Emergency Auxiliary Services: A Bi-Directional Mutual Beneficial Framework for Power Systems and Data Centers

M. Jawad[1], S. M. Ali[2], Z. Ullah[3*], M. U. S. Khan[4], C. A. Mehmood[2], B. Khan[2], A. Fayyaz[2]

1. *Electrical Engineering, COMSATS University Islamabad, Lahore Campus, Pakistan.*
2. *Electrical Engineering, COMSATS University Islamabad, Abbottabad Campus, Pakistan*.
3. *Electrical Engineering Department, UMT Lahore, Sialkot Campus, Pakistan.*
4. *Department of Computer Science, COMSATS University Islamabad, Abbottabad Campus, Pakistan*

∗ **Corresponding Author:** Email: zahid.ullah@skt.umt.edu.pk

Abstract

The power systems are facing a gradual increase in electrical load at different hours of the day and on different buses of the network. This situation can create the problem of demand/supply mismanagement in the power system. Therefore, the power systems need fast auxiliary services to keep power management, stability, and reliability in the network. Conventionally, power systems have own dedicated computing facility for executing auxiliary services, however, data centers are among the largest energy consumption clients for the power systems and have the capability to provide enough computational resources to the power system when required. This paper proposes an Emergency Auxiliary Services (EAS) model for power systems and data centers to work combinedly with mutual benefits. A dynamic Service Level Agreement (SLA) is introduced along with an EAS job scheduling algorithm that motivates data center to run power system jobs on priority and effectively during emergency conditions and maintain data center revenue. The EAS includes Optimal Power Flow (OPF) analysis, bus centrality index, and transmission line centrality index. The simulations are performed on real workload of a data center integrated with the IEEE 30-bus system to assess the performance of the proposed model. The results illustrate that the priority execution EAS on data centers has a minimal impact on overall energy consumption and on other cloud computing jobs' time of execution. Moreover, the dynamic SLA compensates the data center revenue loss due to prior execution of the EAS. Therefore, the SLA encourages the data center operators to execute EAS on priority.

Key Words: Data center, Load flow, Optimization, Scheduling, Smart grid

1. Introduction

Today, one of the biggest problems in the world is energy crises. The problem becomes worst due to the rapid growth in the size of data centers. The data centers of Google consume more than 250 MWh of energy in a month that is more than the per month energy consumption of whole Salt Lake City [1]. The high energy demand of Google's data center is due to the growing usage of the internet and cloud computing services. The data center has hundreds to thousands of servers carrying out run time analysis of the requested data. The intensified data analysis requires increased energy consumption that results in data centers' huge electricity bills, which data center must pay to the power supplying agencies [2].

Due to the increased Transmission Line Failures (TLFs), the voltage at the consumer's side will reduce, resulting in a low voltage profile throughout the power system. The data centers can fulfil the intensive computational requirements of modern power systems. The computational services provided by the data center to the power system for stable and reliable operation under emergency is known as Emergency Auxiliary Services (EAS). Conventionally, a bunch of servers or clusters are dedicated to the power system for the EAS. However, servers' dedication has some major drawback, such as (a) energy loss due to the idle resources, (b) high computing cost due to dedicated servers, and (c) underutilization of computing resources. Moreover, the EAS can arrive at any time instant and the EAS job length is variable; for example, if the EAS is coming after every 1 hour for 10 minutes and 100 CPUs are dedicated, then a total of 100×100W=10kW of power or 36 megajoules of energy will be consumed by the idle CPUs in every hour.

In the past decade, most of the researchers have investigated the methods toward the reduction in power consumption of data centers along with the energy cost optimization. For instance, Pedram in [8] proposed a model to

minimize cost based on electricity price and cloud workload. In [9], the authors studied the energy cost reduction problem of the data center under multi-electricity market environment and renewable energy. The authors in [10] and [11] focused on renewable energy to tackle the energy cost problem. Likewise, the aforesaid problem is handled through the Service Level Agreement (SLA) and workload distribution in Ref. [12]. Moreover, the workload distribution criterion depends on the pricing variation. Furthermore, energy saving relies on the selective and flexible conditions of SLA. In [12], the model was dependent on load variations due to climate conditions, which is not an optimum approach because the climatic condition is not the only monitoring constraint. Wang et. al. [13] approached the problem of cost minimization via deregulated energy cost for data centers. The EAS between the power system and data center was not incorporated in aforesaid research work. A novel EAS scheme is provided for optimizing power flows, increasing the reliability of the power system, minimizing Transmission Line Losses (TLLs), maximizing data center revenue under various pricing, and incorporated a real-time bidirectional SLA for mutual benefits. The researchers also emphases on power consumption estimation of the servers for performing various processing tasks. The analysis of power and energy reduction in multi-core processes using Dynamic Frequency Scaling (DFS) is presented in [14]. In [14], Congfeng Jiang discussed that the utilization factor of the central processing unit is not always an indicator for the calculation of power consumption. The total power consumption on storing various applications is elaborated in [15].

Although the aforementioned models provide optimized solutions for reducing power consumption, they are unable to incorporate the effect of energy reductions in the data center revenues. Moreover, the cost reduction models are not analysed in the above-mentioned energy reduction schemes. Furthermore, the models were unable to consider the effects of power delivered on the data center economics. To the best of our knowledge, the literature studies and reports lack the mechanism that used the data center for the fast execution of the EAS for the power system, such as OPF and identification of endangered buses and transmission lines. Moreover, no SLA between the power system and data center exist to tackle the emergency situation of the power system.

In [16], the authors proposed the Auxiliary Services Model (ASM) for the mutual benefits of the power system and data center for the first time. The idea was to use the computational capability of the date centers for the fast execution of power system jobs instead of using dedicated servers rooms. In the previous paper, the problem was addressed and solved in general terms, such as: (a) how the power system jobs will be scheduled and which job scheduling algorithm will be appropriate, (b) what is the impact of power system jobs on the data center operation, such as power consumption, makespan, number of preempted jobs, queue time, and resource utilization, and (c) what is the impact on the revenue of the data center and how an SLA can be defined between the data center and the power system to maximize data center's revenue and ensure the power system reliability.

The model presented in this paper is the continuation of our previous work to explore the problem further for the emergency scenarios only. The emergency means that the job of the power system cannot be delayed or pre-empted and must be execute on highest priority. This case involves the preemption of other data center jobs as the percentage workload in the data center is often above 100%, which means jobs are in queue to be executed. If certain jobs of the data center will be preempted, then the job delaying penalty will be different due to different types of jobs and SLAs. Therefore, in this paper, we evaluated and compares four known job delaying penalty calculation schemes. Moreover, by observing the historical workload pattern\ distribution of the data center, we further divided the whole days into three-time intervals of the workload, such as low-load, medium-load, and peak-load. Therefore, the data center's job preemption scenarios and penalty cost computations will be different in the aforementioned three-time intervals. Furthermore, the impact of EAS is analysed and compared for different percentage of CPU utilization. Therefore, the scope of the current work has a significant difference from our previous work. The main contributions of our paper are stated as:

- An EAS model is proposed for the mutual benefits of the power system and data center. Under EAS, three auxiliary services are presented for the power system, namely: (a) OPF analysis, (b) Transmission Lines (TLs) centrality, and (c) bus centrality. The study is conducted on standard IEEE 30 bus system to validate the results.
- The performance of the EAS is evaluated on real-world data center workload. The Shortest

Remaining Time First (SRTF) job scheduling algorithm is used for the execution of the workload of data center along with the EAS as evaluated in [16]. The EAS effect on the data center is elaborated during three intervals of the data center workload curve, such as low-load period, medium-load period, and peak-load period. Moreover, the EAS model is validated using four different job delaying criteria of the data center, such as no cost of delaying jobs, equal cost of delaying jobs, number of CPU utilization-based cost of delaying jobs, and execution time-based cost of delaying jobs.

- We defined a dynamic SLA for power systems and data centers that will motivate the data center to provide EAS. The SLA ensures data center's revenue maximization during emergency conditions for the power system. The SLA is tested for capricious data center workload, utility price, and job sizes at different time instances in a day. Moreover, the effect of an incomplete EAS is estimated on the data center revenue.

The rest of our paper is organized as follows. Section 2 details the system modelling. Section 3 presents a network setup. The results are analyzed in Section 4. Section 5 concludes the paper and provides directions of the future work.

2. System Model

The system model consists of a data center module and an EAS module. First, the notations are introduced defining dynamic SLA and revenue model. The architectural view of the proposed system model is shown in Fig. 1. We assume M_{max} homogenous computing machines for the computation of cloud workload based on a predefined SLA with the cloud customers and data center. A significant portion of the earned revenue is used for purchasing the power supply. On the other hand, the power system's revenue is determined by demand-supply management, performance in steady-state, and stability. During an emergency condition, a fast computing workstation is required to compute EAS for reliability considerations.

Therefore, in our system model, this intelligent computing workstation is the data center and a dynamic SLA is defined between the power system and data center to accomplish EAS. The system model is further divided as (a) data center module, (b) EAS model, (c) SLA, and (d) revenue model.

2.1 Data Center Module

The data center module consists of electricity pricing tariffs selection and workload-based power consumption calculation. The revenue of the data center is highly dependent on electricity pricing and power consumption. Therefore, it is essential to include both entities in this section. The power consumption is calculated on an hourly basis with and without the workload of the EAS.

2.1.1 Electricity Price

There are mainly two electricity pricing tariffs (a) regulated and (b) deregulated based on power market [3]. In a regulated power market, a constant hourly electricity price persists throughout the day. Whereas, in the deregulated market, the price is dynamic depending upon the variation in the wholesale electricity market [3]. The deregulated power market mainly offers (a) time-of-use pricing, (b) real-time pricing, and (c) 24-hours ahead pricing.

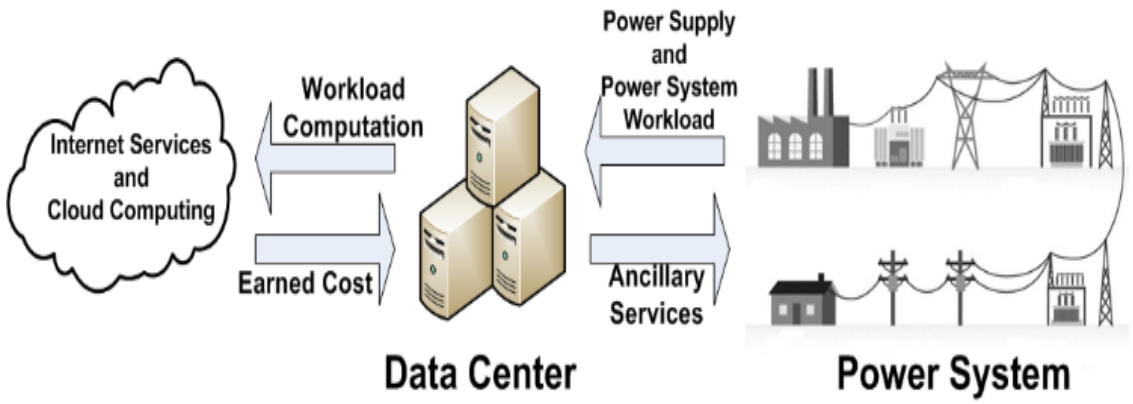


Fig. 1: Architectural overview of the system model

2.1.2 Power Consumption

The data center power consumption includes the power consumption of servers, network equipment, cooling plants, and lighting facilities. The power consumption is calculated using Eq. 1 [8].

$$P = M[P_{idle} + (PUE - 1)P_{Peak} + (P_{Peak} - P_{idle})C] \quad (1)$$

where P_{idle} is the power consumed by an idle server, P_{Peak} is the peak power consumption (averaged) of a server. The term $M \leq M_{max}$ denotes the sum of 'on' servers, C denotes the average CPU utilization of the servers, and the power usage effectiveness of the servers is denoted by Power Usage Effectiveness (PUE) [19].

2.2 Emergency Auxiliary Services

The power system network architecture contains the buses that are connected through TLs. The size of the network is measured by the number of buses present in the system. The power generators and electric loads are directly connected with the buses, which inject and consume power from the Power Transmission Network (PTN), respectively. This PTN topology is appropriate to analyze power flow within the system [20]. The three main EAS are discussed as follows.

2.2.1 Optimal Power Flow Analysis

The OPF provides optimum solution of the economic dispatch, voltage instability, TLLs problem, and directly linked with the cascading failures/ blackouts. In OPF model, the standard power flow equations are used to balance complex power at each bus. The complex power balancing is the equality constraint. The power flow on TLs and bus voltage limitations are inequality. The calculation of OPF requires some known parameters, such as PTN characteristics, generators' limits and generation cost function, and PTN topology [21]. The main objective function of OPF is the minimization of power generating cost that is defined as:

$$min_u \sum_{i=1}^{N} Co_i(P_i) \quad (2)$$

where the cost of bus i denoted by Co_i corresponds to the active power P_i of bus i. The other objective functions are (a) minimizing the changes in control variables where the vector of control variables is denoted by u,

$$\min_u \sum_{i=1}^{N_u} Co_i|u_i - u_i^0|, \quad (3)$$

and (b) TLLs minimization using a function F:

$$F = \sum_{i=(k,j)} \left(G_k \left(V_i^2 + V_j^2 - 2V_iV_j \cos(\delta_{ij})\right)\right) \ \forall\ k \in N_b \quad (4)$$

In Eq. (4), the V_i and V_j are the corresponding voltages of the buses i and j, respectively. The conductance of TL k is denoted by G_k, the N_b denotes total TLs in the network, and δ_{ij} represents the difference in the voltage angles of bus i and bus j.

Constraints of the OPF Algorithm: The power balance equations (equality constraints) are defined as:

$$P_k^G - P_k^L = \sum_{i=1}^{N} V_kV_i[G_{ki}\, cos(\theta_k - \theta_i) + B_{ki}\, sin(\theta_k - \theta_i)].$$

$$Q_k^G - Q_k^L = \sum_{i=1}^{N} V_kV_i\left[G_{ki}\sin(\theta_k - \theta_i) + B_{ki}\cos(\theta_k - \theta_i)\right] \quad (5)$$

Compact Expression: $G(x, y, u) \cong 0$

The P_k^G and P_k^L are the active powers of the generator and load, respectively, Q_k^G and Q_k^L are the reactive powers of the generator and load, respectively, G_{ki} and B_{ki} are mutual conductance and susceptance, respectively, and the phasor angle is denoted by θ_k. The x is the state variables vector and y is the parameter vector. For Inequality constraints, limits are imposed on all control variables as: $\underline{u} \leq u \leq \overline{u}$, the real power flow has an operating limit of $|P_{ij}| \leq \overline{P_{ij}}$, the operating limits on voltages are $\underline{V_j} \leq V_j \leq \overline{V_j}$, and $H(x, u, y) \geq 0$ is the compact expression for inequality constraints.

2.2.2 Transmission Lines Centrality

We calculate TL centrality for detecting faults and avoiding cascading failure in the power system. When a single or multiple TL failures occur, the OPF is required to be calculated again to balance the PTN due to increase in AC power flow on other connected TLs. In TL centrality calculation, all such TLs are identified that can initiate cascade failure in the PTN. The TLs centrality is based on the AC power flow ratio.

2.2.3 Bus Centrality

The bus centrality is an equally important index for PTN identifying most central bus in the PTN [5]. The common method of centrality calculation is the Eigenvector centrality because in this centrality measure, a centrality value γ is assigned to all isolated buses in the PTN [5]. The

mathematical expression of the centrality measure is defined as:

$$PR_i = \sigma \sum_j A_{ij} \frac{PR_j}{O_j^{out}} + \gamma \quad (6)$$

where α represents the damping factor coefficient $0 < \sigma < 1$, the adjacency matrix is denoted by A, $A_{ij} = 1$ means the bus i is directly connected with bus j, zero otherwise, the PR_j in the bus j centrality, which is adjacent to bus i, and out-degree of bus j is presented by O_j^{out}. The out degree indicates the number of buses directly taking power from bus j. The main issue in the expression of Eq. (6) is if any bus has an out degree equal to zero, then the expression will be undefined. Therefore, set O_i^{out} =1 for all such buses having zero out degree.

2.3 Service Level Agreement

The SLA is defined between the power system and data center to minimize the revenue loss for both entities when the EAS will be executed on the data center on priority. The data center's workload has varying job timings and the nature of workload is stochastic [6]. Therefore, some important factors are necessary to calculate for defining SLA, such as (a) how many computing servers of the data center are required to execute EAS at any time instant, (b) time taken by the data center to execute EAS, and (c) total bearable revenue loss for the data center to execute EAS on priority.

If the EAS initiates at peak workload time, the data center must delay some other cloud computing jobs to execute the EAS request on priority. However, if the preemption time of other non-priority jobs is more than a firm time, the revenue of the data center can be affected. Moreover, a mechanism must be determined to predict the acceptable time delay in the execution of EAS on data centers. Furthermore, if the EAS request is delayed, how much revenue loss data center can bear as per the SLA. Therefore, the SLA incorporates all the aforementioned issues. The SLA between power system and data center is similar to Amazon Elastic Compute Cloud (EC2) SLA definition, which states "If the data center face delay in a job by 10% of its total agreed execution time, then data center will pay back a penalty (service credit) of 10% of the agreed amount ($) for the job. Moreover, a 30% penalty will be imposed on the data center, if the job delay time is more than 10% of the agreed time" [4]. The aim of the SLA is to determine a marginal time limit that minimally disturbs power system reliability and operational cost of the data center.

2.4 Revenue Model

The data center revenue mainly revolves around the power consumption cost that can be calculated using Eq. (1). In Eq. (7), a revenue model is defined for the data center to execute EAS. The mathematical expression of the revenue is written as:

$$R = \left(1 - q(\mu)\right)\left[1 + \frac{E}{L}\right] - q(\mu)\left[1 + \frac{E}{L}\right], \quad (7)$$

where μ is the job service rate expressed as:

$$\mu = kM \quad (8)$$

Suppose the time (seconds) taken by a server to finish one job is denoted by T, then in Eq. (8), the $k = {}^{1}/_{T}$ is the jobs per sec and M denotes a number of 'on' servers in the data center. In Eq. (8), job failure probability is denoted by $q(\mu)$. L denotes revenue loss of data center for preempting cloud computing jobs for priority execution of EAS. The data center demands cost L from the power system as an intensive to provide EAS. Further, E is the profit charges to execute EAS. In Eq. (7), the term $\left(1 - q(\mu)\right)\left[1 + \frac{E}{L}\right]$ calculates the total earned revenue for completing EAS within the allocated time. The term $q(\mu)\left[1 + \frac{E}{L}\right]$ refers penalty cost on a data center for delaying EAS.

3. Network Setup

The power system reliability experiment is performed on the IEEE 30-bus system. The one-line diagram of the system is shown in Fig. 2. A typical server in a data center has $\mathcal{P}_{Peak}$ = 213 watts, $\mathcal{P}_{idle}$ = 100 watts as shown in Table 1 [17]. The total capacity of the IEEE 30-bus system has 192.1 MW total generation capacity and total load is 189.2 MW. Another modification was necessary to accommodate the data center' load. Therefore, the basic load of the system is reduced to 184 MW.

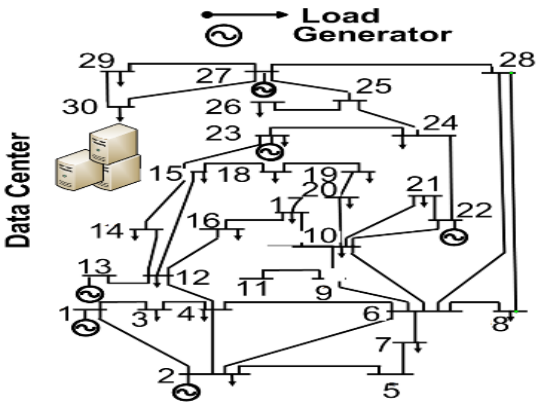


Fig. 2: The Architecture of the modified IEEE 30-bus system

The outage of TL is considered as an emergency condition when the power system required EAS for power balancing and TLLs reduction. The two states model of Markov is used to model TLs outage [7]. The Probability Density Function (PDF) of the exponential random variable is defined as:

$$F(y) = 1 - e^{-py} = x, \quad (9)$$

where Y is the random variable and $1/p$ denotes mean of the variable. The PDF is set equal to a random binary decimal number x. Then Eq. (10) is redefined as:

$$y = -\frac{\ln(1-x)}{p}. \quad (10)$$

Table 1: Power Consumption per Component in a Data Center's Server

Component	Peak Power (W)	Count	Total Power (W)
CPU	40	2	80
Disk	12	1	12
Memory	9	4	36
PC1 Slots	25	2	50
Motherboard	25	1	25
Fan	10	1	10
System Total Power			213

The occurrence of TL failure and maintenance duration is modeled by Eq. (10). Moreover, the electrical load is also random by nature; therefore, the electrical load on all load buses is modeled using a normal distribution with 9.1 MW as a nominal value. For the data center's power consumption data, a real workload profile is used that is collected from the University of New York (Buffalo) [17]. Table 2 lists the server specification. The data center workload profile is on the span 30 days from February 20, 2009, to March 22, 2009. The detailed specifications are given in Table 2.

Table 2: Workload Specifications of Data Center

Time Duration	February 20th, 2009 – March 22nd, 2009
Total jobs execute on data center	22,385
Total distinct servers	1,045
Processor name	1056 Dell PowerEdge SC1425
Processor speed	3.0 GHz or 3.2 GHz
Peak performance	13 TFlop/sec

In the dataset, the jobs were characterized on size and length. The jobs are further divided into three categories based on their length as (a) short (less than 1 hour), (b) long (greater than 1 hour and less than twelve hours), and (c) very long (greater than twelve hours). There are more than 22% of very long jobs present in the workload and are referred to as delay tolerant jobs (flexible deadlines). Moreover, there are total 110 hours in a month's time where workload exceeds 100% and jobs must wait in a queue due to unavailability of computing resources as shown in Fig. 3. Furthermore, the New York Independent System Operator (NYISO) is used to get the electricity price profile for the simulations shown in Fig. 4 [3].

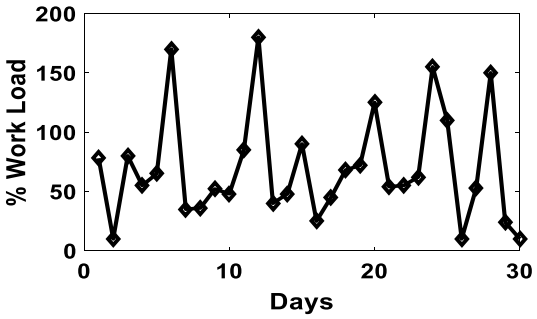


Fig. 3: Total load of the data center in a month

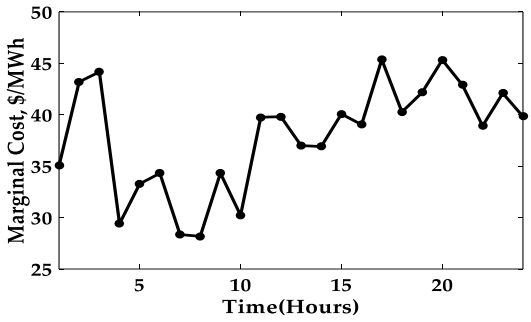


Fig. 4: Real time electricity price profile

The Shortest Remaining Time First (SRTF) job scheduling algorithm is used to execute computational workload on the data center. There may be other job scheduling techniques that can have less workload execution time, but the selection of optimal job scheduling technique does not lie in the scope of this paper. Moreover, four different scenarios are considered for the cost of delaying data center jobs under SLA like (a) no cost of delaying jobs, (b) equal cost of delaying jobs, (c) Number of CPU utilization-based cost of delaying jobs, (d) execution time-based cost of delaying jobs.

No Cost of Delaying Jobs (NCDJ): When the data center delays its own workload jobs for fulfilling power system job requirements, the only benefit data center will get from the power system is the cost that it will lose from delaying own jobs. The data center's own workload jobs will not put any penalty to the data center for the delay.

Equal Cost of Delaying Jobs (ECDJ): In this scenario, when data center's own workload jobs will be delayed, there will be equal penalty cost per minute for every job on the data center. In this case, the data center will lose some revenue than NCDJ.

Number of CPU Utilization based Cost of Delaying Jobs (CCDJ): In this scenario, the delaying job cost penalty will vary from job to job. The delaying cost of each job will depend upon the number of CPU utilization. The jobs executing on a smaller number of CPUs will be selected first for the delay because the job penalty is less compared to the jobs utilizing more CPUs [17].

Execution Time based Cost of Delaying Jobs (ETDJ): In this scenario, the jobs that are started in the near past and have many hours of execution time left will be selected first for the delay. The jobs that will be finished in the next few minutes will be the least prior jobs to be delayed. The reason for this selection criteria is because if a job is delayed, which was executing from a long time will cost more in case of halt or restart [17].

4. Results Analysis

The simulations of the proposed algorithm are performed on an SYS-7047GR-TRF system server that has 96 cores. The simulation results show that the proposed SLA-based EAS model is mutually beneficial for power systems and data centers. When EAS workload is added with the other cloud workload on the data center, the power consumption of cloud workload is used as a reference/ base calculation to compute the increase in power consumption. The average baseline power consumption per hour curve of the data center is plotted in Fig. 5 and the baseline power consumption cost of the data center is presented in Fig. 6. Due to EAS and delay in cloud workload, the excess power consumption adds an extra penalty ($) cost. Moreover, from Fig. 5, three different load periods are observed that are named as (a) peak-load period, (b) medium-load period, and (c) low-load period.

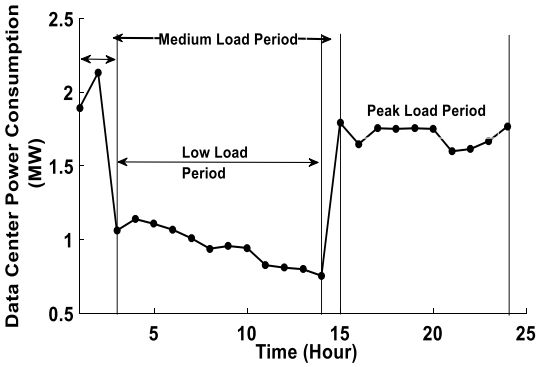


Fig. 5: Data Center per hour average power consumption during a specific month

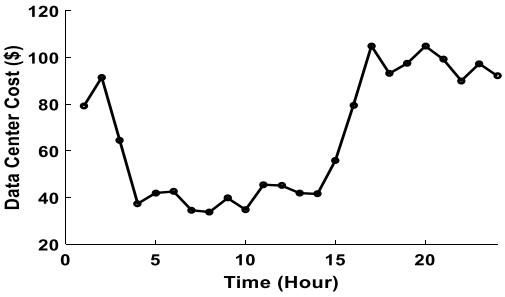


Fig. 6: Data Center power consumption cost per day

The peak-load period is witnessed from 1:00 A.M. to 2:00 A.M. and 15:00 P.M. to 24:00 A.M., medium-load period is also observed during two separated intervals of a day, such as 2:00 A.M., 3:00 A.M., 13:00 P.M. and 14:00 P.M and low-load period is observed from 3:00 A.M. to 13:00 P.M. The model is experimented for all three periods and evaluated the percentage increase in the cost of the data center by performing power system jobs on priority and delaying data center's workload jobs, if not enough servers are available for the EAS.

On a given day in a data center, the total number of jobs during the peak-load hour (16:00 P.M.) is 2112. The 101 jobs are running from more than 48 hours and will be completed in the next 8 hours on an average. The 1322 jobs are running from more than 24 hours and their expected completion time is within next 12 hours on an average, 401 jobs are running from the last 8 hours and they are expected to complete in the next 4 hours on an average. Lastly, the 288 jobs are running from the last 8 hours and their expected completion time are 10 hours on average. Similarly, at a medium-load hour (14:00 P.M.) and low-load hour (10:00 A.M.), the total number of jobs executing are 1459 and 960 respectively. Moreover, at the peak-load hour, there are three

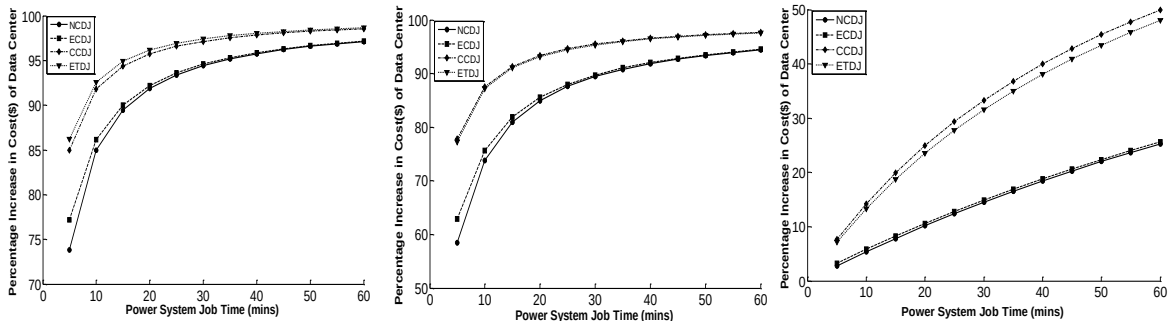
types of jobs with respect to the number of CPU utilization. There are 51 jobs, each one is utilizing more than 30 CPUs, 694 jobs are utilizing 2-10 CPUs each, and 1367 jobs are utilizing 1 CPU each. A similar ratio of CPU utilization is observed during medium and low load hours.

In-network setup, on-demand pricing criteria of Amazon is used to determine the cost ($) of the jobs, such as if a job utilizes 8 CPUs, it cost $0.840/hour [4]. Moreover, for CCDJ and ETDJ, the penalty for delaying jobs are based on the SLA defined in Section 3 [4]. Furthermore, for CCDJ, whenever two or more jobs utilize the same number of CPUs and one of the jobs required to be delayed, the first job in the list will be delayed. Similarly, in ETDJ, when two or more jobs are running for an equal amount of time and one must be delayed, the first job in the list will be delayed. The close observation of Fig. 4 illustrates that electricity price is also high during the peak-load period of the data center. Therefore, the data center workload-based power consumption cost also illustrates a similar pattern to the electricity price profile as shown in Fig. 6.

4.1 Impact of SLA based EAS Model on Data Center

The EAS job timing is varied from five minutes to one hour for testing and validation of the model. The power consumption pattern of the users is always variable over the period of the day and this phenomenon must be considered in the simulations. Particularly, when the model-based algorithm's output back into the power system via a control mechanism. Therefore, the OPF model validity can be found in the 5 minutes' time to approximately 1-hour time frame [18]. In data centers, the servers' utilization for performing the EAS can also vary depending upon the requirements of parallel computing. Therefore, the model is evaluated under 100%, 50%, and 1% data center's server utilization. Experimental results in Fig. 7, Fig. 8, and Fig. 9 represent the percentage increase in the data center cost ($) at peak-load, medium-load, and low-load periods, respectively. The three load periods are discussed under NCDJ, ECDJ, CCDJ and ETDJ scenarios. We calculate the cost ($) of the data center power consumption due to the power system's job execution and delaying of the data center's workload jobs.

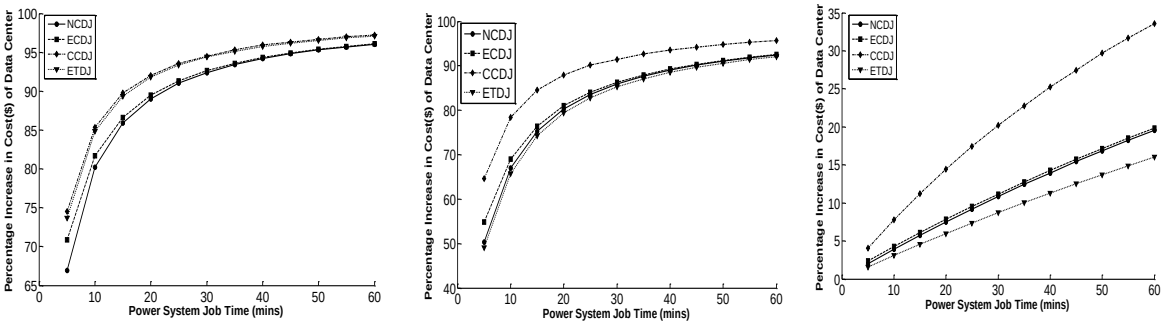
The cost is compared with the base power consumption cost shown in Fig. 6. Moreover, the results presented in Fig. 7, Fig. 8, and Fig. 9 are used for the revenue calculation of the data center. In Fig. 7, the percentage increase in the cost of CCDJ and ETDJ has remained higher than the NCDJ and ECDJ in all three servers' utilization scenarios. The reason for this phenomenon is the variable cost for each job delay in CCDJ and ETDJ. In Fig. 7 (a), the ETDJ approach cost more compared to CCDJ approach because the delay cost of the jobs that were running for more than 48 hours are proved costlier than the jobs utilizing a higher number of CPUs.



(a) 100% Data center servers utilization *(b) 50% Data center servers utilization* *(c) 1% Data center servers utilization*

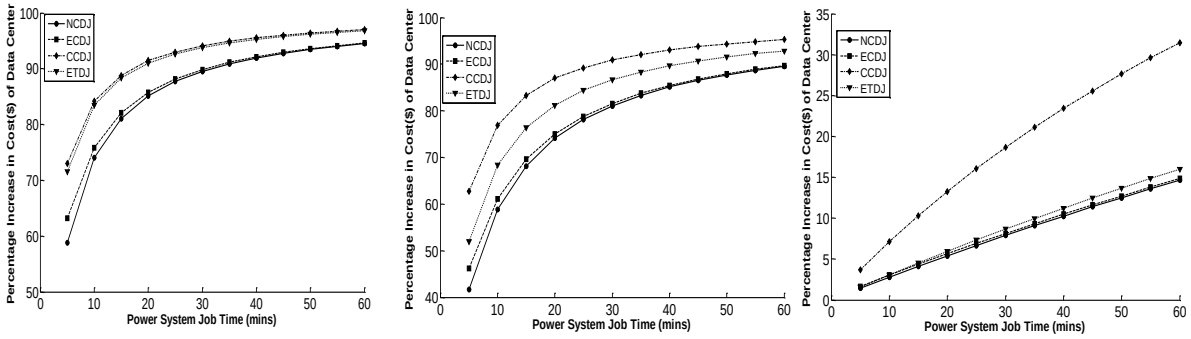
Fig. 7: Impact on data center's cost for executing EAS at peak-load hour.

However, for the 50% server utilization at peak-load, the job delaying cost is similar for both CCDJ and ETDJ because neither the long running jobs were delayed nor the jobs utilizing a higher number of CPUs.The aforesaid description is depicted in Fig. 7 (b). Fig. 7 (c) presents the results of the power system jobs that only require 1% of data center servers for the execution. The power system jobs that have the least execution time in ETDJ cost less compared to the jobs utilizing the least number of CPUs in CCDJ. In Fig. 8, during the medium-load period, delaying jobs using ETDJ cost less compared to CCDJ scheme. The effect is more prominent in the case



(a) 100% Data center servers utilization *(b) 50% Data center servers utilization* *(c) 1% Data center servers utilization*

Fig. 8: Impact on data center's cost for executing EAS at medium-load hour



(a) 100% Data center servers utilization *(b) 50% Data center servers utilization* *(c) 1% Data center servers utilization*

Fig. 9: Impact on data center's cost for executing EAS at low-load hour

.of 50% and 1% servers' utilization, when the ETDJ scheme proves to be less expensive in contrast to NCDJ, ECDJ, and CCDJ. The reason for the aforementioned effect is the jobs that were running from more than 24 hours were kept on running without any delay that results in overall less penalty. Therefore, the cost of the ETDJ scheme appeared to be more than NCDJ and ECDJ but less than CCDJ as shown in Fig. 9.

4.2 Impact of EAS Model on Power System

The cost function for the OPF optimization algorithm is to minimize the TLs power losses. The power loss is calculated with the real power discrepancy between sending and receiving bus of the TL. We arbitrary outage TLs for a random time. Whenever such outage occurs, the situation is known as an emergency and need to recalculate the OPF for power balancing. If the optimization algorithm does not converge, then the blackout/ system failure can occur. Further, if the algorithm converges but the TLLs exceed the threshold of 9.607MW, the system is still considering as in failure state. Fig. 10 illustrates the convergence iterations of the OPF algorithm.

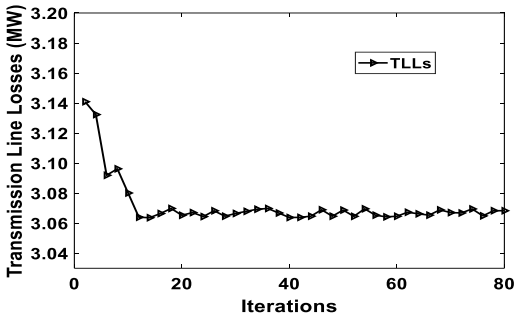
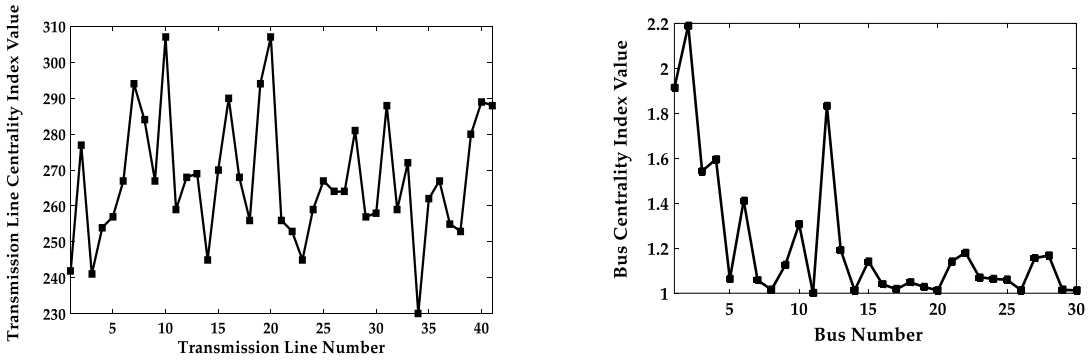


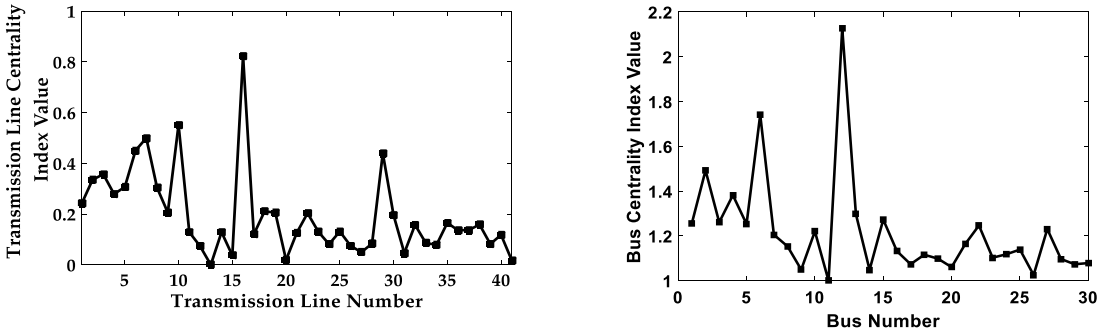
Fig. 10: TLLs convergence using OPF algorithm when 7 TLs are outage

The OPF converges in 5 to 7 iterations in most situations. The least TLLs achieved for 7 TLs outage case is 3.0771MW, which is less than 5% of the total generated power of the system and this power loss is acceptable in any real power system. Moreover, the TL centrality is calculated as shown in Fig. 11 (a). In IEEE 30-bus system, the TL centrality index indicates that TL 10 and TL 20 are more prone to failure because the TL index value exceeds the threshold limit as shown in Fig. 11(a).



(a) Number of TL fails lead to the system (b) Bus Centrality

Fig. 11: The status of the 30-bus system when N-k transmission lines are outage



(a) TL Centrality (b) Bus Centrality

Fig. 12: The status of the 30-bus system after successful execution of the EAS on the data center

Moreover, Fig. 11 (b) shows the bus centralities, the buses with high centrality values are more vigilant to failure because these buses are linked with the TLs having high TL centralities and the electrical load attached on the buses is also high compared to the other buses. Fig. 11 (b) illustrates that Bus 1, Bus 2, Bus 4, and Bus 12 are the most central buses in the 30-bus test system. Therefore, an optimized OPF solution is required from the data center to balance the power flow on endangering TLs. Once the overloading of TLs is balanced, the centrality index of critical buses will also reduce. Fig. 12 (a) shows the optimized TL centrality on TLs after the execution of EAS on the data center. In Fig. 12 (a), it is observed that the TL centrality of TL 10 and TL 20 has reduced into normal operating range. Moreover, Fig. 12 (b) shows the bus centralities after the execution of EAS. The comparison of Fig. 11(b) and Fig. 12 (b) illustrates the observable reduction in the bus centralities. Therefore, it is concluded that the computational capability of the data center can effectively be utilized for the steady-state operation of the power system. The proposed SLA works perfectly if there is no delay in the EAS, however, the delay in EAS will cause revenue loss for the data center.

Fig. 13 illustrates the revenue curve and incentive margin for the data center under the influence of EAS failure. Fig. 13 is the resultant graph of Eq. (10). In Fig. 13, the 100% revenue means, the given incentives E becomes the profit of the data center. The 0% revenue indicates that the revenue loss and incentives E of a data center due to the delay of own cloud jobs equalize each other and earning no extra profit. Moreover, the graph below 0% shows that the data center is not only losing power consumption cost but also losing own revenue that is earned from another workload.

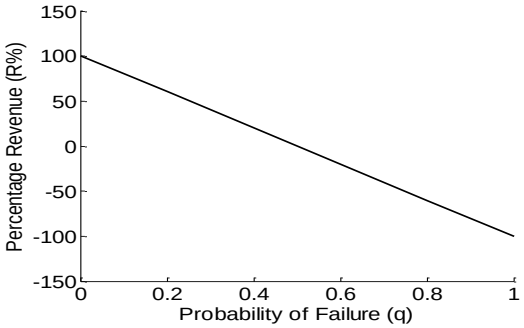


Fig. 13: Relationship of job failure with data center revenue under SLA

Fig. 13 also shows that the bearable minimum failure rate (q) of a single EAS job is 0.48 over the period of a day. If the failure rate is more than 0.48, then the data center will lose revenue.

5. Conclusions and Future Work

An Emergency Auxiliary Services (EAS) model is proposed in the paper for power systems and data centers. The idea is to use the computational capability of the data center to ensure stability of the power system in steady-state rather than using the dedicated servers for the power system operations. The proposed EAS model results also illustrate the data center's revenue maximization. The EAS includes a useful tool for the power system as the converged OPF solution ensure reduction in transmission line losses and indicate health of transmission lines and buses by using centrality concepts. Finally, using experiments on the real-world workload of the data center and IEEE standard bus system, it was concluded that the EAS optimized solution in certain conditions is convex. Because the wide-area power system is very large and required intensive computation for maintaining system reliability. The proposed EAS model will be extended to multiple data centers attached to the power system and selection of data center to fasten the response time of EAS.

6. References

[1] Glanz, J., & Sakuma, P. (2011). Google details, and defends, its use of electricity. The new york times, 8.

[2] Yao, J., Liu, X., He, W., & Rahman, A. (2012, June). Dynamic control of electricity cost with power demand smoothing and peak shaving for distributed internet data centers. In 2012 IEEE 32nd International Conference on Distributed Computing Systems (pp. 416-424). IEEE.

[3] "NYISO Home - NYISO," https://www.nyiso.com/, accessed April 2019.

[4] "Service Level Agreement – Amazon Simple Storage Service (S3) – AWS," https://aws.amazon.com/s3/sla/, accessed April 2019.

[5] Newman, M. (2018). Networks. Oxford university press.

[6] Kiani, A., & Ansari, N. (2018). Profit Maximization for Geographically Dispersed Green Data Centers. IEEE Transactions on Smart Grid, 9(2), 703–711.

[7] Mei-Ling TL (2007). A two-state markov model. Jounal of Communications in Statistics - Theory and Methods, 23(2), 615-623.

[8] Patel, S., & Vyas, K. (2019). Current Methodologies for Energy Efficient Cloud Data Centers. In Information and Communication Technology for Competitive Strategies (pp. 429-437). Springer, Singapore.

[9] Jawad, M., Qureshi, M. B., Khan, U., Ali, S. M., Mehmood, A., Khan, B., & Khan, S. U. (2018). A robust Optimization Technique for Energy Cost Minimization of Cloud Data Centers. IEEE Transactions on Cloud Computing.

[10] Liu, Z., Lin, M., Wierman, A., Low, S. H., & Andrew, L. L. (2011, June). Greening geographical load balancing. In Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems (pp. 233-244). ACM.

[11] Adnan, M. A., Sugihara, R., & Gupta, R. K. (2012, June). Energy efficient geographical load balancing via dynamic deferral of workload. In 2012 IEEE Fifth International Conference on Cloud Computing (pp. 188-195). IEEE.

[12] Wang, P., Rao, L., Liu, X., & Qi, Y. (2012). D-Pro: Dynamic data center operations with demand-responsive electricity prices in smart grid. IEEE Transactions on Smart Grid, 3(4), 1743-1754.

[13] Wang, H., Huang, J., Lin, X., & Mohsenian-Rad, H. (2014). Exploring smart grid and data center interactions for electric power load balancing. ACM SIGMETRICS Performance Evaluation Review, 41(3), 89-94.

[14] Jiang, C. (2012, September). System level power characterization of multi-core computers with dynamic frequency scaling support. In 2012 IEEE International Conference on Cluster Computing Workshops (pp. 73-79). IEEE.

[15] Gurumurthi, S., & Sivasubramaniam, A. (2012). Energy-Efficient Storage Systems for Data Centers. In Energy-Efficient Distributed Computing Systems (pp. 361-376). John Wiley and Sons.

[16] Ali, S. M., Jawad M., Khan M.U. S., Bilal K., Glower J., Smith S. C., Khan S. U., Li K., and Zomaya A. Y. (2017). An Ancillary Services Model for Data Centers and Power Systems. IEEE Transactions on Cloud Computing.

[17] Chandio, A. A., Bilal, K., Tziritas, N., Yu, Z., Jiang, Q., Khan, S. U., & Xu, C. Z. (2014). A comparative study on resource allocation and energy efficient job scheduling strategies in large-scale parallel computing systems. Cluster Computing, 17(4), 1349-1367.

[18] Zimmerman, R. D., Murillo-Sánchez, C. E., & Thomas, R. J. (2010). MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. IEEE Transactions on power systems, 26(1), 12-19.

[19] Horner, N., & Azevedo, I. (2016). Power usage effectiveness in data centers: overloaded and underachieving. The Electricity Journal, 29(4), 61-69.

[20] Murty, P. S. R. (2017). Power systems analysis. Butterworth-Heinemann.

[21] Rahul, J., Sharma, Y., & Birla, D. (2012, November). A new attempt to optimize optimal power flow based transmission losses using a genetic algorithm. In 2012 Fourth International Conference on Computational Intelligence and Communication Networks (pp. 566-570). IEEE.