen-month-long time series voltage collected from one cell battery. This cell suffered a deep battery discharge at the beginning of July 2015, and another started in August. This leads to the degradation of the whole battery pack, yet such the aging status are too messy and excessive to be captured. Tab. 1 lists the number and percentage of some selected status categories, and Fig. 8 also shows the frequency distribution among all the 105 categories. We can see that the distribution is highly skewed: the most popular category is Alert, at about 28.09%, the second is Faulty cell, at about 20.42%; and the third is Discharging, at about 10.70%.

We take three status as examples to further investigate the correlations between status and battery remaining lifetime, which are *low float voltage*, *discharge* and *fault cell* shown in Fig. 9(a), (b) and (c), respectively. We count the specific status number for each battery until the batteries are replaced, and pick up the top-30 batteries with the maximum number of status. From July 28th, 2014 to February 17th, 2016, there are 576 days in our dataset, where the remaining

Rank	Category	Count	Percentage
1	Alert	1479914	28.09%
2	Faulty cell	1075533	20.42%
3	Discharge	563469	10.70%
4	Low float voltage	457767	8.69%
5	Too high	292966	5.56%
6	Low	263625	5.00%
14	Power outage	48605	0.92%
26	Failure	9100	0.17%
27	Voltage too low	7060	0.13%
30	Low cell voltage	2830	0.05%
34	Generator on charge	960	0.02%
66	No power	23	4.36E-06
69	AC out	10	1.89E-06
70	Generator running	10	1.89E-06

Table 1: LIST OF BATTERY STATUS CATEGORIES

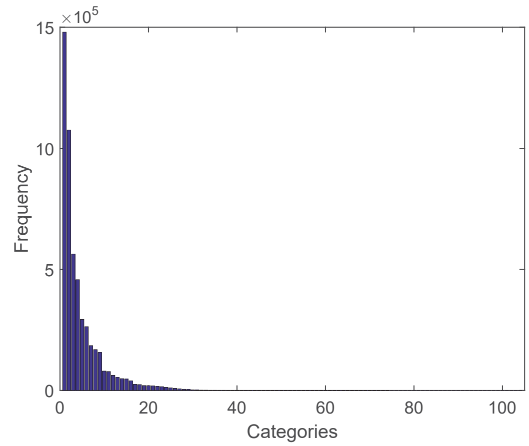
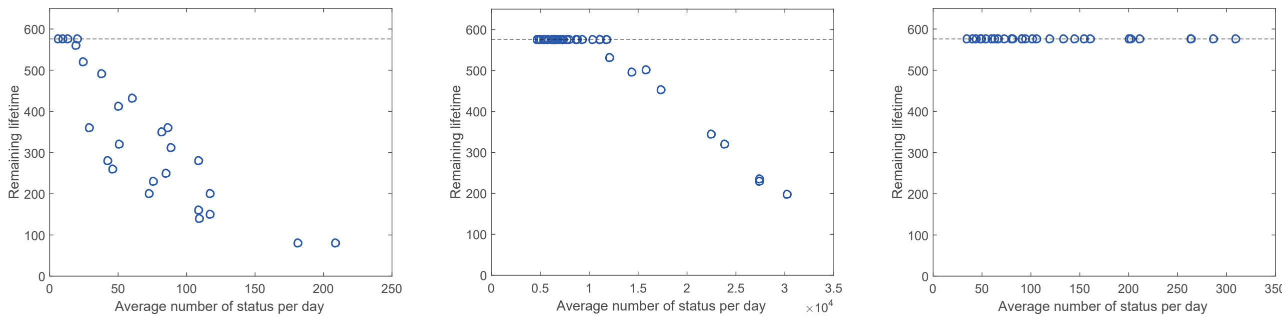


Figure 8: Distribution of battery status categories

lifetime of most batteries in our dataset is longer than 576 days, therefore dash lines represent that those batteries on it have longer remaining lifetime than 576 days. Fig. 9(a) and (b) plot the correlation between *low float voltage*, *Discharge* and remaining lifetime. This clearly demonstrates that there exists a strong correlation between battery remaining lifetime and *low float voltage*, as well as between battery remaining lifetime and *Discharge*. We further plot the remaining lifetime against the number of *fault cell* status in the system in Fig. 9(c), which does not show any noticeable correlation between them. This implies that the remaining lifetime is only affected by some specific status. The observations suggest that the diverse status have different influences on the battery working conditions, thus it is necessary to discriminatingly differentiate these status for the accurate prediction.

3. BATPRO: DESIGN SKETCH

Based on our measurement and analysis, the battery working conditions are correlated with the historical voltage and status logs. Yet there still remain several challenges to precisely predict the battery lifetime, especially when there are massive status in logs with the noises existing. In this section, we propose BatPro, a battery profiling framework, to predict the battery working conditions.



(a) Low float voltage status (b) Discharge status (c) Fault cell status

Figure 9: Correlation between the remaining life and the number of different status

3.1 Problem Statement

Our observations suggest that the working condition of batteries can be predicted by its relevant historical voltage and status, which motivates us and serves as the basis for the prediction of battery remaining lifetime. Before we proceed with the detailed solutions for the BatPro framework, we first summarize the key notations in Tab. 2. The monitoring systems collect raw messages from all batteries with the voltage and status, and we select the float voltage from the table named historybattery, where each v_i^t having the voltage value of battery i at t time. Now we have voltage set $\mathcal{V} = \{\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_N}\}$ and status set $\mathcal{X} = \{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}\}$, where N is the number of batteries. Each voltage records $\mathbf{v_i} = \{v_i^1, v_i^2, ..., v_i^T\}$ is a sequence of voltage records in time series for battery i, where v^T is a value of voltage at time T. Our goal is to extract and profile the aging trends for the working conditions of batteries, and predict the remaining lifetime of batteries. Since the remaining lifetime is predicted base on the voltage as mentioned, the voltage aging extraction problem can be defined as follows:

Given the battery data with voltage $\mathcal{V}$ and historical status $\mathcal{X}$ from time slot 1 to T, we formulate this problem to predict the remaining lifetime of batteries based on the predictive voltage. Let $\mathbf{v_i}' = \{{v'}_i^{T+1}, ..., {v'}_i^{T+K}\}$ define the predictive voltage in the following K time slots, which is compared with the real-world voltage $\mathbf{v_i} = \{v_i^{T+1}, v_i^{T+2}, ...v_i^{T+K}\}$. The whole objective function can be written the following way:

$$\min \sum_{i=1}^{N} \sum_{j=T+1}^{T+K} (v_i^j - {v'}_i^j)^2 \tag{1}$$

With the domain knowledge and our observations that the voltage is the criterion of the battery condition, the remaining lifetime can be easily predicted with the voltage threshold θ for pre-defined battery failure.

3.2 BatPro Framework

To solve the problem, we propose the BatPro framework to predict the working conditions of batteries shown in Fig. 10, which consists of the following stages: data preparation, voltage decomposition, voltage prediction and remaining lifetime prediction. In the BatPro framework, we first filter out a large amount of noise in historical voltage and only utilize the float voltage to reduce the interference on the voltage from other batteries activities, e.g., battery discharging. We also pull the relevant textual status from the database, as mentioned in the prior section. While the historical status are correlated with working conditions of batteries, a single status record is not a reliable factor for the prediction. Therefore, the BatPro framework decomposes the historical float voltage in time series into the aging and fluctuation terms. In the aging term, we propose a concept of penalty to describe the influences of a group of status on the voltage aging trend, which will be further illustrated later. Based on the voltage aging term, the framework utilizes the learning algorithm with status data to predict the future aging trend, which is further combined with the predictive voltage fluctuation term by ARIMA. With the predictive voltage in coming months, the BatPro framework can then estimate the remaining lifetime of batteries.

3.3 Profiling Strategy

N	the number of batteries
$\mathcal{V}$	voltage space
$\mathbf{v_i}$	the voltage of battery i
$\mathbf{v_i}_a$	the aging term of battery i
$\mathbf{v_i}_f$	the fluctuation term of battery i
v_i^t	the voltage of battery i on time t
l	the length of time segment
v_{ia}^k	the aging term of battery i on time segment k
s_i^k	the voltage slope of battery i on time segment k
y_i^k	the voltage penalty of battery i on time segment k
$\mathcal{X}$	status space
$\mathbf{x_i}$	a status set for battery i
$x_{i,j}$	j-th status in i-th battery group
m_i	the number of status for battery group i
$\mathcal{Y}$	penalty labels space
$\mathbf{y_i}$	a set of labels for battery i
$y_{i,j}$	j-th penalty label in i-th battery group
n_i	the number of penalty label for battery group i
θ	pre-defined battery failure threshold
r_i	remaining life time for battery i

Table 2: Summary of Notations

As mentioned in Fig. 10, we decompose a given time series $\mathbf{v}$ into the aging term $\mathbf{v}_a$ and the fluctuation term $\mathbf{v}_f$, as following shows:

$$\mathbf{v} = \mathbf{v}_a + \mathbf{v}_f \tag{2}$$

To ensure that the trend v_a is monotonic, we can simply write the constraint of the objective function as:

$$v_a^t \geq v_a^{t+1} \tag{3}$$

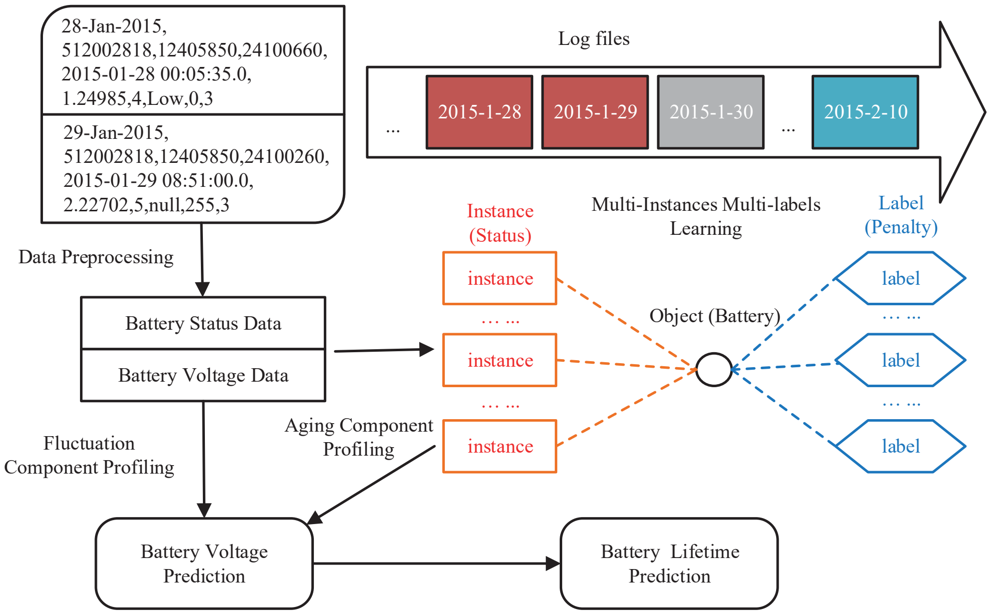


Figure 10: Overview of the BatPro framework

Let v_f^t describe the variance of the voltage, which is increasing fluctuation with the growth of t. We further extend the objective function of the optimization problem in Eq. 1 as minimizing the reconstruction error $||\mathbf{v} - \mathbf{v'}_a - \mathbf{v'}_f||^2$ and ensuring flatness of the fluctuation term over time, which is subject to the monotonicity constraint in Eq. 3. In the next subsections, we predict the voltage based on aging term and fluctuation term separately, and combine them to estimate the battery remaining lifetime.

3.3.1 Aging Term

To extracting the voltage aging term and make the fluctuation term stationary, we segment the float voltage in time series, and ensure the mean value difference of each segment to be as small as possible. More specifically, given the voltage $\mathbf{v_i}$ in a time series, we partition it into segments as shown below:

$$\mathbf{v_i} = \{\underbrace{v_i^1, v_i^2, ..., v_i^l}_{1st\ seg}, \underbrace{v_i^{l+1}, ..., v_i^{2l}}_{2nd\ seg}, ..., \underbrace{v_i^{(k-1)l+1}, ..., v_i^{kl}}_{kth\ seg}, ...\} \quad (4)$$

where l is the length of the segment and each segment k has initial value v_{ia}^k. We use the polynomial regression to fit each segment and the slope value s_i^k to define the voltage aging term in time segment k. The aging term can be defined as:

$$\mathbf{v_i}_a = \{(v_{ia}^1, s_i^1), (v_{ia}^2, s_i^2), ..., (v_{ia}^k, s_i^k), ..\} \quad (5)$$

Then we propose a concept voltage penalty y_i^k to represent the acceleration on the aging term:

$$y_i^k = s_i^k - s_i^{k-1} \quad (6)$$

For battery i in time segment k, we have the aging term (v_{ia}^k, s_i^k) as well as y_i^k. We utilize the historical battery status in the logs to predict the penalty y_i^{k+1} for the aging term $\mathbf{v_i}_a$. Given the training examples with a collection of historical battery status and voltage penalty, we can formulate a predictive problem as constructing a status classifier for predicting voltage penalty. One of the major challenges is that the monitoring system collects the status for each battery group as mentioned in Section 2, where the cell battery in the group has the exactly same status log, despite the cell battery has the unique voltage data. Thus we predict the voltage penalty for a battery group rather than every cell battery, while the remaining life time of cell battery can be estimated based on the voltage penalty value of the battery group.

The problem can be transformed to a MIML problem [24], where each training example is associated with not only multiple instances but also multiple labels. Instead of receiving a set of independently labeled instances as in the standard classification, our MIML model shown in Fig. 10 receives a set of bags which are simultaneously associated with multiple labels. In our problem, each status is regarded as an instance and a voltage penalty is defined as a label. The concept bag denotes a set of status $\mathbf{x_i}$, where each bag $\mathbf{x_i}$ has m_i instances $\mathbf{x_i} = \{x_{i,1}, x_{i,2}, ..., x_{i,m_i}\}$. Given the bags obtained from different batteries, the goal is to build a classifier that will label the unseen bags with a voltage penalty correctly.

Formally, let $\mathcal{X}$ denote the instance space and $\mathcal{Y}$ represent the set of labels. We denote the training data by $\{(\mathbf{x_1}, \mathbf{y_1}), (\mathbf{x_2}, \mathbf{y_2}), ..., (\mathbf{x_N}, \mathbf{y_N})\}$ that consist of N examples, where $\mathbf{y_i}$ is the penalty label associated with $\mathbf{x_i}$. As mentioned, each battery has n_i cells with different voltage aging trend, where $\mathbf{y_i} = \{y_{i,1}, y_{i,2}, ..., y_{i,n_i}\}$ is a subset of all possible labels and $\mathbf{y_i} \in \mathcal{Y}$. Then the objective is to learn a function $f_{MIML} : 2^{\mathcal{X}} \to 2^{\mathcal{Y}}$ from a given data set $\{(\mathbf{x_1}, \mathbf{y_1}), (\mathbf{x_2}, \mathbf{y_2}), ..., (\mathbf{x_N}, \mathbf{y_N})\}$ to accurately predicts the penalty label $\mathbf{y_i}$ for each bag $\mathbf{x_i}$.

Based on MIMLfast [11], we propose a modified approach with an effective approximation of the original MIML problem. Specifically, to utilize the relations among multiple labels, MIMLfast first learns a shared space for all the labels from the original features, and then trains label specific linear models from the shared space. To identify the key instance to represent a bag for a specific label, the classification model is trained on the instance level, and then select the instance with maximum prediction. To make the learning efficient, stochastic gradient descent is used to optimize an approximated ranking loss. In the testing phase, different with MIMLfast which returns a subset of all possible labels with the prediction value, the BatPro framework obtains the penalty label for each bag by selecting the one with the largest prediction value. This modification enables us to easily find the unique penalty label for each battery with the highest confidence score.

3.3.2 Fluctuation Term

Voltage variance is also a significant feature in predicting the battery life. The fluctuation of float voltage is not easy to capture from the original time series using human eyes. With the domain knowledge and regular maintenance policies, we know that most batteries chronically keep in float-charging status, which periodically make batteries charge and discharge with weak current among the fixed intervals. Time series approach is suitable for describing such kind of stochastic process and also easy to establish the forecasting model to forecast the fluctuation term of voltage. ARIMA [2] is one of the most popular time series models for predicting future values of a time series, which has been widely used in financial, economic and social scientific fields. ARIMA model can be represented as a linear combination of the past observations and past errors, which consists of three parts, i.e., an Autoregressive (AR) model, a Moving Average (MA) model and an integrated part. The ARIMA(p, q) model is written as follows:

$$Y(t) = \sum_{i=1}^{p} \beta_i Y(t-i) + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \tag{7}$$

where Y(t) is the number of views in the time t. $\beta_1, ..., \beta_p$ are the parameters of Autoregressive (AR) model, and $\theta_1, ..., \theta_q$ are the parameters of the Moving Average (MA) model. $\epsilon_1, ..., \epsilon_{t-1}$ are white noise error terms, which are generally assumed to be Gaussian random variables with zero mean and constant variance. In our BatPro framework, we utilize a partial ARIMA model to represent the fluctuation term.

$$v_f^t = \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \tag{8}$$

Given the time series of float voltage $\mathbf{v_i}$ in the past several days, it can make a fine-grained prediction for the voltage's future evolution by leveraging the trend, periodicity and autocorrelation exhibited in the history information. The BatPro framework can thus use it to extract and forecast the fluctuation term $\mathbf{v_i}_f$ based on the voltage $\mathbf{v_i}$.

3.3.3 Remaining Lifetime

With the voltage aging trend term $\mathbf{v}_a$ and fluctuation term $\mathbf{v}_f$, we can predict the future voltage $\mathbf{v}'$ by Eq. 2. With the domain knowledge and our observations that the voltage is the criterion of the battery condition, we say the working condition of battery i is good, when the float voltage value is higher than pre-defined threshold θ, which can be written as follows:

$$v_{ia} + v_{if} > \theta \tag{9}$$

With the voltage $v_a^t + v_f^t$, the aging term s^t and penalty y^t of the voltage at t time, thus the remaining lifetime r_i for battery i can be calculated as follows:

$$r_i = \frac{v_{ia} + v_{if} - \theta}{s_i + y_i} \tag{10}$$

4. EXPERIMENTS AND DISCUSSION

This section presents the evaluation of our BatPro framework. The BatPro framework provides the precise prediction for battery working conditions, which is evaluated against three well-known prediction approaches in time series, i.e., Linear Regression [3], ARIMA [2] and Wavelet [7].

4.1 Experiment Setup

Figure 11: Clusters for Hive-based data processing

The massive data analysis requires lots of time and capacity, therefore we move the original 323 GB data from Sybase ASE Database[3] onto our cloud platform to make querying over the archived data efficient, which is shown in Fig. 11. Our platform consists of eight Dell PowerEdge R430 Server, each equipped with two Xeon E5-2630 v3 2.4GHz 8 core, 256 GB DDR3 RAM and a 10 Gbits/sec Network Interface Card (NIC). Hyper-Threading is enabled for the CPU so that each CPU core can support two threads (virtual CPU core). All the physical servers are inter-connected through a NETGEAR 16-port 10-gigabit switch. We use a dedicated physical machine to host the master node, rather than a virtual machine, so as to ensure fast response time with minimized resource contention. Other physical machines that host virtual slave nodes run the latest Xen Hypervisor (version 4.1.3). For the operating systems running on VMs (both Domain0 and DomainU), we use the popular Ubuntu 12.04 LTS 64 bit (kernel version 3.11.0-12). All the VMs are allocated 8 virtual CPU cores and 4 GB memory. We run Hive 1.2.1 [19], a data warehouse infrastructure built on top of Apache Hadoop 1.2.1 and suitable for data management and analytical queries with HiveQL, an SQL-based query language.

4.2 Results

In this section, we evaluate the prediction accuracy of BatPro framework using real-world trace collected from China Mobile and demonstrate the performance in estimating the

[3]www.sybaseproducts.com/

	40	60	80	100	120
BatPro	0.0030	0.0083	0.0053	0.0098	0.0060
ARIMA	0.0154	0.0222	0.0290	0.0390	0.0451
LR	0.0372	0.0386	0.0387	0.0398	0.0400
Wavelet	0.0870	0.0868	0.0787	0.0794	0.0798

Table 3: RMS error between actual and predictive value

remaining battery lifetime. To evaluate the prediction quality, we run the experiment on the real-world data with 1691 batteries data, 1282 examples in the training set and 409 examples in the test set.

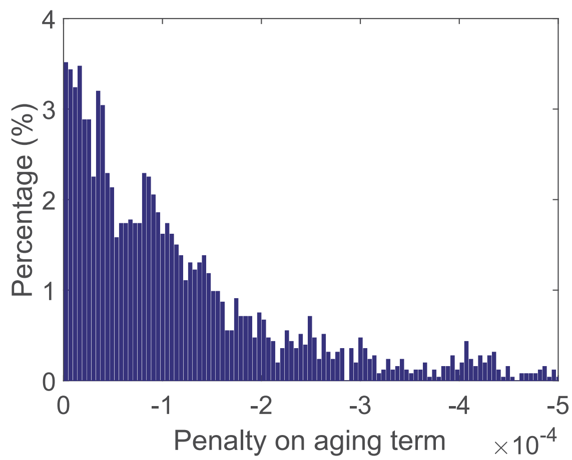


Figure 12: The distribution of penalties on aging term

We calculate the differences between the slopes of the prior 365 days and latter 120 days to generate a penalty labels for the battery aging terms. Fig. 12 summarizes the distribution of the penalty trends that are extracted from the dataset, where the range is from 0 to -5×10^{-4} and most batteries are in the slow deterioration status. In Tab. 3, we compare the BatPro approach with Linear Regression (LR) [3], ARIMA [2] and Wavelet [7] and compute the root-mean-square error between the predictive voltage value and actual data. Our BatPro approach can precisely predict the battery voltage with the RMS error less than 0.01 v. As Fig 13 shows, we can see that our method performs best among the majority of the compared schemes with the smallest error between the extracted trend and the ground-truth trend. Especially when the slope is small and the aging phenomenon is tiny, our method performs much better than LR, ARIMA, and Wavelets. Moreover, the aging trend extracted by our scheme is monotonic and satisfies the nature of the aging behavior, while LR, ARIMA, or Wavelets does not have such advantage as they have no monotonic constraint in extracting the trend.

As the battery replacement is the consequence of a battery failure [18], we choose 112 batteries with the replacement records in the log to verify the prediction accuracy for remaining lifetime. We count the survival days for each battery after the failure alert by the different approaches, i.e., BatPro, ARIMA, Wavelet and Linear Regression. Fig. 14 plots the relation between the percentage and the survival days after the failure alert. We can see that 87% batter-

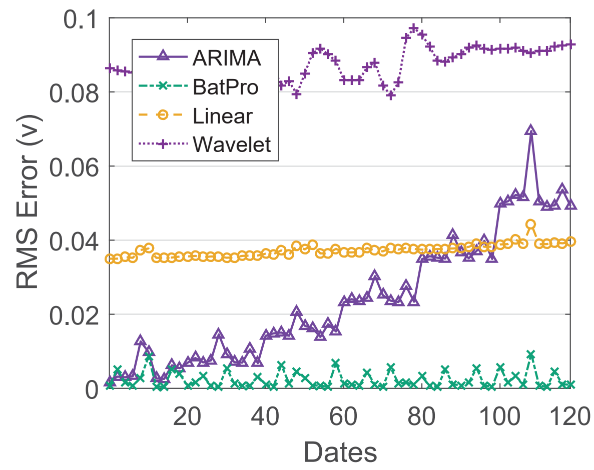


Figure 13: Root-mean-square Errors

ies are replaced in three months after BatPro's alters, which demonstrates the BatPro framework is a strong predictor for battery working conditions. We also observe that there are still a small number of batteries not replaced after our failure alerts. We have a close look at those batteries, and find out there are redundant battery groups in their mobile network base stations. Therefore repairing engineers postponed the maintenance service for those batteries, which also demonstrates the BatPro framework can help the maintenance engineers to detect the potential issues in the battery groups.

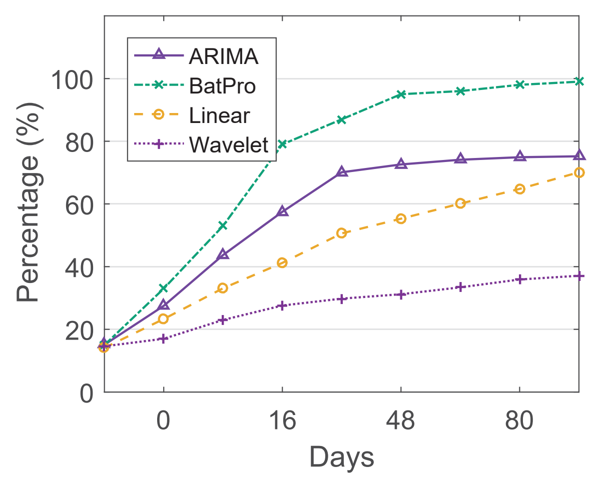


Figure 14: Survival days after our failure alert

4.3 Case Study: Insight into a Battery

In this section, we select one instance, i.e., one battery, from the large-scale experiments as a case study to intuitively demonstrate the effectiveness of the BatPro framework. We applied BatPro and ARIMA approach on the 365 days battery data to predict the voltage and battery working conditions in the future 120 days. The result is shown in Fig. 15, which is consistent with our assumption that status could bring permanent impacts to battery working conditions. ARIMA only uses single battery's float voltage, and it cannot model the effects of historical status. Taking account of the effects of status, our BatPro framework is able to quickly respond to those status and provides an accurate predictive trend. In Fig. 15, we can clearly see that the BatPro framework can successfully predict the decreasing trend of battery voltage and predict the working condition of this battery in the next three months. As we know, the major drawback of ARIMA model is that it needs a relatively long period of history information for prediction. In the experiment, ARIMA works well to predict the fluctuation terms based on the voltage data in prior 365 days, although it cannot estimate the aging terms accurately. With a predefined threshold of the battery, we can then easily predict the battery remaining lifetime, as demonstrated in the previous subsection.

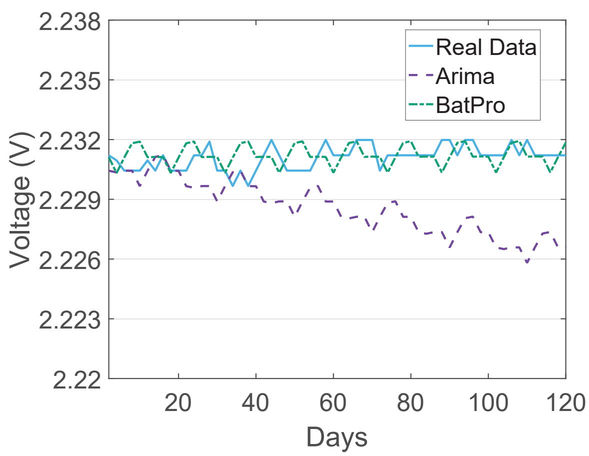


Figure 15: Voltage predict

5. RELATED WORK

In this section, we discuss two categories of researches that are closely related to our work, namely, the prediction in time series and the analysis of the log status.

5.1 Time Series Analysis

The time series research has attracted significant efforts in recent decades, due to its importance to all technical systems. Here, we discuss only some recent and relevant works. One of the simplest approaches is smoothing or filtering, which can be done using different techniques, including wavelets [7] and moving average processes [9]. Another intuitive approach for trend extraction is fitting a linear or higher level model to the time series data, while the common way of fitting a model is a method of least squares [3]. AutoRegressive Integrated Moving Average (ARIMA) [2] is a model-based technique that allows extracting trend and seasonal components from the time series. This approach can be used for both modeling and forecasting, and it is especially useful when there is a certain seasonal component. The limitation of this model is that the parameters must be provided by the user, which may be very subjective and lead to inadequate results. Singular Spectrum Analysis (SSA) [5] is a parameter free decomposition techniques for time series, which allows the extraction of the alleged trend, seasonal and noise components from time series.

Djurdjanovic et al. [4] proposed a comprehensive framework named Watchdog agent for analysis and monitoring of the systems based on the time series analysis. Their focus was more on the methodology of data collection and analysis rather than the methods of decomposing time series and extracting necessary features. Luo et al. [15] presented an approach to find the correlation between actual status and time series in order to diagnose incidents. Their approach matched the status with certain subsequences of time series in order to give a real explanation of the time series shape. Herodotou et al. [10] proposed a probabilistic method in data centers failure monitoring, which is domain specific and initially builds a network model. Recently, a novel time series analysis technique proposed by Ulanova et al. [20], which allows the decomposition of the time series into trend and fluctuation components of complex physical systems.

5.2 Status in Logs

There are many performance diagnosing techniques that rely on specific types of data, where logs can be used to further understand the system conditions and analyze the common patterns. Li et al. [14] is aimed at discovering and understanding the common patterns, but did not touch the proactive prevention of errors. Gu et al. [8] proposed an online failure forecast system by observing the stream system's status, which uses an ensemble of decision tree classifiers and is applicable in an online setting. Fu et al. [6] used finite state automata to model sequential dependencies between messages and detect anomalies. Makanju et al. [16] proposed IPLoM by creating status descriptions based on clustering text messages in the logs, but interpreting them requires unavailable domain knowledge. Xu et al. [23] [22] combined textual messages in console logs to construct performance features and conducted the Principal Component Analysis (PCA) [12] to detect anomalies. Similarly, Nagaraj et al. [17] modeled status logs and state logs into performance features to infer the anomalies between components. Kimura et al. [13] proposed a modeling and status extraction method of network log data using a tensor factorization approach, yet they did not focus on detection of anomalies. Most recently, Sipos et al. [18] considered log-based predictive analysis in order to monitor the conditions of the operating equipment, yet they ignored the possibility of fine temporal patterns to simplify the model.

Different with the prior approaches, the BatPro framework focuses on a large amount of batteries and utilizes not only the signal value but also the status data to comprehensively predict the battery working conditions.

6. CONCLUSION

In this paper, we proposed BatPro, a battery profiling framework, to precisely extract the features that cause the working condition degradation of the battery group. Based on the framework, we formulated the prediction models for both battery voltage and lifetime and proposed a series of solutions to yield accurate outputs. By real world trace-driven evaluations, we demonstrate that our BatPro approach can achieve much higher prediction accuracy on the battery voltage and lifetime, which can serve as a basis towards a cost-effective battery maintenance on minimizing the service interruptions in mobile network base stations.

7. ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for valuable and insightful comments. This research is supported in part by an NSERC Discovery Grant, a Strategic Project Grant, an RTI grant, and an E.W.R. Steacie Memorial Fellowship. Feng's work is partly supported by the Start-up Grant from the University of Mississippi and NSF I/UCRC 1539990.

8. REFERENCES

[1] Balshe, W. Power system considerations for cell tower applications. *Cummins Power Generation* (2011).

[2] Box, G. E., Jenkins, G. M., and Reinsel, G. C. Linear nonstationary models. *Time Series Analysis, Fourth Edition* (1976), 93–136.

[3] Charnes, A., Frome, E., and Yu, P.-L. The equivalence of generalized least squares and maximum likelihood estimates in the exponential family. *Journal of the American Statistical Association 71*, 353 (1976), 169–171.

[4] Djurdjanovic, D., Lee, J., and Ni, J. Watchdog agent–an infotronics-based prognostics approach for product performance degradation assessment and prediction. *Advanced Engineering Informatics 17*, 3 (2003), 109–125.

[5] Elsner, J. B., and Tsonis, A. A. *Singular spectrum analysis: a new tool in time series analysis*. Springer Science & Business Media, 2013.

[6] Fu, Q., Lou, J.-G., Wang, Y., and Li, J. Execution anomaly detection in distributed systems through unstructured log analysis. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on* (2009), IEEE, pp. 149–158.

[7] Grossmann, A., and Morlet, J. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM journal on mathematical analysis 15*, 4 (1984), 723–736.

[8] Gu, X., Papadimitriou, S., Yu, P. S., and Chang, S.-P. Online failure forecast for fault-tolerant data stream processing. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on* (2008), IEEE, pp. 1388–1390.

[9] Hamilton, J. D. *Time series analysis*, vol. 2. Princeton university press Princeton, 1994.

[10] Herodotou, H., Ding, B., Balakrishnan, S., Outhred, G., and Fitter, P. Scalable near real-time failure localization of data center networks. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (2014), ACM, pp. 1689–1698.

[11] Huang, S.-J., and Zhou, Z.-H. Fast multi-instance multi-label learning. *arXiv preprint arXiv:1310.2049* (2013).

[12] Jolliffe, I. *Principal component analysis*. Wiley Online Library, 2002.

[13] Kimura, T., Ishibashi, K., Mori, T., Sawada, H., Toyono, T., Nishimatsu, K., Watanabe, A., Shimoda, A., and Shiomoto, K. Spatio-temporal factorization of log data for understanding network events. In *INFOCOM, 2014 Proceedings IEEE* (2014), IEEE, pp. 610–618.

[14] Li, T., Liang, F., Ma, S., and Peng, W. An integrated framework on mining logs files for computing system management. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (2005), ACM, pp. 776–781.

[15] Luo, C., Lou, J.-G., Lin, Q., Fu, Q., Ding, R., Zhang, D., and Wang, Z. Correlating events with time series for incident diagnosis. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (2014), ACM, pp. 1583–1592.

[16] Makanju, A. A., Zincir-Heywood, A. N., and Milios, E. E. Clustering event logs using iterative partitioning. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (2009), ACM, pp. 1255–1264.

[17] Nagaraj, K., Killian, C., and Neville, J. Structured comparative analysis of systems logs to diagnose performance problems. In *Presented as part of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)* (2012), pp. 353–366.

[18] Sipos, R., Fradkin, D., Moerchen, F., and Wang, Z. Log-based predictive maintenance. In *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining* (2014), ACM, pp. 1867–1876.

[19] Thusoo, A., Sarma, J. S., Jain, N., Shao, Z., Chakka, P., Anthony, S., Liu, H., Wyckoff, P., and Murthy, R. Hive: a warehousing solution over a map-reduce framework. *Proceedings of the VLDB Endowment 2*, 2 (2009), 1626–1629.

[20] Ulanova, L., Yan, T., Chen, H., Jiang, G., Keogh, E., and Zhang, K. Efficient long-term degradation profiling in time series for complex physical systems. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2015), ACM, pp. 2167–2176.

[21] Vetter, J., Novák, P., Wagner, M., Veit, C., Möller, K.-C., Besenhard, J., Winter, M., Wohlfahrt-Mehrens, M., Vogler, C., and Hammouche, A. Ageing mechanisms in lithium-ion batteries. *Journal of power sources 147*, 1 (2005), 269–281.

[22] Xu, W., Huang, L., Fox, A., Patterson, D., and Jordan, M. Online system problem detection by mining patterns of console logs. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on* (2009), IEEE, pp. 588–597.

[23] Xu, W., Huang, L., Fox, A., Patterson, D. A., and Jordan, M. I. Mining console logs for large-scale system problem detection. *SysML 8* (2008), 4–4.

[24] Zhou, Z.-H., Zhang, M.-L., Huang, S.-J., and Li, Y.-F. Multi-instance multi-label learning. *arXiv preprint arXiv:0808.3231* (2008).