Scalable User Assignment in Power Grids: A Data Driven Approach *

Bo Lyu[\$], Shijian Li[†], Yanhua Li[†], Jie Fu[†], Andrew C. Trapp[†], Haiyong Xie[\$], Yong Liao[\$]
[†] Worcester Polytechnic Institute (WPI), MA USA
[\$] China Academy of Electronics and Information Technology, Beijing, China

ABSTRACT

The fast pace of global urbanization is drastically changing the population distributions over the world, which leads to significant changes in geographical population densities. Such changes in turn alter the underlying geographical power demand over time, and drive power substations to become over-supplied (demand $\ll$ capacity) or under-supplied (demand $\approx$ capacity). In this paper, we make the first attempt to investigate the problem of power substation-user assignment by analyzing large-scale power grid data. We develop a **S**calable **P**ower **U**ser **A**ssignment (SPUA) framework, that takes large-scale *spatial power user/substation distribution data* and *temporal user power consumption data* as input, and assigns users to substations, in a manner that minimizes the maximum substation utilization among all substations. To evaluate the performance of our SPUA framework, we conduct evaluations on real power consumption data and user/substation location data collected from a province in China for 35 days in 2015. The evaluation results demonstrate that our SPUA framework can achieve a 20%–65% reduction on the maximum substation utilization, and 2 to 3.7 times reduction on total transmission loss over other baseline methods.

CCS Concepts

•**Energy distribution** → **Power networks;** •**Mathematics of computing** → *Combinatorial optimization;* •**Theory of computation** → Distributed algorithms;

Keywords

Power grid; user assignment; optimization

1. INTRODUCTION

Electricity has become an indispensable necessity in our daily lives, powering the machines that keep our homes, businesses, schools and hospitals safe, comfortable and convenient. As the fast development of sensors, monitoring devices, such as smart meters, a large amount of power grid data are generated over time, including temporal energy consumption data, spatial user/substation distribution data, and so on. All these heterogeneous data sources offer new research and technological opportunities, and enable intelligent solutions for various applications in power grids.

A power grid consists of a network of power plants and power substations that provide electricity power to a wide range of power users. Each power substation has a certain power capacity, that limits the total power demand it can serve. However, the fast pace of global urbanization leads to significant changes on geographical population densities, thereby altering the underlying geographical power demand over time. For example, from large-scale power consumption data from many cities in China, the rapid expansion of urban population sizes has driven regional power demands to the capacity limits of the nearby power substations. On the other hand, as the population density changes over time, some power substations cover power users that are 300 km away, leading to high transmission losses. We are thus motivated to investigate how to reduce substation power utilization, and prevent them from being overloaded or over-supplied.

In this work, we make the first attempt to investigate the power user assignment problem in large scale power grid. The design goal is to have a scalable solution to assign each power user to one substation, while minimizing the maximum substation utilization. We develop an **S**calable **P**ower **U**ser **A**ssignment (SPUA) framework, which takes the spatial power user/substation distribution, and temporal user power consumption data as input, and performs optimal user assignment to substations, that minimizes the maximum substation utilization among all substations. Our main contributions are summarized as follows.

- We formulate the power user assignment problem using integer programming, which is NP-hard. We employ a 2-approximation algorithm to solve the problem via linear programming (LP) relaxation.
- To evaluate the performance of our SPUA framework, we conduct evaluations on real power consumption data and user/substation location data collected from a province in China for 35 days. The evaluation results demonstrate that our SPUA framework can achieve a 20%-65% reduction on the maximum substation utilization, and 2 to 3.7 times reduction on total transmission loss. Moreover, *we plan to make the power grid dataset available to facilitate the re-*

*This work was primarily done when Bo Lyu was a visiting scholar at Worcester Polytechnic Institute.

SIGSPATIAL'16, October 31-November 03, 2016, Burlingame, CA, USA

DOI: http://dx.doi.org/10.1145/2996913.2996970

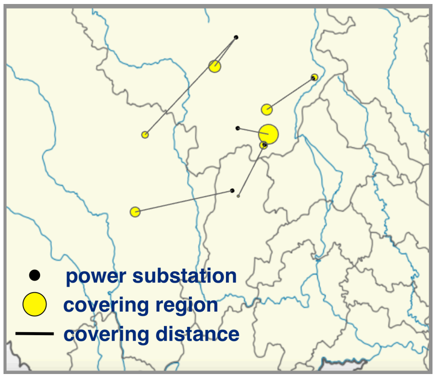

Figure 1: Long-distance coverage

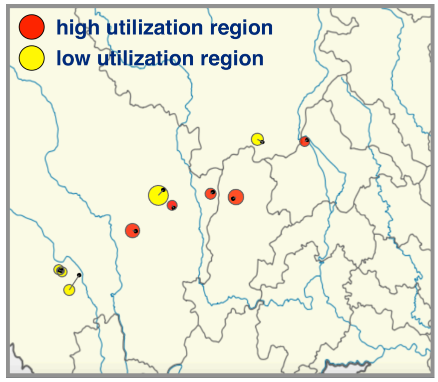

Figure 2: Under- and over-supplied substations (peak hours)

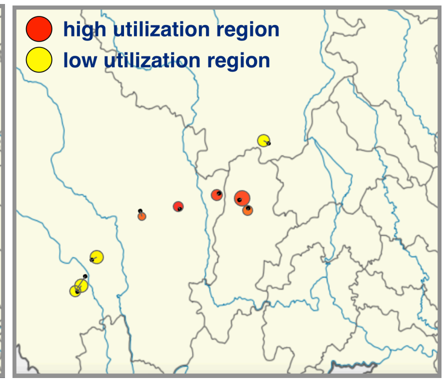

Figure 3: Under- and over-supplied substations (valley hours)

search community once the paper gets accepted.

The rest of the paper is organized as follows. Section 2 presents the datasets and motivations. Section 3 presents the framework and detailed methodology of SPUA. Section 4 presents evaluation results on a real large-scale power consumption dataset. The paper is concluded in Section 5.

2. MOTIVATIONS

In this section, we describe the dataset we use and motivate the power user assignment problem.

2.1 Datasets

The dataset we use for this study includes (1) power user profiles, (2) power substation profiles, and (3) temporal user power consumption data. The datasets were collected from a province in China during March 10th – April 13th in 2015.

Power user locations. The dataset contains in total 6.3 million unique users, with their unique *user IDs*. Note that users include 6.16 million residential users and 0.14 million commercial and industrial users.

Power substation locations and capacities. At the time of data collection, there were 783 power substations deployed in the province. Each substation has a *substation ID*, *address* and *substation capacity*, namely, the maximum electrical power it can provide per hour. We parsed the addresses into locations in latitude and longitude using BAIDU Geo-Coding APIs, and cross-validated using Google Geo-Coding APIs. There are about 25% user records with missing or incomplete addresses, which were therefore eliminated from the dataset.

Temporal user power consumption data. This dataset contains both the user-substation assignment information and the dynamic power usage for each individual user. Each user with a user ID *uid* is uniquely assigned to a substation *sid*. Moreover, the dataset contains the power usage for all users over 35 days. For each user, the dataset records the total daily power consumption, and the power consumptions for peak hours (9AM-1PM and 9PM–1AM), plain hours (1PM–9PM), and valley hours (1AM–9AM), respectively.

2.2 Motivations

A power grid consists of a network of *power plants* and *power substations*. A power plant is an industrial facility for the generation of electric power, which contains one or more generators. A power substation as a part of an electrical generation, transmission, and distribution system, transforms voltage from high to low, or the reverse. Moreover, a power substation could serve a group of power consumers.

By analyzing the datasets, we obtained interesting observations: Due to the global urbanization and human mobility, the population size and density change geographically over time, which drives the needs to upgrade the power grid network infrastructure for two main reasons, including long distance user coverage and over- and under-supplied power substations.

Long distance user coverage. The electrical power transmission incurs certain transmission cost. The longer the user is from the substation, the more power transmission loss [3]. Studies have shown that the power transmission is proportional to the square of transmission distance. From the real data, we observe that many users are covered by a long distance power substation, rather than a nearby one. Figure 1 shows five power substations in the province that cover users who are 300 km away from the substation or more.

Over- and under-supplied power substations. A power substation when being designed and deployed has a certain capacity, namely, a maximum amount of electrical power can be provided per unit time (e.g., one hour). Over time, the power demand of some power substations may increase drastically, and exceed the substation capacity, leading to under-supplied scenario. On the other hand, the population density may decrease in the regions covered by some power substations, which would lead to over-supplied scenario, where the substation utilization becomes lower. For example, Figure 2 and 3 show the substations with highest and lowest power utilization during peak and valley hours, respectively. For those busy power substations, they are primarily located in regions with high population densities, such as downtown of Urumqi City.

Motivated by these observations, we aim to develop a scalable power user assignment framework, that assigns each user to a power substation by analyzing large-scale power consumption data, which maintaining low substation utilizations. Next, we define the power user assignment problem [1].

[1]Besides distribution automation through reassigning the users to substations, there are alternative methods to tackle the above two challenges, including upgrading/degrading the substation capacity or deploying/removing new power substations. However, those methods are more costly in terms of redeployment cost [1], and reassignment of users and substations are still needed after applying these methods. Thus, in this paper, we focus on the solution based on reassigning users to substations.

3. FRAMEWORK

Figure 4 presents our scalable power user assignment (SPUA) framework. It takes three datasets as inputs, including power user profiles, power substation profiles, and user power consumption. The whole framework consists of two stages (highlighted as two dashed boxes): (1) user aggregation, and (2) user assignment.

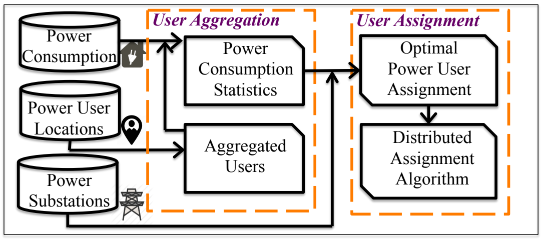

Figure 4: Scalable power user assignment

3.1 User Aggregation

In a real power grid system, due to various system constraints it is not possible to assign individual users to just any substation. For example, users on the same distribution line or transformer have to be assigned/switched to the same power substation. To consider such constraints, we aggregate the power users with the same or close locations to an aggregated super user, and conduct the user assignment for aggregated users.

We use a granularity of 0.0005 degrees in latitude and longitude, roughly 50 meters distance, to aggregate users. It is worth mentioning that we only aggregate residential users (who tend to have lower amounts of power consumption), not commercial or industrial users. After the aggregation, we extracted $m = 21,801$ aggregated users from 6.3 million individual users. For simplicity and conciseness, we will use power users to refer to aggregated power users throughout the remainder of this paper.

Given a group of individual users who form an aggregated user, we sum up all power consumed by individual users to extract the power consumption for the aggregate user. For each aggregated user $j \in U_a$, we extract the average hourly power consumption $d_j \in D_a$ during peak hours.

3.2 User Assignment

3.2.1 Problem Formulation

Given a set of substations S with capacity C, (aggregated) users U_a, together with the average user peak hour demand D_a, we are now in a position to formulate the power user assignment problem, with the goal of minimizing the maximum power substation utilization. Given a user $j \in U_a$, the total hourly power consumption for assigning it to substation $i \in S$ is $p_{ij} = d_j + \alpha d_j^2 \mathsf{dist}_{ij}$, which contains d_j the actual average hourly power consumption during the peak hours and $\alpha d_j^2 \mathsf{dist}_{ij}$ the transmission loss incurred by transmitting d_j amount of power from the substation i to user j. Note that we use the average hourly user power demand during peak hours $D_a = [d_j]$ instead of over all 24 hours, because the highest power utilization of substations in general occurs during peak hours. The transmission loss is a product of a system factor α, the (Euclidean) distance dist_{ij} (in kilometers) between station i and user j, and the square of user j's hourly power consumption in peak hours d_j^2. Thus, the substation power utilization ℓ_i is the ratio between the total user power demand with the operation cost by transmission loss p_{ij} and the substation capacity c_i, namely, $\ell_i = \sum_j p_{ij} x_{ij} / c_i$. Each $d_j \in D_a$ is extracted from the past power consumption data in the user aggregation stage. Let ℓ be the maximum substation power utilization. We denote a decision variable x_{ij} as a binary indicator variable, indicating that a user $j \in U_a$ is assigned to a station $i \in S$ when $x_{ij} = 1$, and $x_{ij} = 0$ otherwise. We aim to find the optimal assignment of all x_{ij} values that leads to the smallest possible ℓ. This problem is formally formulated as below.

$$\text{min:} \quad \ell \tag{1}$$

$$\text{s.t.:} \sum_{j \in U_a} \frac{p_{ij}}{c_i} x_{ij} \le \ell, \qquad \forall i \in S, \tag{2}$$

$$\sum_{i \in S} x_{ij} = 1, \qquad \forall j \in U_a, \tag{3}$$

$$x_{ij} = \{0, 1\}, \quad 0 \le \ell \le 1, \qquad \forall i \in S, j \in U_a. \tag{4}$$

The objective function eq.(1) is to minimize the maximum utilization ℓ for all power substations. The constraint in eq.(2) indicates the power substation capacity constraint, namely, for a substation $i \in S$, the substation power utilization ℓ_i is no more than the maximum power utilization ℓ. The validity constraint in eq.(3) indicates that any power user is covered by exactly one power substation.

3.2.2 Optimal Power User Assignment

The above integer linear programming (ILP) problem can be viewed as a *makespan scheduling problem with unrelated machines* or *scheduling on unrelated parallel machines*. The problem is NP-hard and has been extensively studied in the literature, with a variety of approximation algorithms proposed that employ LP-rounding approaches. In this study, we adopt the approximation solution algorithm proposed in [2] based on LP-rounding. Other algorithms can be chosen, depending on the specific requirements on the error bound and complexity.

Lemma 1. *Algorithm 1 assigns each power user in U_a to one substation in S, and the maximum substation utilization ℓ obtained by such assignment is no more than $2\ell^*$, where ℓ^* is the optimal objective value to the problem eq.(1)–(4).*

Algorithm 1 Approximate Power User Assignment Algorithm

Input: U_a, S, D_a, α, dist_{ij};
Output: $x_{ij} \in \{0, 1\}$, ℓ;
for $j \in S_a$ **do**
    $y_{ij} = 1$, if $i = \text{argmin}_{i \in S}\{p_{ij}/c_i\}$, and 0, otherwise;
$\beta = \max_i \sum_{j \in U_a} p_{ij} y_{ij} / c_i$;
Binary search ℓ in $[\beta/n, \beta]$ for smallest ℓ that $LP(\ell)$ has a feasible solution $[x_{ij}]$;
Construct bipartite graph H and find perfect matching M;
Round in $X = [x_{ij}]$ all fractionally set jobs according to the matching M;

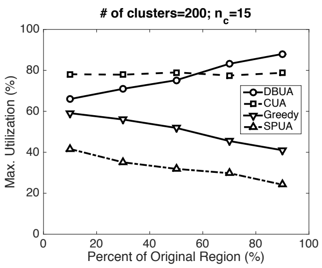

Figure 5: Max. utilization vs problem scale

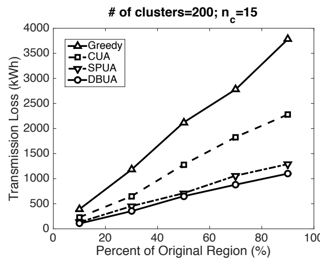

Figure 6: Transmission loss vs problem scale

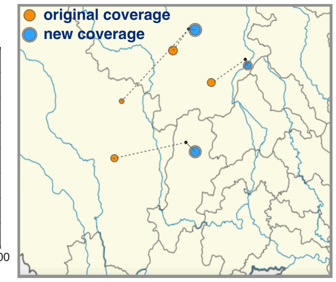

Figure 7: Reduced covering distance

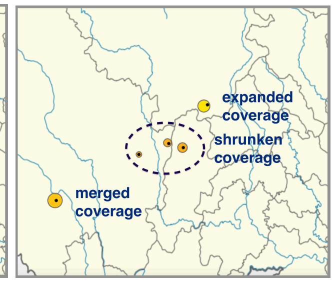

Figure 8: Balanced substation utilization

4. EVALUATIONS

We evaluate our proposed SPUA using two performance metrics, including maximum substation utilization (max. utilization) and total transmission loss (in kWh). We primarily compare our SPUA method with three baseline algorithms, including the current user assignment (CUA), Distance-based user assignment (DBUA), and greedy method (Greedy).
(1) Current user assignment (CUA). This baseline algorithm employs the substation-user assignments observed from the real dataset.
(2) Distance-based user assignment (DBUA). This baseline algorithm simply assigns each user to its closest substation.
(3) Greedy method (Greedy). The idea behind this baseline algorithm is that we want to incrementally assign users to substations, so as to keep each substation with the relatively same utilization.

To evaluate the salability of the proposed framework, we change the problem scale by choosing sub-regions with varying sizes, i.e., from 10% to 90% size of the entire dataset. For each size, e.g., 10%, we randomly generate 100 sub-regions, and take the average of the result from each region, to reduce the effect of randomness.

As shown in Figures 5, we observe that our SPUA method has the lowest maximum substation utilization comparing all baseline methods, with a significant improvement ranging from 20% (over Greedy) to 65% (over DBUA at the scale of 90% original region size). As the size of the sub-region increases from 10% to 90%, the maximum substation utilization decreases with our SPUA method and Greedy method. The reason is that a larger underlying sub-region generally contains a larger number of users and substations, thus allows larger flexibility for SPUA and Greedy to assign and shift users across substations, leading to lower maximum substation utilization. Since the user assignment with CUA does not change with the sub-region scale, the maximum substation utilization stays the same over sub-region sizes.

Similarly, when looking at the total transmission loss (in kWh), our SPUA always achieves lower total transmission loss over CUA and Greedy methods (as shown in Figure 6), with 2 to 3.7 times reduction. Notice that DBUA method has a slightly lower (about 30–190kWh) total transmission loss (per hour) than SPUA method, which is because DBUA is designed by nature to assign the nearest substations to users, thus leading to the lowest total transmission loss. However, comparing to the significant improvement (up to 65% reduction) of maximum substation utilization over DBUA method (from Figure 5), such a small increase on transmission loss is completely reasonable.

To demonstrate the practicability of the proposed framework, we look into the user assignment results obtained by SPUA vs the current assignment from the data. Figure 7 visualizes three substations with particularly long distance coverage in the existing user assignment. The black dots are the substations, and the orange circles are the current covering regions. Due to the high transmission loss, SPUA method re-assigns users from orange to blue circles, which are nearer in proximity. Comparing to Figure 2, Figure 8 illustrates that SPUA balances the substation utilization across substations to circumvent the over- and under-supplied problems. For over-supplied substations, SPUA either merges some of them, or expands their coverage to achieve higher utilization. For under-supplied substations, SPUA reduces the covering range to decrease the substation utilization.

5. CONCLUSION

In this paper, we study the problem of how to judiciously assign each power user to a substation, such that the maximum substation utilization is minimized. We develop a data-driven scalable power user assignment (SPUA) framework that takes heterogeneous power grid data as inputs, including temporal power consumption data and spatial power user/substation distribution data, and performs optimal user assignment. The observation from evaluations motivates us to further investigate various power grid planning problems, including the power plant and substation deployment, as well as roll-out strategies of substation-user assignment.

6. ACKNOWLEDGMENTS

This work is partly supported by the CAEIT Innovation Funds under Grant 14500601, 14500701 and 16502901. Haiyong Xie is also supported by the New Century Excellent Talents Program. Yanhua Li is partly supported by a gift funding from Pitney Bowes Inc.

7. REFERENCES

[1] C. E. Commission. The value of distribution automation. Technical report, Navigant Consulting, Inc., 2009.

[2] J. K. Lenstra, D. B. Shmoys, and É. Tardos. Approximation algorithms for scheduling unrelated parallel machines. *Mathematical programming*, 46(1-3):259–271, 1990.

[3] S. M. Mazhari and H. Monsef. Dynamic sub-transmission substation expansion planning using learning automata. *Electric Power Systems Research*, 96:255–266, 2013.