Coordinated Priority-aware Charging of Distributed Batteries in Oversubscribed Data Centers

Sulav Malla[*†] Qingyuan Deng[*] Zoh Ebrahimzadeh[*] Joe Gasperetti[*]
Sajal Jain[*] Parimala Kondety[*] Thiara Ortiz[*] Debra Vieira[*]
[*]*Facebook, Inc.* [†]*University of South Florida*
{sulav, qdeng, zoh, jgasperetti, sajalj, pkondety, thiaraortiz, sparkeygrl}@fb.com

***Abstract*—Data centers employ batteries for uninterruptible operation during maintenance and power failures, for example, when switching to diesel generator power after a utility power failure. Depleted batteries start to recharge once the input power is back, creating a sudden power spike in the power hierarchy. If not properly controlled, a sustained power overload can potentially trip circuit breakers, leading to service outages. Power overloads due to battery recharging are even more likely in oversubscribed data centers where the power infrastructure is aggressively provisioned for high utilization. The problem caused by simultaneous recharging of batteries in a data center has not been extensively studied and no real-world solutions have been proposed in the literature.**

In this paper, we identify the problem due to battery recharging with case studies from Facebook's data centers. We describe the solutions we have developed to coordinate charging of batteries without exceeding the circuit breaker power limit. We explain in detail, the variable battery charging algorithm built into the distributed battery charger hardware deployed in Facebook data centers, and the system design considerations necessary on a large scale. The new variable charger is able to reduce battery recharge power by up to 80%. We further leverage individual battery charging control mechanism to coordinate the charging process such that we charge the batteries according to priorities of applications running on the servers supported by the batteries. We evaluate our coordinated priority-aware battery charging algorithm by building a prototype in a Facebook production data center as well as through simulation experiments using production power traces. Our results show that we are able to meet reliability service level agreements by using our battery recharging algorithm, while satisfying given power constraints.

I. Introduction

The increasing popularity of Internet services such as social media, search, online shopping, and content streaming, as well as the industry's migration towards cloud-based services has accelerated the demand for data centers. Data centers need to be highly reliable as services running on them are expected to be always available and a power outage can incur significant service interruption [17]. Indeed, redundant backup power infrastructure, such as diesel generators and uninterruptible power supply (UPS) with batteries, are built into the data center to ensure that power to the IT equipment (servers, storage, and network switches) inside the data center is always available.

In case of a utility power failure, the switch over from utility to diesel generator generally takes 10 to 20 seconds. Such brief power failures also happen during maintenance, when switching from a normal power source to a reserve power source, in data centers with redundant power supply devices for higher reliability (tier III/IV data centers [3]). Batteries are designed to power the IT equipment without any interruption during such transitions. Batteries may be centrally located at the generator level or distributed across the power hierarchy as in the case with modern data centers. For example, Facebook [37], Google [50], and Microsoft [12], [44] data centers have distributed batteries at the rack level. Advantages of having distributed batteries, in contrast to having a centralized UPS is that they are more efficient (no AC-DC-AC double conversions), more fault tolerant (no single point of failure), and scale naturally with the number of IT equipment/racks deployed. Batteries ensure that there is no downtime during brief power source switches or failures. Once the input power is back, the depleted batteries start to recharge. The additional power draw to recharge the battery can create a sudden power spike in the data center power hierarchy. In the case of Facebook data centers, we have found that the battery recharge power spike can be up to 25% of the server power consumption and last for more than 30 minutes. If the power consumption by servers is already high, the power spike due to battery recharging can create a sustained *power overloading* (power draw exceeding the power limit) of circuit breakers, potentially tripping them. For example, a 30% power overdraw at a circuit breaker for more than 30 seconds could trip it [47] causing a serious power outage and service disruption. Interestingly, batteries which are supposed to prevent power outages can themselves be the cause of subsequent power outages.

An approach to account for battery recharge power is to statically allocate and reserve the power hierarchy for the anticipated worst-case battery recharge scenario. However, this is expensive (data center power infrastructure usually costs \$10–\$20 per Watt [10], [20] to build) and wasteful as we may have to allocate 25% of the data center power budget for battery recharge power which will be stranded most of the time. In contrast, since data center power is a scarce resource, power is generally oversubscribed to increase the average utilization of the power infrastructure and better amortizes the high capital expense as well as time overhead of building the data center [9], [21], [36], [47]. *Power oversubscription* refers to the deployment of more IT equipment under a circuit breaker than allowed by its power limit. The problem due to battery recharge is even more severe in oversubscribed data

centers where power utilization is generally high.

Another approach to this problem is to reduce the server power during the battery recharge period, for example, through power capping of servers [19], [33] or by suspending low priority jobs [36], to make room for the battery recharge power. However, such server power reduction leads to service performance degradation. Alternative to server power capping, we have developed a new variable battery charging mechanism into our distributed batteries whereby the batteries recharge at a rate proportional to the energy discharged from them. In addition, we have also developed a control mechanism into the battery chargers to manually override the charging current. This allows software control policies to be developed for coordinated charging of distributed batteries.

Existing works that have explored battery charging and discharging in the data center have mostly focused on peak power shaving [1], [2], [5], [12], [13], [18], [26], [27], [40]. These works dual-purpose the battery, that was originally designed to provide backup power during brief power failures, to be used during peak power demand for cost savings. However, the problem due to battery recharging has been largely ignored in the literature. In this paper, we tackle the problem of battery recharging and make the following main contributions:

1) We identify the problem created by recharging of distributed batteries, especially in aggressively oversubscribed data centers.
2) We design a new variable battery charger hardware mechanism, which reduces the battery charging power significantly while meeting the charging time constraints.
3) We deploy and evaluate the variable battery charger in real production data centers.
4) Finally, we propose a first software-based coordinated priority-aware battery charging algorithm, which takes data centers' power availability as well as service priorities into consideration, and evaluate it by building a prototype as well as through simulations using real production power traces.

II. Background

In this section, we describe the data center power hierarchy and its oversubscription. We then define/discuss about open transition and how they result into the battery recharge problem along with a couple of case studies.

A. Data Center Power Hierarchy

We describe the power hierarchy of a typical data center at Facebook built according to the Open Compute Project [29] design. As shown in Fig. 1, power from the local utility arrives at an on-site substation where high-voltage to medium-voltage transformation takes place. Medium-voltage switch gears (MSG) then distributes power to different buildings (a site typically can have several data center buildings). Each data center building is designed for 30 MW of critical power (for IT load) and is divided into four suites (server rooms), each of 7.5 MW capacity, which has servers, storage, and networking

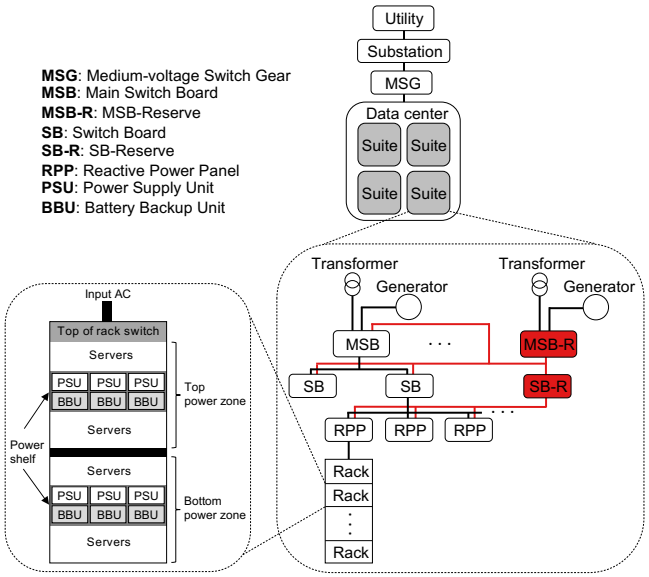

Fig. 1. Power hierarchy of a typical Facebook data center.

equipment arranged in rows of racks. A suite has multiple levels of circuit breakers, forming a power hierarchy tree, to distribute power to individual racks. At the top level is the main switch board (MSB) which has input from the utility (via a medium-to-low-voltage transformer), a backup diesel generator, and a reserve MSB (MSB-R). A typical suite is powered by several MSBs, each rated at 2.5 MW of critical power. The backup diesel generator is automatically used in case the utility power fails. The MSB-R is used in case an MSB needs to be disconnected from the critical power path for maintenance or in case the MSB fails. An MSB supplies power to 2 to 4 switch boards (SB) rated at 1.25 MW of critical power. The MSB-R also powers the reserve SBs (SB-R). MSB-R and SB-R are redundant components (providing N+1 redundancy), generally used during maintenance to isolate other circuit breakers from the power path.

SBs in turn provide power to 190 kW reactor power panels (RPP) located at the end of each row which supplies power to the row through an overhead busway. Racks (rated at 12.6 kW) in a row connect to the overhead busway through a tap box. A rack has two independent power zones, each with a power self containing 3 power supply units (PSU). PSU converts the input AC power to DC power appropriate for the IT equipment in the rack. Each PSU is also connected to a battery backup unit (BBU) below it, which is utilized when the input power to the rack fails. PSU starts charging the discharged batteries once the input power is back. BBUs are sized to provide up to 90 seconds of power to the rack.

B. Power Oversubscription and Dynamo

Allocating the power budget of a data center according to the nameplate rating (maximum power draw) of servers is wasteful since servers do not consume peak power all the time. Server power varies with its utilization which generally

exhibit diurnal and weekly cycles. The probability of a group of servers reaching the aggregate peak power (every server consuming peak power simultaneously) is even lower because of statistical multiplexing of individual server power [8], [36]. Furthermore, data centers are expensive and time consuming to build. These factors motivate a data center operator to oversubscribe their facility. For example, Facebook data centers are built for 30 MW of IT load. Given that a rack is rated at 12.6 kW, a data center can accommodate only 2,380 racks. However, we aggressively deploy much more racks into these data centers due to the rapid growth in computing needs over the last several years. Looking at the 20 largest Facebook data centers, we find that we have, on average, 47% more racks inside the data center, with one data center oversubscribed by as high as 70%.

Power oversubscription enables efficient use of our data centers. However, there is always a risk of reaching different "choke points" of the power infrastructure [36] – for example, overloading circuit breakers (MSB, SB, and RPP) and tripping them. Some kind of peak power capping system is required to prevent peak aggregate power and ensure safe power oversubscription. In our data centers, we use Dynamo [47], a real-time power monitoring and control system, to prevent tripping of circuit breakers that leads to extended power outages. Dynamo monitors the power consumption of each server as well as all the circuit breakers at multiple levels in a data center. Upon detecting a power overload at a particular circuit breaker, Dynamo automatically caps the power consumption of servers (according to priority of services running on those servers) under the overloaded circuit breaker, protecting it from tripping. Interested readers can refer to [47] for further details regarding Dynamo.

C. Open Transition and Power Outage

Typical tier III and tier IV data centers (most of the commercial data centers, including Facebook data centers) have redundancy built into the power infrastructure for high availability of power [3]. In addition to diesel generators providing backup power during utility failure, each level of the power delivery hierarchy can have redundant components (an alternate power path) which enables maintenance of different power devices without any downtime. For example, at Facebook data centers, MSG are 2N redundant while MSB and SB are N+1 redundant. During the maintenance of an MSB (or SB/RPP), it is removed (de-energized) from the critical power path and replaced by MSB-R (or SB-R). The switch over from the primary power device to the reserve power device (and vice versa), causes a brief power unavailability for the subset of racks that draw power from the component undergoing maintenance. We refer to the short power unavailability, usually caused by the transfer of electrical power from one source to another, as *open transition*.

During an open transition, the input power is not available and IT equipment must rely on energy storage devices, such as batteries, for continuous power supply. Once the input power is back, the depleted batteries need to recharge which can be a source of sudden power spike in the power hierarchy, lasting for several minutes (as discussed next in our case studies). In concurrently maintainable data centers, there are open transitions due to planned preventive/corrective maintenance, or unexpected utility power failures. Open transition can occur at different levels (RPP, SB, MSB, MSG, etc.) in the power hierarchy, or even at multiple data center levels during utility failure or substation maintenance (transient loss of utility power or switching over from the utility feed to the generators). If a newly constructed data center building is being added to an existing site, MSG/substation level open transitions are usually needed on remaining buildings to "merge" the load. Furthermore, open transitions are frequent events. Power devices undergo annual preventive maintenance, as well as any corrective maintenance if an immediate fix is needed. For example, with hundreds of MSB across Facebook data centers, an MSB level open transition takes place almost every workday. Hence, open transitions are the norm rather than an exception in large scale data centers.

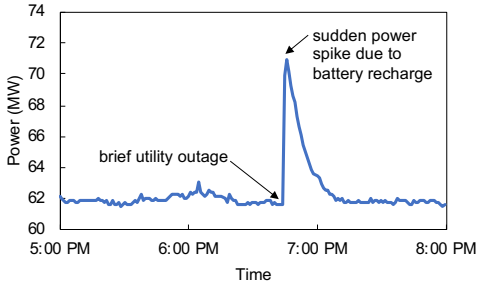

Fig. 2. A brief utility outage in a data center region causes the batteries in the entire region to recharge leading to a sudden power spike of 9.3 MW.

An open transition, which generally lasts for under a minute, is different from a power outage. *Power outages* are rare instances (such as, tripping of circuit breakers, or, diesel generators failing to start during an open transition) when the IT equipment in the racks lose power, potentially leading to service disruption. Similar to open transition, power outages can also occur at different levels in the power hierarchy, however, unlike open transition, power outages are production impacting incidents and generally last for hours.

D. Battery Recharging Case Studies

We discuss two recent events where battery recharging has caused problems in our data centers.

Case I – Utility outage during a thunderstorm: In August of 2019, a series of thunderstorms in one of the Facebook data center regions caused a brief utility power failure. The utility had a brief voltage sag for less than one second (a very short black out) so no diesel generators were engaged. However, during the brief loss of AC power, racks in three data centers fell back to battery power and started charging them after the utility power was back. Fig. 2 shows the total IT load of the region during the time period. The power consumption was 61.6 MW just before the utility outage. We can see in Fig. 2, a sudden power spike of about 9.3 MW (15% increase) due

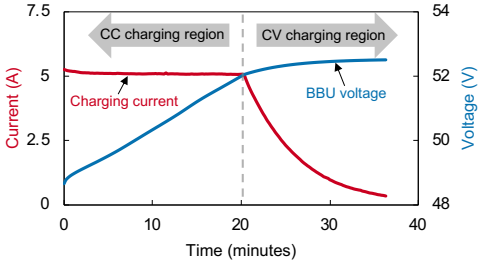

Fig. 3. Charging of a battery backup unit (BBU) after a full 90 seconds discharge.

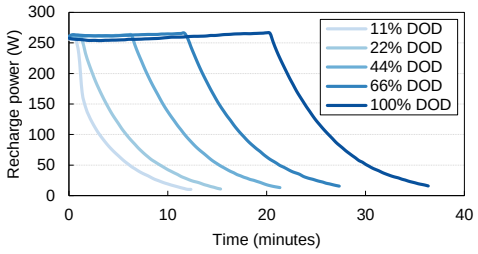

Fig. 4. Recharge power versus time for different depth of discharge (DOD) of the BBU.

to the recharging of batteries (we explain why this happens when we describe the battery charger in Section III) occurred immediately after the utility outage, which lasted for about 25 minute. Multiple circuit breakers within the three data centers reached their power limit, but Dynamo immediately capped servers under the overloaded circuit breakers to prevent them from tripping, preventing a potential regional outage.

Case II – Substation maintenance gone wrong: During a planned maintenance in September of 2019, at a substation of a data center region, maintenance personnel misidentified the component to work on and accidentally tripped a circuit breaker feeding power to one of the data center buildings. Due to the loss of utility feed, all MSBs within the data center carrying the IT load, automatically switched to their corresponding backup diesel generators. Battery recharging after the open transition caused the power on each MSB to suddenly increase by more than 20% from their existing levels. Once again, Dynamo had to step in and prevent a potential data center outage. More than ten thousand servers had to be power capped (causing large service degradation) to prevent circuit breakers at different levels from tripping.

In both of the case studies above, a recurring theme is that an open transition causes power consumption to spike due to recharging of batteries. The power spike is significant enough to trip circuit breakers. Fortunately, the safety net provided by Dynamo prevents such outages by power capping servers. However, throttling of servers causes degradation in their performance. In the next section, we discuss in detail the battery recharging process and a new variable charger we designed and deployed to tackle this problem. The new variable charger reduces battery recharging power according to the depth of battery discharge. It also allows software control of the battery charging rate, such that, in case of a power overload due to battery recharging, we can throttle the battery recharging power, rather than the server power, to prevent outages without any impact on service performance.

III. Battery Design

A. Original Battery Charger

The server racks at Facebook data centers are designed for a maximum IT load of 12.6 kW according to the Open Compute Project's Open Rack V2 design specification [30]. The rack is divided into two identical power zones, each of which is powered by 3 power supply units (PSU) connected to corresponding 3 battery backup units (BBU) in a 2+1 redundant architecture. BBUs are designed to handle open transitions (generally less than 45 seconds). When the rack input power is lost (mostly caused by open transitions), BBUs are discharged by the PSUs to maintain the IT load, for up to 90 seconds. The service level agreement (SLA) that the BBU needs to meet is to be able to power the racks for up to 90 seconds in the worst-case discharge scenario of peak power draw by the servers.

We use Li-ion batteries to meet our battery requirement, due to their high-power discharge, high energy density, and long cycle life [34]. The charging and discharging of BBU is done by the corresponding PSU. Li-ion batteries are generally charged using a two-step, constant current-constant voltage (CC-CV), method [39]. The battery is initially charged using a constant current (CC) up to a predefined voltage and then charged using a constant voltage (CV) until the charging current drops below a predefined current. Fig. 3 shows the charging process of a BBU after a full discharge[1]. We can see that the charger in the PSU initially charges the BBU at a constant 5 A charging current in the CC mode until the BBU voltage increases to 52 V. This takes about 20 minutes. The design choice of a constant 5 A charging current for the CC mode of BBU recharging was based on the fact that it is an ideal charge rate for the Li-ion cells and a simple Li-ion battery charger would suffice to charge the BBU. After the CC mode, the charger transitions to the CV mode where it maintains a constant voltage of 52.5 V until the charging current drops below 400 mA. The charging current decays rapidly in the CV mode and the entire charging sequence is complete in about 36 minutes as shown in Fig. 3.

The BBU recharge power profile over time for different depth of discharge (DOD) is shown in Fig. 4. As expected, the time to fully charge the BBU decreases with its DOD. Two interesting observations can be made here:

1) The decrease in total charge time is primarily due to the shortening of the CC phase of the charging process while the difference in time spent in the CV phase, for different DOD, is small (less than 4 minutes).

[1]We refer to the discharge of a BBU at 3,300 W of IT load for 90 seconds as a full discharge or 100% depth of discharge (DOD). In our lab experiments, we vary the DOD of BBU (energy discharged from BBU) by varying the time we discharge it.

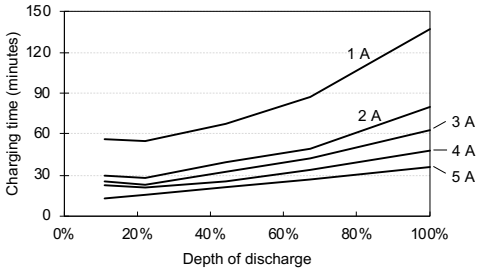

Fig. 5. BBU charging time versus depth of discharge for varying charging currents.

2) The initial charging power is about 260 W and is independent of the DOD of the battery. This is because the charger always starts charging in the CC mode before transitioning to the CV mode.

The initial recharge power for a rack can be up to 1.9 kW (6 BBUs per rack and including electrical losses), which is 25% of the power budget for most racks (all racks have an assigned statistically expected "power budget" according to the number/type of servers and services running on them, which is usually lower than the maximum power consumption of 12.6 kW). Furthermore, the fact that charging always starts at the maximum rate, even for short discharge (as shown in Fig. 4), result in the worst-case battery recharge power every time. Budgeting for the battery recharge power at each rack would require up to 25% more data center power capacity, which would have been left stranded most of the time.

In practice, we usually do not encounter a full BBU discharge due to (1) *shorter discharge time*: almost all open transitions are less than 45 seconds long, and (2) *lower discharge rate*: battery discharge rate depends on rack IT load during the discharge event which may be lower than the peak power. But the fact that battery recharge power peaks despite the DOD causes problems. Thus, we revisit our original battery charger design to develop a new variable charger which charges according to the DOD of the battery.

B. Variable Charger Design

A problem with the original battery charger is that the BBU is charged at the maximum power (5 A current) regardless of the amount of energy discharged from the BBU. Since a static power budget allocation for BBU recharge power is expensive and wasteful, we do not budget for the battery recharge power into the rack power budget. The power to recharge the BBU may not be available (due to high power utilization from IT load) at the time when the battery begins to charge after a power loss. The additional recharge power can overload the circuit breaker, leading to server power capping as illustrated in our case studies in Section II-D. Next, we explore charging batteries with lower charging current in an effort to decrease the battery recharge power.

In our lab, we experiment with different initial CC mode charging current of the BBU to study its impact on charging time and recharge power. We find that the maximum recharge

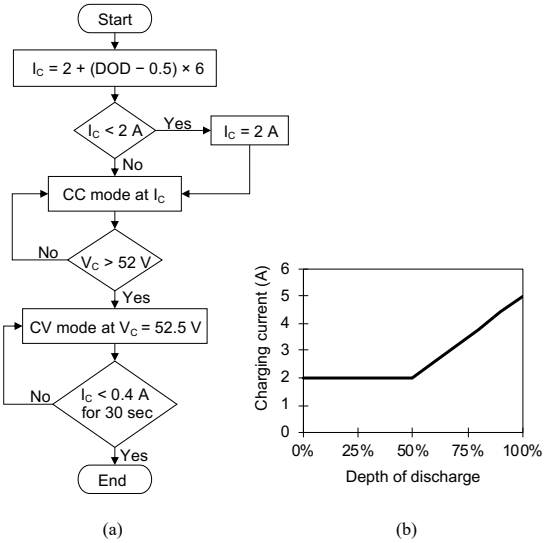

Fig. 6. (a) Flowchart of the variable current CC-CV charging logic for BBU. (b) Selection of the CC mode charging current according to the depth of discharge of the BBU as given by Eq. (1).

power decreases in proportion to the decrease in charging current with the trade-off that the time to fully charge the BBU is increased. Fig. 5 shows the charging time for different DOD of BBU when using a charging current from 1 A to 5 A. We observe that the time to charge a BBU decreases with (1) the decreasing DOD of the BBU and (2) the increasing charging current. We further observe that the charging time remains constant below a certain DOD (for example, below 22% DOD). This is due to the fact that BBUs are charged in mostly CV mode for lower DOD where change in charging time is very small. We also observe that 1 A charging current has a considerably high charging time since we are charging at a very low rate. We do not experiment with charging current less than 1 A as it is the lower end of the recommended constant current charging range for Li-ion batteries [6].

In our experiments, we found that the worst-case charge time for the original 5 A charger is within 45 minutes. We proceed to design the new variable charger with the objective to always charge the battery within the 45 minutes time period. Experiment results in Fig. 5 show that, to keep the charging time within 45 minutes, we can use a lower charging current if the DOD is less than the fully discharged case. For example, if the BBU is 70% discharged, a 4 A charging current could charge the BBU back in 40 minutes, while if the BBU was less than 50% discharged, a 2 A charging current would suffice to charge it back at around the same time. Based upon this observation, we come up with a linear formula to calculate the initial charging current depending upon the depth of discharge of the BBU, such that the charging time is within the 45 minutes bound, as

$$I_C = \begin{cases} 2 + (\text{DOD} - 0.5) \times 6 & \text{if DOD} \geq 50\% \\ 2 & \text{if DOD} < 50\%. \end{cases} \quad (1)$$

where I_C is the charging current (in Ampere) and DOD is the

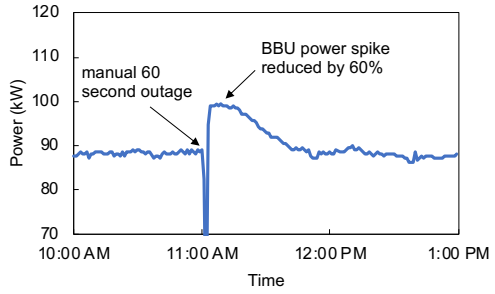

Fig. 7. Power consumption of RPP during the testing of the new variable charger in production.

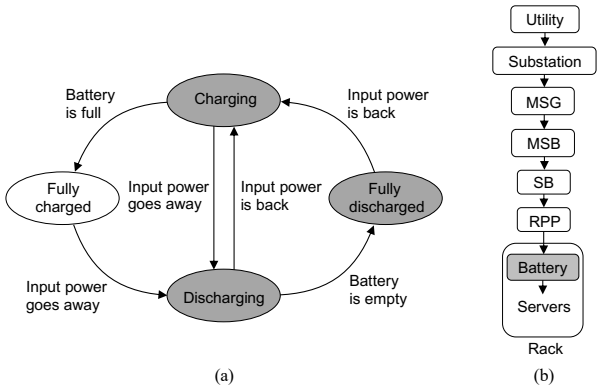

Fig. 8. (a) State transition diagram of batteries in the rack. (b) Major components in the critical power path to a rack.

depth of discharge of the BBU. We limit the charging current to a minimum of 2 A for DOD less than 50% and linearly increase the charging current up to 5 A for DOD greater than 50% as shown in Fig. 6 (b). The overall variable charging logic is shown in the flowchart in Fig. 6 (a) where I_C and V_C are the charging current and voltage respectively.

Manual override: The new variable charger calculates the energy discharged from the BBU during a discharge event and automatically sets the charging current between 2 A to 5 A depending on the DOD of the battery. The recharge power is decreased by as much as 60% (if DOD is less than 50%). In addition to the automatic behavior, we add a *manual override* feature to the battery charger to set the charging current between 1 A and 5 A, whereby a power monitoring and control system (like Dynamo) can set the charging current.

Production validation: To validate the functionality of the variable charger, the RPP circuit breaker powering a test row was manually opened for 60 seconds and closed back, causing an open transition for all the 14 racks in the row. The power consumption of the RPP (sum of input power to the racks) during the test is shown in Fig. 7. Since the BBUs in the racks were discharged by less than 50% (20% DOD on average), they started charging at 2 A. The power use increased by about 10 kW due to the recharging of the BBUs. If it had been the original charger, the power spike would have been more than 26 kW. The new variable charger was able to reduce the battery recharging power by 60% while charging the BBUs back in about 45 minutes. We have already deployed the variable charger to our production data centers and it has helped us in carrying out regular data center maintenance with much greater ease.

IV. Coordinated Battery Charging

The new variable charger is able to lower the battery recharge power for most cases. However, the decision to select the BBU charging current is made locally at the rack level by measuring the DOD of the battery, independent of the aggregated power use at the circuit breakers. We need a higher-level coordination for two major reasons. First, the locally selected charging current can still overload the circuit breakers. We need coordination to charge the rack batteries by taking the power constraint at circuit breakers into consideration, such that we avoid power overloading or capping as much as possible. Second, different racks can have different priorities due to the importance of services running on them. For example, racks running stateful workloads (such as database servers) require much stronger power availability guarantee, preferably having battery backup power source ready all the time. On the contrary, racks running stateless compute workloads (such as web services) may not require such strong guarantee. If we have to reduce the battery charging current to avoid overloading of circuit breakers, we may want to reduce or even postpone charging current of lower priority racks before impacting higher priority racks.

We categorize racks into three priorities, P1 (high), P2 (medium), and P3 (low) based upon the services running on them. Rather than charging all the racks the same way, we define different charging time SLA for the racks based upon their priority, to meet certain reliability goals.

A. Reliability of Racks

A key question we need to answer is: *How does the charging time of batteries impact the reliability of racks?* Services at Facebook are designed around the BBU being able to power the racks for 90 seconds in case of a power loss. For example, if the input power to a rack is lost and not restored within 45 seconds, services prepare for the power outage, such as, by flushing kernel buffers to disk, re-routing web requests, or promoting master database shard away from affected servers. If batteries are in the charging process, meeting the SLA of 90 seconds of power is not guaranteed in case of an open transition or a power outage, which can lead to the services being in an inconsistent state. Hence, for the rack to have battery redundancy, the battery must be fully charged. We quantify this through the *availability of redundancy (AOR)* metric, the fraction of time the rack battery is fully charged.

BBU in the racks could be in one of the four states, fully charged, charging, discharging, or fully discharged, and the transition between these states is as shown in Fig. 8 (a). To measure AOR, we would need to know the time spent in each state, which depends upon (1) the frequency of rack input

power loss, (2) the duration of the power loss, and (3) the time to charge the battery. Major components in the critical power path to the rack are shown in Fig. 8 (b) and a rack will lose power if any of the components fail. Next, we measure the mean time between failure (MTBF)[2] and mean time to repair (MTTR) for the different ways in which the components in the power path may fail by studying the past maintenance and outage data from 2017 to 2019. The failure of rack input power can be categorized into the following four major types.

1) *Utility failure*: Whenever the utility power fails, racks lose power during the open transition from the utility to the diesel generator and again during the back transition (from the diesel generator to the utility), once the utility power is back. We use the MTBF and MTTR of the industrial utility supply from the IEEE standards [16].
2) *Corrective maintenance*: Corrective maintenance work at various levels in the data center power hierarchy is needed to ensure smooth and safe operation. Most of the power devices in Facebook data center have N+1 redundancy and maintenance work requires an open transition from the primary to the reserve power device with another back transition to the primary power device after the maintenance is complete.
3) *Annual maintenance*: In addition to corrective maintenance, periodic preventive maintenance is carried out annually for MSB, SB, and RPP.
4) *Power outages*: While all three failure types described above result in brief rack input power loss, the servers are unaffected as batteries power them during the open transition. However, there are rare power outages for the IT equipment causing service disruption. Power outages usually happen at the MSB, SB, or RPP level (failures above MSB would cause the generators backing the MSB to take over the load).

The MTBF and MTTR of the different components per failure type is summarized in Table I.

Monte Carlo simulation: Considering each component and failure type as an independent block in a series system, we can simulate the state transition diagram in Fig. 8 (a) through Monte Carlo methods to calculate the AOR of rack power as the fraction of time we are in the fully charged state. We use the failure/repair data in Table I to model rack input power failures/repairs. We assume that all failures and repairs are independent and exponentially distributed as done in prior research [11], [41], except for annual maintenance, which we model as normally distributed with $\mu = 1$ year and $\sigma^2 = 41$ days (from maintenance dataset). Note that utility failure and maintenance results in at least two open transitions, one when the utility fails/maintenance starts, and another when the utility is back/maintenance completes, while power outages result in an extended period of rack input power loss until repair is complete. We assume open transitions to be exponentially distributed with a mean of 45 seconds.

[2]All MTBF are calculated by normalizing the number of failures observed by the total number of components during the observation period. For example, if 1 out of 10 MSB fails during an observation period of 1 year, MTBF for MSBs will be 10 years.

TABLE I
COMPONENT FAILURE AND REPAIR TIMES.

Failure type	Component	MTBF (hours)	MTTR (hours)
Utility failure	Utility [16]	6.39×10^3	0.6
Corrective maintenance	Sub/MSG	5.87×10^4	8.0
	MSB	4.12×10^4	20.2
	SB	1.51×10^5	8.7
	RPP	6.31×10^5	5.5
Annual maintenance	MSB	8.76×10^3	12.8
	SB	8.76×10^3	7.4
	RPP	8.76×10^3	9.9
Power outage	MSB	2.93×10^5	6.4
	SB	5.20×10^5	4.6
	RPP	6.25×10^6	10.9

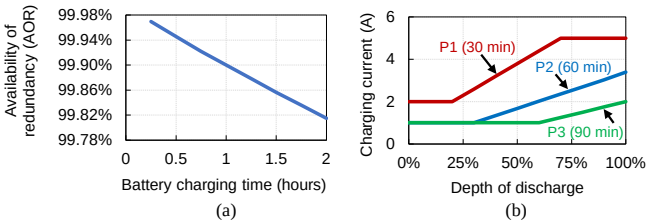

Fig. 9. (a) Availability of redundancy (AOR) of rack power for different battery charging times. (b) Charging current required to satisfy the SLA for the three rack priorities, according to the depth of discharge of the BBU.

We simulate failures and repairs for 10^5 years and repeat the simulation for different battery charge times. Results are shown in Fig. 9 (a) where we can observe that the AOR decreases linearly as battery charging time increases. With this observation, we assign each rack priority a target AOR value and the corresponding battery charging time SLA as shown in Table II. For example, P1 racks are the highest priority and the SLA is to charge the batteries within 30 minutes which will ensure that the racks will meet AOR of 99.94%. Our choice of the AOR value for the racks is limited by the current battery charger design (hardware limitation), which can charge between the range of 1 A and 5 A charging current. In the future, we plan to explore postponing of battery charging, which would allow us to further relax the AOR for lower

TABLE II
CHARGING TIME SLA FOR DIFFERENT RACK PRIORITY.

Rack priority	AOR	Loss of redundancy (hr/year)	Charging time SLA
P1 (high)	99.94%	5.26	30 minutes
P2 (normal)	99.90%	8.76	60 minutes
P3 (low)	99.85%	13.14	90 minutes

priority racks. The general solution framework presented in this paper would apply to future cases, regardless of the AOR values or the number of rack priority levels.

We can calculate the charging current required to meet the charging time SLA for the three rack priorities, according to the DOD of BBU, by linearly interpolating the BBU charging time data in Fig. 5, as shown in Fig. 9 (b).

B. Coordinated Control Architecture

In an effort to coordinate the battery charging process, we add functionality to the existing Dynamo architecture. Dynamo primarily consists of a light-weight *agent*, which runs on every server, and a set of distributed *controllers* mirroring the power hierarchy, which monitors the power form every server as well as circuit breakers in the power hierarchy. Dynamo agents can read server power as well as perform power capping/uncapping upon request. The lowest level of controllers, the *leaf-controller*, is responsible for protecting the RPP (lowest level circuit breaker). There is a leaf-controller for every RPP that directly communicates with agents under that RPP to continuously monitor the aggregate server power (as well as read directly from the RPP power meter) and compare against the power limit of RPP to detect overloading. *Upper-level controllers* (that protect MSB and SB) communicate with leaf-controllers to aggregate power at the corresponding circuit breaker. Whenever a controller detects a power overload, it issues server power capping requests to agents under the circuit break to prevent it from tripping. Following functionality and control logic were added to the Dynamo architecture, such that, in the case of power overloading caused due to battery charging, we can reduce the battery recharge power as a first line of defense before taking the more aggressive step of capping servers that impact service performance.

Dynamo agent: We built a new type of Dynamo agent that runs on the top-of-rack (TOR) switch in each rack. It can communicate with the power supply units (PSU) in the rack to read the input and output power (IT load) of the rack as well as charging/discharging power of BBU. The agent can also issue a *manual override* command to the PSU to change the charging current between 1 A and 5 A, as allowed by the new variable charger. The agent by itself does not perform any action but acts as a request handler waiting for the controller to issue the read/write commands.

Dynamo controller: The leaf-controller reads in additional information, such as, rack power and BBU recharging power, from the agents running on the TOR switch. The controller can detect loss of input power during open transition and the power BBUs recharge at. The DOD of the battery is estimated from the length of the open transition and IT load of the rack during the power loss. Additionally, all controllers also keep track of the priority of racks under the circuit breaker.

C. Priority-aware Charging Algorithm

Our goal is to meet the charging time SLA for all the racks, during a battery charging event. Fig. 9 (b) shows the current a rack needs to be charged at, to meet the charging time SLA, depending upon the DOD of the battery as well as the rack priority. However, more importantly, we need to also make sure that we do not overload the circuit breaker when charging the rack batteries. Hence, we design a new priority-aware control policy to satisfy the charging time SLA for the racks as long as there is available power to meet the demand (we refer to the difference between the power limit and the power use of a circuit breaker as *available power*).

During a battery discharge event, due to rack input power loss, the leaf-controller monitors the IT load of the racks and calculates the energy discharged (DOD of batteries) from each rack under it. At the beginning of the charging sequence, the controller calculates the SLA charging current for all racks based upon the DOD of the battery and the rack priority as shown in Fig. 9 (b). Starting from the rack with the *highest priority* and *lowest DOD*, we satisfy the SLA charging current for the rack as long as there is available power. This order ensures that SLA for higher priority racks are met first while maximizing the number of racks that meet the SLA within the same priority group (since a rack with the lowest DOD will require the lowest charging current to meet the SLA). Our *highest-priority-lowest-discharge-first* battery charging algorithm is summarized in Algorithm 1. Finally, the charging current overrides are sent to the corresponding racks.

The initial charging current calculation (using Algorithm 1) and setting of charging current would be done by the leaf-controller. However, the leaf-controllers as well as the upper-level controllers would all monitor the power use of their corresponding circuit breaker for the entire charging period. If a power overload is detected, the controller starts setting the charging current to the minimum of 1 A for racks in the reverse (*lowest-priority-highest-discharge-first*) order. In the extreme event that the overloading of circuit breakers cannot be avoided even after charging all the racks at the minimum charging current, as a last resort, we start capping servers to reduce the IT load and prevent circuit breakers from tripping.

Algorithm 1 Highest-priority-lowest-discharge-first battery charging algorithm

Input: Available power, Battery DOD/priority of all racks
Output: Charging current for all racks

1: **for all** racks **do**
2: Initialize the charging current to the minimum of 1 A
3: Calculate the SLA charging current from DOD and rack priority as shown in Fig. 9 (b)
4: **end for**
5: Sort the racks according to priority and then by DOD
6: **while** available power > 0 **do**
7: Satisfy the SLA for the rack with the next highest priority and lowest energy discharge by setting its charging current as calculated in step 3
8: **end while**

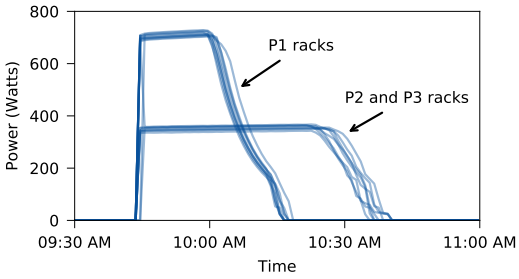

Fig. 10. Battery recharge power of 17 racks in a test row after an open transition.

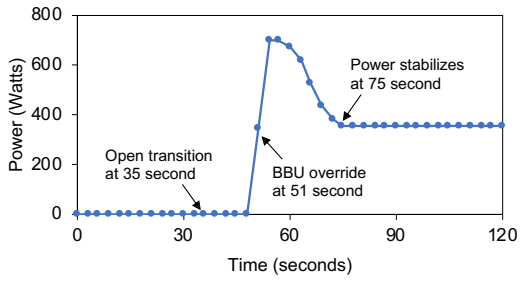

Fig. 11. Battery recharge power of a rack undergoing override of the BBU charging current from the leaf-controller.

V. Evaluation

We have built a working prototype system in a production data center (which will be deployed across all data centers after further testing and validation). We present results from this real prototype system. Furthermore, we evaluate our priority-aware battery charging algorithm through simulation using real rack power traces from a production data center.

A. Prototype Experiment

The Dynamo agent that runs on the TOR switch of every rack can read different types of power readings from the PSUs in the rack. In racks with the new variable battery charger, the agent can also override the charging current for the BBUs. In this experiment, we demonstrate the working of a prototype Dynamo controller that we have built. We have an experimental Dynamo controller setup in a data center suite whose SBs are going to be transferred from the reserve MSB to the normal MSB after the completion of a planned maintenance.

Fig. 10 shows the battery recharging power of racks in a row being monitored by a leaf-controller. The leaf-controller is monitoring and protecting an RPP that powers the row. This particular row has 9 P1 racks, 5 P2 racks, and 3 P3 racks, for a total of 17 racks. The open transition occurred at 09:43 AM for about 5 seconds and the DOD of BBU in the racks was less than 5%. All 17 racks started charging at the 2 A charging current, the default current selected by the new variable charger. However, almost immediately, the leaf-controller calculates the SLA charging current for all the racks according to their priority, which in this case is 2 A for P1 racks and 1 A for P2 and P3 racks (from Fig. 9 (b)). Since, the power consumption at RPP is not constrained, the leaf-controller overrides all racks with the calculated SLA charging current. As shown in Fig. 10, P1 racks are charging at 2 A (about 700 W recharge power) while P2 and P3 racks are charging at 1 A (about 350 W recharge power). Also, the P1 racks complete the charging process in about 30 minutes while the P2 and P3 racks are fully charged within an hour.

A more fine-grain power reading of BBU recharge power from one of the racks that gets overridden with 1 A charging current is shown in Fig. 11. The open transition starts at 35 seconds, which is detected by the leaf-controller from the rack input power going to zero (not shown in the figure). Upon detecting the first BBU recharge power, the leaf-controller performs the SLA charging current calculation and overrides the BBU charging current. We can see that the BBU power stabilizes to the override value after about 20 seconds of the command being issued.

B. Simulation Experiment

The prototype experiment we carried out above during a planned maintenance event, represents a normal operation scenario. However, many corner case scenarios (for example, scenarios leading to deeper battery discharge, or extreme power constraints) rarely happen in production data center environments and are difficult and/or risky to recreate in a production environment. Thus, through simulation, we perform a more comprehensive evaluation and also compare the results of the original 5 A charger, the new variable charger, and the coordinated priority-aware charging.

1) Experimental setup: We simulate a Dynamo controller and open transitions at the MSB level. We perform sensitivity analysis of the charging algorithm under different power constraints (available power), DOD of batteries, and rack priority distribution.

Production rack power trace: We collect rack power trace at 3 second granularity for racks under an MSB and replay the power trace in our simulation. This particular MSB has 89 P1 racks, 142 P2 racks, and 85 P3 racks, for a total of 316 racks. The actual power limit for the MSB is 2.5 MW, however, we experiment with different power limits in our simulation to study the effect of varying available power under different possible utilization and oversubscription scenarios. We vary power limit only for the MSB and assume that all lower-level circuit breakers have enough available power to charge the batteries (for latest Facebook data centers power is normally constrained at the MSB level because of generator capacity). The aggregate power consumption of the MSB for a week is shown in Fig. 12, where we can observe that the power use exhibits diurnal cycles between the range of 1.9 MW to 2.1 MW. We simulate open transitions at the first peak in the trace as this is when the available power for battery recharging is most constrained.

BBU charging power profile: Batteries charge in the CC-CV two-step process. We consider the CC phase of the BBU as having a constant power draw (proportional to the charging

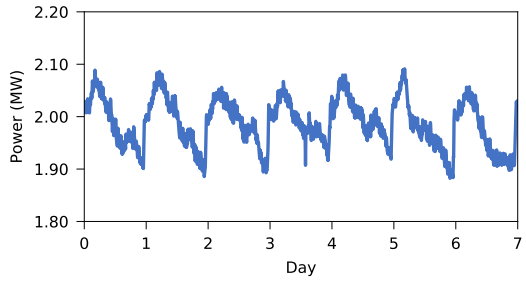

Fig. 12. Aggregate power of MSB used for evaluation.

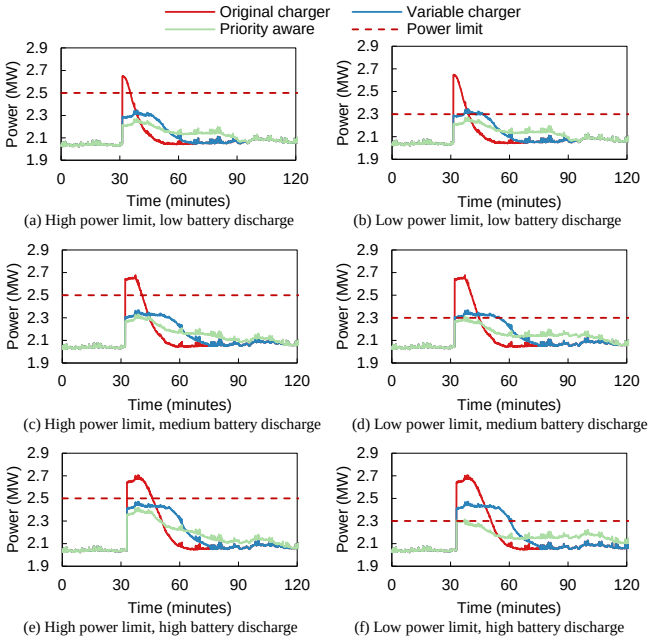

Fig. 13. MSB power use in the case of the original charger, new variable charger, and priority-aware charging for varying power limit and battery discharge.

current), while we found that the CV phase can be approximated by fitting an exponential function of the form Ae^{Bt}. For example, for a fully discharged rack batteries charging at 5 A charging current, CC power would be a constant 1.9 kW and the CV power would be approximated with the $1.9e^{-0.18t}$ kW function. We use the battery charging times for different DOD and charging current from our lab experiment in Section III-B (Fig. 5). The DOD of batteries depend upon the IT load of the rack and the length of open transition. In our simulation, we vary the DOD of the batteries by varying the length of open transition. We experiment with three levels of battery discharge: (1) low discharge, (2) medium discharge, and (3) high discharge, where the average DOD of the BBU is 30%, 50%, and 70%, respectively.

2) Coordinated battery charging results: The primary purpose of the coordinated priority-aware battery charging algorithm is to protect the circuit breakers from overloading due to the battery recharge power spike. The original 5 A charger and the new variable charger work locally at the rack level without any coordination. We compare the coordinated priority-aware charging with the original 5 A charger and the new variable charger to demonstrate why coordination is necessary. We experiment with a high power limit of 2.5 MW (the actual power limit) and a low power limit of 2.3 MW (a probable scenario of low available power) at three levels of battery discharge. Results for the six cases are shown in Fig. 13. We can see that the original charger would cause power overloading of the MSB for all of the cases since the initial power spike is very high. The new variable charger is better in the sense that the initial power spike is reduced by 60% for most of the cases (if BBUs are less than 50% discharged). However, for the low power limit cases, even the new variable charger would cause overloading of the circuit breaker. Server power capping is required in such cases to reduce the IT load. Table III shows the maximum server power capping required (magnitude and percentage of IT load) for the six cases, under the different deployment scenarios. Both the original charger as well as the variable charger would lead to server power capping (as high as 20%), resulting in performance degradation. The problem gets worse when battery discharge is higher and/or power limit is lower.

TABLE III
MAXIMUM SERVER POWER CAPPING REQUIRED FOR THE SIX CASES IN FIG. 13 (A)–(F).

Case	Original charger	Variable charger	Priority-aware
(a)	149 kW (7%)	0 kW (0%)	0 kW (0%)
(b)	349 kW (17%)	45 kW (2%)	0 kW (0%)
(c)	178 kW (9%)	0 kW (0%)	0 kW (0%)
(d)	378 kW (18%)	68 kW (3%)	0 kW (0%)
(e)	205 kW (10%)	0 kW (0%)	0 kW (0%)
(f)	405 kW (20%)	171 kW (8%)	0 kW (0%)

On the other hand, the coordinated priority-aware charging algorithm would avoid power loading (no server power capping) in all the six cases. This is due to the fact that we constantly monitor the available power of the circuit breaker and lower the battery charging rate if power overloading is detected. Only during the extreme case that power overloading would occur even after all the batteries are charging at their minimum rate, would we resort to server power capping. In this particular experiment, server power capping would begin if the available power was less than 120 kW (power limit was below 2.2 MW). Thus, coordinated charging can prevent circuit breakers from overloading while minimizing the performance degradation from power capping. The trade-off is that our solution would slow down the battery charging process and compromise the redundancy. However, we prefer to relax the redundancy provided by the batteries to minimize performance degradation.

3) Priority-aware battery charging results: Our battery charging algorithm not only protects the circuit breakers from overloading, but also does it in a priority-aware way. Whenever power is constrained, the battery charging rate of lower priority

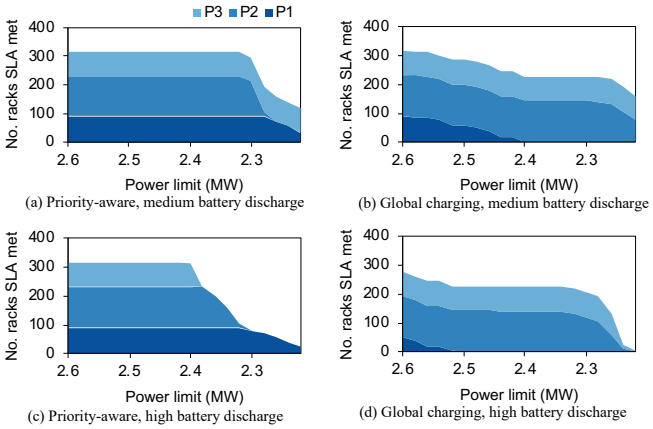

Fig. 14. Performance of priority-aware charging algorithm and global charging algorithm in terms of number of racks whose charging time SLA are met for different battery discharge.

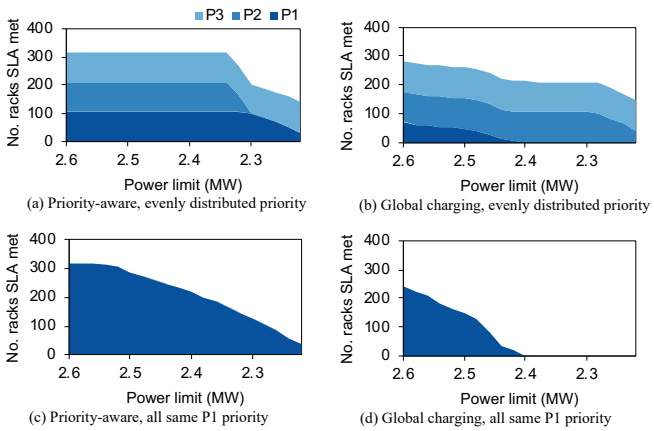

Fig. 15. Performance of priority-aware charging algorithm and global charging algorithm in terms of number of racks whose charging time SLA are met for different rack priority distribution.

racks are reduced first before impacting the higher priority racks. We compare the priority-aware charging algorithm with a baseline global charging algorithm. The global charging algorithm only looks at the available power during a charging event and charges all the racks at the same rate to prevent power overload. While the global charging algorithm also coordinates charging of racks to prevent circuit breakers from overloading, it does not consider the priority (or battery DOD) of racks. We compare the priority-aware charging algorithm with the global charging algorithm in terms of the number of racks that can meet the charging time SLA for varying power limits.

Different battery discharge: Fig. 14 shows the number of racks (disaggregated by priority) that meet the SLA for medium and high battery discharge cases when the power limit is gradually decreased from 2.6 MW to 2.2 MW. We can observe that our priority-aware charging algorithm satisfies the charging time SLA for P1 racks as long as possible when the power limit is decreased. P3 racks are the ones affected first before reducing the charging current for P2 and P1 racks as seen in Fig. 14 (c). Note that in Fig. 14 (a), P2 racks seems to be affected before P3 racks because, even though the P3 racks are charging at the minimum rate (due to the current hardware limitation), their SLAs are met. On the contrary, P1 racks are the first ones to get penalized by the global charging algorithm, followed by P2 racks. This is because higher priority racks have higher charging current demand (to meet stricter SLA), but all racks are charged with the same charging current regardless of their priority.

Different rack priority distribution: We repeat the same experiment by varying the rack priority distribution. Once with evenly distributed priority (each third of the racks has P1, P2, and P3 priority) and another with all racks having the same P1 priority. Fig. 15 shows the result for the case of medium battery discharge. The result for the evenly distributed rack priority is similar to the previous experiment (again, P3 rack SLA is satisfied even when charging at the minimum rate). In the case of all racks having the same P1 priority, our priority-aware charging algorithm performs superior to the global charging algorithm. For example, the average number of racks that meet the SLA for priority-aware charging in Fig. 15 (c) is 208, about three times higher than the baseline in Fig. 15 (d). This is due to the fact that the lowest-discharge-first order selects racks with the lowest DOD (which require the lowest charging current to meet the SLA) to be satisfied first. This maximizes the number of racks that meet the SLA for the given available power.

VI. Related Work

Various works in the literature have looked at controlling the charging or discharging of batteries in data centers. Govindan et al. [10] were among the first to propose using UPS batteries in data centers for peak power shaving. The basic idea is to charge the battery using utility power when the power demand is low and use it to supplement the utility power when the power demand is high. The benefit of peak power shaving is twofold. (1) Saving on capital expenses (Cap-Ex) [1], [2], [12], [13], [18]: existing power infrastructure can be oversubscribed (under-provisioned) by installing more IT equipment to save on Cap-Ex. (2) Saving on operational expenses (Op-Ex) [5], [26], [27], [40]: we can save on Op-Ex by hiding peak demands to the utility, since the electricity cost includes a significant peak power use component in addition to the energy usage component. These works are orthogonal to server throttling or workload scheduling techniques for peak power shaving which impact performance.

Another line of work looks at minimizing the total data center electric bill by utilizing the batteries [14], [15], [43], [48], [49]. The basic idea is to use battery power when/where the electricity price is high while charging the battery when/where the electricity price is low. Electricity price differences may be in time (utility charging a varying real-time price) or location (many utilities charging varying prices to geographically distributed data centers). A related area of research explores using

batteries to participate in demand response programs [22]–[25], [28], whereby a utility is actively offering incentives to reshape consumer's demand. These works study the feasibility of demand response participation and its impact on availability of the batteries. There are also works that focus on integrating on site production of intermittent renewable energy [4], [7], [35], [38], [42], such as, solar and wind using batteries.

All of the above works have mainly focused on repurposing the batteries from its original intended use of providing backup power during an open transition. Our work differs from them since we focus on the primary use of the battery to serve as a fail-over mechanism. Additional work exists that look into a multitude of factors, such as, UPS battery placements [11], appropriate sizing of batteries [44], alternate energy storage technologies [45], using batteries in colocation data centers [31], and battery aging issue [20]. However, all prior works have ignored the problem caused by simultaneous recharging of batteries, which is a common phenomenon. There are control solutions proposed for the conventional centralized UPS systems [32], [46], but they do not directly apply to the case of distributed power supply/batteries since the power constraint is at an upstream power device.

To the best of our knowledge, our work is the first to identify and highlight the importance of the problem caused by distributed battery recharging. Furthermore, we describe our solution in detail, the mechanism as well as the policy, which we have successfully tested in our data centers.

VII. Conclusion

In this paper, we have identified the problem caused by recharging of distributed batteries in oversubscribed data centers after an open transition, with case studies of past events that have occurred in Facebook data centers. Rather than statically allocate the data center power budget for battery recharge power, we proposed a novel variable battery charger which allows us to efficiently utilize and allocate the power budget for IT equipment. We have designed, developed, and implemented the variable battery charger in production to reduce the impact of power spike from battery recharging.

Lastly, we developed and evaluated a priority-aware charging algorithm which utilizes the manual override feature of the new battery charger to coordinate the battery charging process. Our coordinated priority-aware battery charging algorithm was shown to minimize performance degradation (obviates the need for up to 20% server power capping) as well as maximizes the number of highest priority racks that satisfy the charging time SLA without exceeding the power limit of circuit breakers.

Acknowledgment

We would like to thank our colleagues at Facebook, Bahram Davalian, Greg Epstein, Erik Meijer, Scott Michelson, Aaron Miller, Ramesh Swaminathan, Brian Spaulding, Neha Tibrewal, Kaushik Veeraraghavan, Carole-Jean Wu, and Yang Xia for their contribution, feedback, and suggestions on this paper. We would also like to thank the anonymous reviewers for their valuable feedback in improving this paper.

References

[1] B. Aksanli, E. Pettis, and T. Rosing, "Architecting efficient peak power shaving using batteries in data centers," in *Proceedings of the IEEE 21st International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS'13)*, 2013, pp. 242–253.

[2] B. Aksanli, T. Rosing, and E. Pettis, "Distributed battery control for peak power shaving in datacenters," in *Proceedings of the International Green Computing Conference (IGCC'13)*, 2013, pp. 1–8.

[3] L. A. Barroso, U. Hölzle, and P. Ranganathan, "The datacenter as a computer: Designing warehouse-scale machines," *Synthesis Lectures on Computer Architecture*, vol. 13, no. 3, pp. i–189, 2018.

[4] T. Chen, Y. Zhang, X. Wang, and G. B. Giannakis, "Robust workload and energy management for sustainable data centers," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 651–664, 2016.

[5] M. Dabbagh, A. Rayes, B. Hamdaoui, and M. Guizani, "Peak shaving through optimal energy storage control for data centers," in *Proceedings of the IEEE International Conference on Communications (ICC'16)*, 2016, pp. 1–6.

[6] S. Dearborn, "Charging Li-ion batteries for maximum run times," *Power Electronics Technology*, vol. 31, no. 4, pp. 40–49, 2005.

[7] W. Deng, F. Liu, H. Jin, C. Wu, and X. Liu, "Multigreen: Cost-minimizing multi-source datacenter power supply with online control," in *Proceedings of the 4th International Conference on Future Energy Systems (e-Energy'13)*, 2013, pp. 149–160.

[8] X. Fan, W.-D. Weber, and L. A. Barroso, "Power provisioning for a warehouse-sized computer," in *Proceedings of the 34th Annual International Symposium on Computer Architecture (ISCA'07)*, 2007, pp. 13–23.

[9] X. Fu, X. Wang, and C. Lefurgy, "How much power oversubscription is safe and allowed in data centers," in *Proceedings of the 8th ACM International Conference on Autonomic Computing (ICAC'11)*, 2011, pp. 21–30.

[10] S. Govindan, A. Sivasubramaniam, and B. Urgaonkar, "Benefits and limitations of tapping into stored energy for datacenters," in *Proceedings of the 38th Annual International Symposium on Computer Architecture (ISCA'11)*, 2011, pp. 341–351.

[11] S. Govindan, D. Wang, L. Chen, A. Sivasubramaniam, and B. Urgaonkar, "Towards realizing a low cost and highly available datacenter power infrastructure," in *Proceedings of the 4th Workshop on Power-Aware Computing and Systems (HotPower'11)*, 2011, pp. 7:1–7:5.

[12] S. Govindan, D. Wang, A. Sivasubramaniam, and B. Urgaonkar, "Leveraging stored energy for handling power emergencies in aggressively provisioned datacenters," in *Proceedings of the 17th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'12)*, 2012, pp. 75–86.

[13] S. Govindan, D. Wang, A. Sivasubramaniam, and B. Urgaonkar, "Aggressive datacenter power provisioning with batteries," *ACM Transactions on Computer Systems*, vol. 31, no. 1, pp. 2:1–2:31, 2013.

[14] Y. Guo and Y. Fang, "Electricity cost saving strategy in data centers by using energy storage," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 6, pp. 1149–1160, 2013.

[15] Y. Guo, Z. Ding, Y. Fang, and D. Wu, "Cutting down electricity cost in internet data centers by using energy storage," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM'11)*, 2011, pp. 1–5.

[16] IEEE Std 3006.8, "IEEE recommended practice for analyzing reliability data for equipment used in industrial and commercial power systems," 2018.

[17] M. A. Islam, X. Ren, S. Ren, A. Wierman, and X. Wang, "A market approach for handling power emergencies in multi-tenant data center," in *Proceedings of the IEEE International Symposium on High Performance Computer Architecture (HPCA'16)*, 2016, pp. 432–443.

[18] V. Kontorinis, L. E. Zhang, B. Aksanli, J. Sampson, H. Homayoun, E. Pettis, D. M. Tullsen, and T. Simunic Rosing, "Managing distributed UPS energy for effective power capping in data centers," in *Proceedings of the 39th Annual International Symposium on Computer Architecture (ISCA'12)*, 2012, pp. 488–499.

[19] C. Lefurgy, X. Wang, and M. Ware, "Server-level power control," in *Proceedings of the 4th International Conference on Autonomic Computing (ICAC'07)*, 2007, pp. 4–4.

[20] L. Liu, H. Sun, C. Li, T. Li, J. Xin, and N. Zheng, "Managing battery aging for high energy availability in green datacenters," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 12, pp. 3521–3536, 2017.

[21] S. Malla and K. Christensen, "A survey on power management techniques for oversubscription of multi-tenant data centers," *ACM Computing Surveys*, vol. 52, no. 1, p. 1, 2019.

[22] A. Mamun, I. Narayanan, D. Wang, A. Sivasubramaniam, and H. K. Fathy, "A stochastic optimal control approach for exploring tradeoffs between cost savings and battery aging in datacenter demand response," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 1, pp. 360–367, 2018.

[23] A. Mamun, I. Narayanan, D. Wang, A. Sivasubramaniam, and H. Fathy, "Multi-objective optimization of demand response in a datacenter with lithium-ion battery storage," *Journal of Energy Storage*, vol. 7, pp. 258–269, 2016.

[24] A. Mamun, D. Wang, I. Narayanan, A. Sivasubramaniam, and H. Fathy, "Physics-based simulation of the impact of demand response on lead-acid emergency power availability in a datacenter," *Journal of Power Sources*, vol. 275, pp. 516–524, 2015.

[25] L. Narayanan, D. Wang, A.-A. Mamun, A. Sivasubramaniam, and H. K. Fathy, "Should we dual-purpose energy storage in datacenters for power backup and demand response?" in *Proceedings of the 6th Workshop on Power-Aware Computing and Systems (HotPower'14)*, 2014.

[26] N. Nasiriani, G. Kesidis, and D. Wang, "Optimal peak shaving using batteries at datacenters: Characterizing the risks and benefits," in *Proceedings of the IEEE 25th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS'17)*, 2017, pp. 164–174.

[27] N. Nasiriani and G. Kesidis, "Optimal peak shaving using batteries at datacenters: Charging risk and degradation model," in *Proceedings of the International Conference on Computing, Networking and Communications (ICNC'18)*, 2018, pp. 58–62.

[28] A. Oleksiak, W. Piatek, K. Kuczynski, and F. Sidorski, "Reducing energy costs in data centres using renewable energy sources and energy storage," in *Proceedings of the 5th International Workshop on Energy Efficient Data Centers (E2DC'16)*, 2016, pp. 5:1–5:8.

[29] Open Compute Project, https://www.opencompute.org/, 2019, [Online; accessed November 8, 2019].

[30] Open Compute Project, https://www.opencompute.org/wiki/Open_Rack/SpecsAndDesigns, 2019, [Online; accessed November 8, 2019].

[31] D. S. Palasamudram, R. K. Sitaraman, B. Urgaonkar, and R. Urgaonkar, "Using batteries to reduce the power costs of internet-scale distributed networks," in *Proceedings of the 3rd ACM Symposium on Cloud Computing (SoCC'12)*, 2012, pp. 11:1–11:14.

[32] X. Pei and Y. Kang, "Short-circuit fault protection strategy for high-power three-phase three-wire inverter," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 3, pp. 545–553, 2012.

[33] P. Ranganathan, P. Leech, D. Irwin, and J. Chase, "Ensemble-level power management for dense blade servers," in *Proceedings of the 33rd Annual International Symposium on Computer Architecture (ISCA'06)*, 2006, pp. 66–77.

[34] T. B. Reddy, *Linden's handbook of batteries*. McGraw-hill New York, 2011, vol. 4.

[35] C. Ren, D. Wang, B. Urgaonkar, and A. Sivasubramaniam, "Carbon-aware energy capacity planning for datacenters," in *Proceedings of the IEEE 20th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS'12)*, 2012, pp. 391–400.

[36] V. Sakalkar, V. Kontorinis, D. Landhuis, S. Li, D. D. Ronde, T. Blooming, J. Kennedy, C. Malone, and P. Ranganathan, "Data center power oversubscription with a medium voltage power plane and priority-aware capping," in *Proceedings of the 25th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'20)*, 2020.

[37] P. Sarti and Z. Ebrahimzadeh, "BBU battery backup module 3600W for V2 power shelf," https://www.opencompute.org/wiki/Open_Rack/SpecsAndDesigns, Open Compute Project, Tech. Rep., 2015.

[38] N. Sharma, S. Barker, D. Irwin, and P. Shenoy, "Blink: Managing server clusters on intermittent power," in *Proceedings of the 16th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'11)*, 2011, pp. 185–198.

[39] W. Shen, T. T. Vo, and A. Kapoor, "Charging algorithms of lithium-ion batteries: An overview," in *Proceedings of the 7th IEEE Conference on Industrial Electronics and Applications (ICIEA'12)*, 2012, pp. 1567–1572.

[40] Y. Shi, B. Xu, B. Zhang, and D. Wang, "Leveraging energy storage to optimize data center electricity cost in emerging power markets," in *Proceedings of the 7th International Conference on Future Energy Systems (e-Energy'16)*, 2016, pp. 18:1–18:13.

[41] B. R. Shrestha, T. M. Hansen, and R. Tonkoski, "Reliability analysis of 380V DC distribution in data centers," in *Proceedings of the IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT'16)*, 2016, pp. 1–5.

[42] C. Stewart and K. Shen, "Some joules are more precious than others: Managing renewable energy in the datacenter," in *Proceedings of the Workshop on Power-Aware Computing and Systems (HotPower'09)*, 2009, pp. 15–19.

[43] R. Urgaonkar, B. Urgaonkar, M. J. Neely, and A. Sivasubramaniam, "Optimal power cost management using stored energy in data centers," in *Proceedings of the ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS'11)*, 2011, pp. 221–232.

[44] D. Wang, S. Govindan, A. Sivasubramaniam, A. Kansal, J. Liu, and B. Khessib, "Underprovisioning backup power infrastructure for datacenters," in *Proceedings of the 19th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'14)*, 2014, p. 177–192.

[45] D. Wang, C. Ren, A. Sivasubramaniam, B. Urgaonkar, and H. Fathy, "Energy storage in datacenters: What, where, and how much?" in *Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS'12)*, 2012, p. 187–198.

[46] B. Wei, A. Marzàbal, J. Perez, R. Pinyol, J. M. Guerrero, and J. C. Vásquez, "Overload and short-circuit protection strategy for voltage source inverter-based UPS," *IEEE Transactions on Power Electronics*, vol. 34, no. 11, pp. 11 371–11 382, 2019.

[47] Q. Wu, Q. Deng, L. Ganesh, C. Hsu, Y. Jin, S. Kumar, B. Li, J. Meza, and Y. J. Song, "Dynamo: Facebook's data center-wide power management system," in *Proceedings of the ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA'16)*, 2016, pp. 469–480.

[48] J. Yao, X. Liu, and C. Zhang, "Predictive electricity cost minimization through energy buffering in data centers," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 230–238, 2014.

[49] H. Zhou, J. Yao, H. Guan, and X. Liu, "Comprehensive understanding of operation cost reduction using energy storage for IDCs," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM'15)*, 2015, pp. 2623–2631.

[50] J. Zipfel, "Google joins Open Compute Project to drive standards in IT infrastructure," https://cloud.google.com/blog/products/gcp/google-joins-open-compute-project-to-drive-standards-in-it-infrastructure, 2016, [Online; accessed March 15, 2020].