Grid-Aware Placement of Datacenters and Wind Farms

Abstract*—Due to the growing concern for huge energy consumption, Internet serivce providers tend to build data centers with renewable energy support to reduce cost and the carbon footprint. Prior work about "green" data center placement issues didn't consider the potential impact on the elecitricy grid by penetrating large loads and intermittent power generations from renewable resources. In this paper, we propose a holistic framework which incorporates together the cost of data centers, renewable energy power plants and the utility grid. By considering various factors such as land, building, hardware costs as well as the losses on the grid network, we try to solve an optimization problem to minimize the overall cost for both service providers and utility grid operators. XW: As discussed, we are assuming that the IT service provider want to build data centers and also renewable power plants to support these datacenters. Our results show that losses of the regional grid would have remarkable effect on the decision of selecting locations for data centers and renewable energy power plants. Furthermore, co-location choices for distribution load generation sources and loads as data centers are not always better, even comparing only the losses on the grid, which is quite not intuitive. Our work gives a way to seek for optimal locations in the capacity planning stage and also shows the importance for cloud service companies and grid operators to collaborate for reduction of overall cost.

I. Introduction

As reported recently, the energy consumption of data centers keeps growing while more and more enterprises and organizations are building their own data centers [1], [2]. The increasing speed of the data center energy consumption is approximately 10-12% per year recently [3]. The carbon emission and environment pollution issues attract insights of seeking for clean energy resources. Many IT companies such as Google [4], Apple [5], and McGrawHill [6] are trying to build their data centers together with renewable energy power plants.

Usually, generation of sustainable energy like solar and wind will be closely related to the weather condition at certain locations. As far as we know, there might be different choices for building the data center and the power plants. Both on-site and off-site generation are possible approaches [7]. By grid-centric approaches, the renewable energy will be generated at locations with sufficient renewable sources and pumped into the grid. On the other hand, co-location and self-generation approaches sit the data center and the power plant at the same location to facilitate management and avoid long-distance losses. Since it seems no approach is perfect, we argue that companies essentially expect benefits from the investments by building up data centers and green power plants. However, since the renewable energy generation is mostly intermittent and sometimes might bring great penetration current into the electricity grid, the capacity should be carefully planned and will be limited from the perspective of grid operators. For example, some tiny failures may make an area of grid system completely out of power [8].

Since data centers are becoming quite large loads for the electricity grid, they are supposed to have a significant impact on the operation of power grid [9]. Large data center loads might increase the grid load and also lead to significant load variability of the electricity system. Especially, the emergence of the renewable energy sources with intermittent nature brings fluctuations onto the electricity networks. Furthermore, grid losses are on of the largest expenses for the power system operators [10]. EIA (U.S. Energy Information Administration) [11] has estimated that the electricity transmission and distribution losses all over the U.S. is about 6% of the electricity that is transmitted and distributed each year (averaged from 1990 to 2012). Thus, reduction of these losses for the grid can greatly affect the total operational costs.

In this case, companies who want to build data centers with either on-site or off-site renewable energy plants have to get the permission from the grid operators first in order to reduce the unexpected influence to the grid operation. Since the grid losses will also finally turn into expenses for end-users, service providers may want to collaborate with grid operators to minimize the overall cost when planning locations and capacity for the data centers. In the current literature, data center placement issues have been mentioned in some prior work [12]–[15], and there are also some research focused on the capacity planning of green data centers [16], [17]. Nevertheless, these work hasn't considered the impact of data center placement on the grid itself, which might also lead to comparable costs as other costs for data centers.

In this paper, we attempt to set up a different point of view, by combining the consideration for cloud service providers and energy companies together and aiming at the minimization of the overall cost for both. First, we investigate the impact of data center placement and its importance to the grid by studying a region of grid network. We pay special attention to the data center size, data center locations and the variation of renewable power generations. Second, we formulate the optimization framework, which incorporates the costs of data centers, renewable power plants and power grid into one objective. We also try to solve it under necessary constraints by using several different approaches. Then we conduct a case study in the New England area of the United States, by sitting and provisioning data centers and power plants at different locations, with the purpose of minimizing the overall cost. Results show that grid losses can have remarkable impact on the decision of selecting best locations for green data centers. Furthermore, the co-location choices by sitting the data center and green plants together don't show advantages despite of the reduced line cost and distribution cost, which is not that intuitive as prior work thought.

Contributions of this paper. The main contributions of this paper includes: (i) it quantifies the potential impact of data center placement on the bus network of power grid, and (ii) it proposes a framework for smart placement of both data centers and renewable energy plants in the power grid network. (iii) it formulates an optimization problem and gives a solution approach to find out good choices for sitting and provisioning data centers when considering grid costs.

To the extent of our knowledge, there is no previous work considering the jointly placement issues of data centers and green power plants while caring about the grid operational costs together. The remainder of the paper is organized as follows. Section II first quantifies the potential of data centers together with wind farms by placing them into different buses of the grid network system. In Section III, we describe the optimization framework in detail, showing the integration of various parameters of the entire problem. Section IV evaluates the costs and illustrates breakdown of the total cost by different kinds of strategies. In Section V, we present some prior work related to this paper. Finally, the conclusion is given in Section VI.

II. Metric choice: Data center placements in renewable power grid

Planning optimal data center location, requires comprehensive metric definition. Most of the studies carried out so far neglect the impact of renewable powered data centers on the transmission grid. When the penetration of renewable powered data centers was small, the impact of these on the transmission grid was insignificant. However, with several large data center companies opting for renewable energy source it becomes imperative to study their impact on the transmission grid.

A. Impact of data center on renewable power grid

We will show that choosing optimal data center location within the renewable power grid could be beneficial to both the grid operators and the data center owners. Specifically data centers located at strategic places in the grid could help minimize i) overloading of transmission lines; ii) grid voltage variations outside the acceptable range iii) system losses. We will demonstrate this by considering a real world system, i.e., the New England Independent System Operator (ISO) spanning most of North Eastern region of United States and some parts of Canada.

1) Overloading of transmission lines: The transmission lines (referred as branch hereafter in the paper) are used to transport power from the large generators to the load. The power carrying capacity is limited to protect the line from over heating, mainly due to the line resistive losses i.e., I^2R, where I is the current flowing through the branch and R is the resistance of the branch.

A transmission line has typically two ratings: short term and long term capacity rating. During certain wind and system load (including data center load) conditions, some of the transmission lines could get overloaded. If this happens during the normal operation of the electric grid then one of the following will be done: i) if an electronic power flow controller is available then it is used to control the power flow through the overloaded line; or ii) in extreme situations, the overloaded line is disconnected which may result in power supply interruption to the loads.

If major transmission lines are getting overloaded quiet often annually then new lines are planned and built. This solution is very expensive and takes a long time (typically 7 to 10 years). The need for such expensive grid retrofits may be minimized by planning the location of the data center.

2) Voltage variation in electric grid: The voltage magnitude varies in the electric grid and needs to be maintained within a narrow range (for example +/- 5% of nominal) so that there is no damaged caused to the sensitive electronic loads. However, sometimes/days the change in renewable power output could cause the voltage to vary beyond the acceptable limits. Such over/under voltage problems can be mitigated by appropriately locating the data center load.

3) System losses: Historically, the electric grid was designed to have large central generating stations that are located far away from the load centers. The power from these central sources would be transmitted to the data center over transmission lines. While designing such a grid, the generator location and the transmission line voltage level as well as the path would be optimized to minimize the line losses.However, today with renewable power being distributed the scenario has changed, the generation sources are distributed and they may be located near the load centers. In order to transfer power from these renewable source we still use the existing transmission lines that were planned and built about 50 years ago or earlier. This may result in higher line losses and sub-optimal power transmission between generation sources and loads. Since we cannot re-design the entire electric grid to minimize line losses, we need to leverage the flexibility we have in locating new loads i.e., data centers in our specific case.

B. Simulation study

In order to study the impact of data center location in a renewable power grid we consider the New England ISO transmission network. We choose this system due to the following reasons:

- Wind power expansion in New England: The system studies carried by the New England ISO state that in this region there is a potential of integrating up to 12 Giga-Watts of wind power. Given this enormous interest to install new wind farms, this region may have a great potential in future to accommodate data centers powered by wind farms.
- Transmission network upgrades: A study carried out by the New England ISO showed that they could potentially integrate wind resources to meet up to 24% of the region's total annual electric energy needs in 2020 if the system includes transmission upgrades. If these transmission upgrades can be limited then the development of new wind farms becomes more economical.
- Positive impacts of wind power in New England ISO: Introducing large amounts of low-marginal-cost wind generation tended to depress the spot price and reduce

the price differential for bulk power between day and night. Also the study results demonstrated that there was only a relatively small increase in the use of existing pumped-storage hydro power for large wind penetrations, largely because the flexible natural-gas-fired generation fleet provided most of the system balancing.

We will show in our study, that within the New England ISO, the wind power penetration can be increased much more cost effectively by strategically locating the data center. Intuitively, it may appear that co-locating data center loads and wind farms would solve the problem of building new transmission lines, limit the voltage variation and also minimize losses. However, our study shows that this is not indeed true and co-locating wind farm and data center may not be the optimal solution always.

1) Model of the New England renewable power data center: Before we carry out the case studies, we will describe the approximations we made and the models we used for each sub-system we considered in our study.

- New England transmission system model:
  For our study we considered a widely used [18] model of the New England transmission network. A single line diagram of this test system is shown in Figure 1. As shown in the figure, the model lumps all the generators, loads and transmission lines in the New England ISO region to 10 generators, 19 loads and 46 lines and transformers. The 10 generator buses are numbered from 30-39 in Figure 1. Specifically, bus 39 represents the aggregation of a large number of generators interconnected to rest of US/Canada.

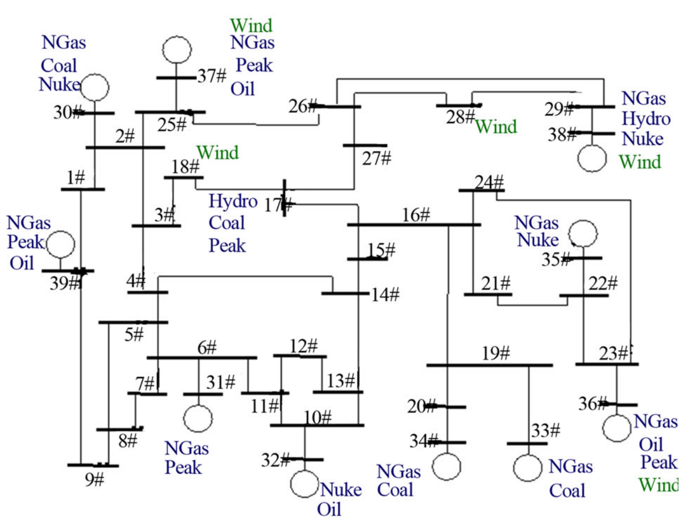

Fig. 1: New England 39 bus test system

- Data center model:
  Based on the geographical mapping we aggregate all the data centers in the New England ISO region into six data centers, each data center for one state. In order to estimate the size of an "aggregated" data center in a certain state, we sue the follow equation:

$$L_i = \frac{n_i * 9.8GW}{1278} \tag{1}$$

  where L_i is the aggregated load of the ith state, n_i is the number of datacenters reported in that state, 9.8GW is the upper bound of total electricity used by US datacenters in 2010, according to the report [2], and 1278 is the number of datacenters in US collected and reported in [19]. However, according to [2], for the summarized load of data centers, there is an increase of 56% from 2005-2010. Hence, we are assuming the increasing percentage from 2010-2014 is 56%*0.8=45%, and after adjustment we are using $L'_i = 1.45L_i$ as the data center load for the target grid system.
  The state wise aggregated data centers are mapped to different buses according to their geographical locations, as seen from Table I. Note that the load size given in the table represents the total load of data centers in the entire state.

TABLE I: Background data center load and location settings

DC No.	State	Number of DCs	Estimated size(MW)	Mapped Bus No.
DC1	Connecticut	12	133.43	6
DC2	Maine	3	33.36	29
DC3	Vermont	4	44.48	25
DC4	Rhode Island	3	33.36	20
DC5	New Hampshire	4	44.48	16
DC6	Massachusetts	27	300.21	4

- Wind farm model:
  The wind farms connected to bus 18, 28, 36, 37 and 38 (Figure 1) are lumped models of several wind farms within a geographical region. The locations and capacity settings of the five wind farms are presented in Table II. We assume that each farm can be represented by $'n'$ identical wind turbines, where n= total farm rated capacity/individual wind turbine rating. This approximation does not change any of our results because we are interested in studying the global impact of wind farm powered data centers on the electric grid. Also, since most of the wind turbines in this region are the GE 1.5MW machines, we consider the GE machine wind speed versus power characteristics [20]. The wind speed versus turbine output characteristics is commonly referred to as the power curve (Figure 2). From Figure 2 we can see that the cut-in wind speed, i.,e the wind speed at which the turbine starts producing power, is 5m/s and the cut-off wind speed is 25m/s, beyond this the turbine will be shut down for safety reasons. The wind turbine produces rated output (1.5MW) between wind speeds of 13-25m/s.

C. Case study

We have used the above described model and carried out the following case studies to assess the impact of data center location on the renewable power grid.

- Case 1: Existing New England system

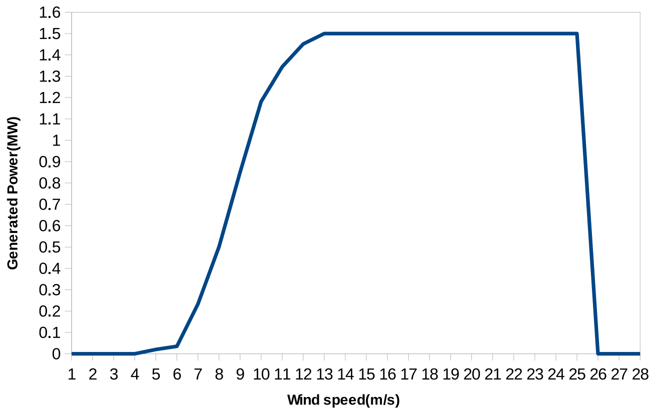

Fig. 2: Power curve of the GE 1.5MW wind turbine

TABLE II: Wind farm settings

WF No.	Bus No.	State	Capacity(MW)
WF1	18	?	100
WF2	28	?	90
WF3	36	Vermont	90
WF4	37	Maine	90
WF5	30	Massachusetts(?)	90

- Case 2: New England system with one additional data center (co-located with a wind farm)
- Case 3: New England system with one additional data center (located away from the wind farm)

For assessing the impact, we use the three metrics described earlier (section II-A) and calculate the system losses, power flows through the transmission lines and the bus voltages. The system losses, line flows and the bus voltages are calculated by solving the power flow or loadflow equations. The loadflow equations mathematically model the power balance (i.e net load +losses = total generation in the electric grid). Within an electric grid the power can be easily measured at the loads and at generators. Also, some generators have the capability to regulate the voltage at a bus at a constant preset reference value. The power flow equations are used to calculate the bus voltages (magnitude and angle), for a given network and a set of load and generation powers. The loadflow equation for a generic n bus network with k branches is given below.

$$P_i = \Sigma_{j=1}^{n}(|Y_{ij}||V_i||V_j|cos(\theta_{ij} + \delta_j - \delta_i) \quad (2)$$

$$Q_i = -\Sigma_{j=1}^{n}(|Y_{ij}||V_i||V_j|sin(\theta_{ij} + \delta_j - \delta_i) \quad (3)$$

where P_i and Q_i are real and reactive powers at the i^{th} bus; $|V_i|\angle\delta_i$ is the voltage magnitude and angle at the the i^{th} bus; $|Y_{ij}|\angle\theta_{ij}$ is the admittance of the branch between i^{th} and j^{th} bus. For a given power P_i and Q_i at the i^{th} load bus, the above powerflow equations are used to solve for the voltage magnitude and angle at the i^{th} bus. Since the above equations, i.e., real and reactive powers are non-linear function of voltage they are solved iteratively using Newton Raphson method. Once the bus voltages are calculated the line flows and system losses are computed.

D. Results and discussion

Here, we show the simulation results of the three cases and give some discussion about them. First, the total system losses are compared over the three cases under the condition of three wind speed settings, as shown in Figure 3. The background load of the electricity grid is set to normal (6254MW in total,summarized over all the buses), as described previously. Three categories of wind speed settings are: LOW (4-6m/s), MEDIUM (7-11m/s), and HIGH (>11m/s). The size of the additional data center in Case 2&3 is set to 200MW. Specifically, for Case 2 the data center is located at bus 18 (co-located with wind farm 1) and for Case 3 the data center is located at bus 10 (away from all wind farms).

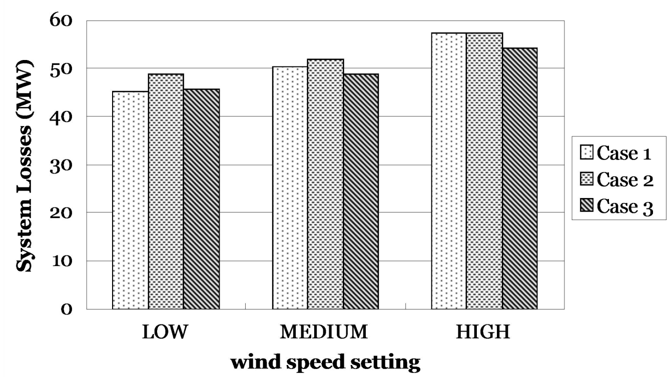

Fig. 3: Results of system losses

By comparing the values of system losses from Fig. 3, it can be observed that the co-location case (Case 2) could lead to more losses than Case 1 and Case 3. Take MEDIUM setting for example, the system losses of Case 3 is about 6% less than Case 2, which illustrates system loss can be reduced by placing the data center correctly and co-location is not necessarily the best choice.

Next,in order to investigate the impact of data center placement on line capacity violations and voltage violations, we change the scenario to peak system load (6885MW in total) and conduct the experiments again. Evaluation results are illustrated in Table III, which shows that there is one overloaded branch from bus 4 to bus 5 in Case 2. This highlights that different placement choice of the additional data center can make some particular lines overloaded. During our experiments, although there are only a few cases where lines are overloaded, annually the number of times and how long these over-loads could occur depends on the frequency and duration of occurrence of a particular wind speed and load condition. If it is too frequent or more persistent then the over-loads could be a serious problem and might require building new transmission lines. XW: need to add costs of building new line? I haven't found the reference yet.

Even though we have provided one set of results for a particular set of condition, we have simulated different wind speed and load conditions. We found that in general the system loss magnitude could change. However, in most cases we saw that co-location is not the optimal choice for minimum system loss. In Section III we will describe a method for determining

TABLE III: Results of overloaded lines

Case No.	# of overloaded lines	List of Overloaded Lines
1	0	None
2	1	bus 4 - bus 5
3	0	None

the data center location that will correspond to minimal system loss annually and will cover all the wind speed and system load conditions.

III. Framework for smart placement

The previous section highlights that placing data centers and wind farms has obvious impact on the grid trasmission network system. Thus, when a service provider is looking for best locations to estabish a network of green data centers, the grid operational cost should also be included as one of the most important factors into consideration. Another observation is that co-location of the green data center and the green energy plant seems not always be the best choice. In this case, we try to study how to seek the best locations for both the data center and the green energy plant, which are designed to be connected to the same grid network.

By grid-aware placement, we attempt to efficiently select a set of locations for one or more data centers to support a given amount of computational power, as well as one or more green power plants (e.g. solar, wind or others) to provide a given power capacity. The main goal is to minimize the overall cost for data centers, green power plants and also the grid network operation. The next subsections defines some important parameters in the selection procedure. Then, the cost model and the entire optimization problem will be formulated and described.

A. Parameters for the placement framework

Table IV gives all of the parameters defined in our framework. These parameters relate to costs, revenue and losses over different geographical locations. We generally classify them into three categories: data-center-related, renewable-related and grid-related.

1) Data-center-related costs: Given a target capacity of how large a data center is planned to build, we can calculate the capital cost first. The capital cost of a data center at location l, denoted as $DC\_CAPcost(l)$, can be broken down into land cost ($DC\_landCost(l)$), building cost ($DC\_buildingCost(l)$) and hardware cost($DC\_hwCost(l)$). Land cost can be computed by land price at that location and the accommodated data center capacity. Building cost depends on the maximum capacity of the data center, too. Hardware cost contains both server cost and switch cost, which can be computed given the total capacity of the data center. Besides, hardware cost also includes costs for building lines connecting the data center to the Internet backbone and the transmission grid, denoted as $costLineNet(l)$ and $costLineGrid(l)$ here respectively. We assume them independent of the data center size, but dependent on the location l of the data center.

Besides capital costs, operating and managing a data center also incurs operational cost, including costs for using the network bandwidth and the electricity from the power grid. We denote them as $DC\_netCost(l)$ and $DC\_energyCost(l)$ respectively, which both depend on the location l of the data center. Furthermore, the energy cost for the data center is also related to the varying demands of the data center workload over different time period, denoted as $demand(l, t)$ hereafter.

2) Renewable-related costs: The costs for building and running a renewable power plant also include capital costs and operation costs. The capital cost of a type r plant at location l, denoted as $RE\_CAPcost(l, r)$, can be broken down into land cost ($RE\_landCost(l, r)$), building cost ($RE\_buildCost(l, r)$) and line cost ($costLineNet(l)$) for connecting to the transmission grid. Specially, if we are also building a data center in the same location, the connection line to the grid could be shared by the plant, and in this case the line cost could be saved. Similarly, the land cost for renewable power plant depends on the needed area and the land price at that location. The building cost mainly depends on the desired capacity of the plant.

Regarding operational costs, we assume that the human and labor costs for maintenance and operation are same over different locations. Thus, by operating the renewable power plant, we only consider the possible revenue it could bring by generating electricity power and transmitting power to the grid. Here, we regard the revenue as a negative cost for the power plant, denoted as $RE\_OPrev(l, r)$, which is closely related to the power generation efficiency at the location and the energy selling price there.

3) Grid-related costs: When connecting the data centers and the renewable energy plants to the utility grid, we are adding both generation and consumption components into the network. This will change the power flow of the whole network, and thus the total losses of the transmission network will be different. Thus, the transmission loss ($transLoss(t)$) will be affected by the generating and consuming power amount of data centers and renewable plants, and also the buses they are connected to, as we showed in Section II.

On the other hand, during the transmission process the line capacity might be violated if the power flow exceeds the limit. We use $numLineVio(t)$ and $numVolVio(t)$ to denote the number of line capacity violations and voltage violations during power transmission. The grid operator will try to avoid such violations when operating the power grid system.

B. Optimization problem formulation

Using the parameters shown in Table IV, we can formulate the optimization problem as shown in Figure 4. The problem is set up from the perspective of both IT company and the grid operator, who want to collaborate for building up green datacenters, with the purpose of minimizing the summarized cost including data center cost, green power plant cost and the grid cost.

Denote $\mathcal{L}$ as the set of all candidate locations, $\mathcal{T}$ as the set of all time epochs, $\mathcal{R}$ as the set of all types of renewable energy. The input of the optimization problem is listed as follows: (1) the total computational capacity of all the data centers to set up, denoted as $CapacityDC$; (2) the parameters of each location in $\mathcal{L}$ during each time epoch in $\mathcal{T}$ such as

TABLE IV: Parameters for placement framework. Each location l is a possible location from the set $\mathcal{L}$, and each t denotes a time epoch during a time period T. Each r is a type of renewable energy from the set $\mathcal{R}$.

Symbol	Meaning	Unit
$dcCapacity$	desired power capacity for computing in DC	kW
$wfCapacity$	desired power production capacity of wind farm	kW
$pLand(l)$	land price at l	\$/m^2
$PUE(l,t)$	PUE at l during t	
$maxPUE(l)$	maximum PUE at l	
$dcArea$	land area needed per kW of DC capacity	m^2/kW
$cLinePow(l)$	cost to layout power line from l to the closest power plant	\$
$cLineNet(l)$	cost to layout optical fiber from l to closest network backbone	\$
$pBuildDC(c)$	price of building a datacenter with c power capacity	\$/kW
$serverPow$	server peak power demand	W
$switchPow$	switch peak power demand	W
$servsPerSwitch$	number of servers per switch	servs/switch
$pServer$	price of a server	\$/server
$pSwitch$	price of a network switch	\$/switch
$pNetBWServ$	cost of external network bandwidth per server	\$/serv-month
$pEnergy(l)$	grid electricity price at l	\$/kWh
$powDemand(t)$	avg computing power demand of DC during t	kW
$\beta(l,t)$	avg generation efficiency of wind energy at l during t	%
$wfArea$	land area needed per kW wind energy	\$/m^2
$pBuildWF$	price of building a wind power plant	\$/kW
$revEnergy(l)$	revenue for selling wind energy to grid at l	\$/kWh
$transLoss(t)$	avg transmission loss in grid during t	kW
$pGridLoss$	the price for transimission losses per kWh	\$/kWh

prices, PUE (Power Utilization Efficiency), demand, power generation efficiency and so on; (3) the minimum availability constraint for the data center network. (XW: to limit the number of data centers) The outputs of the problem is the lowest cost found and the corresponding locations for data centers and renewable power plant, as well as the capacity provisioned at each location for data centers or green power plants (if any).

Equation 1 in Fig.4 shows the optimization objective of our defined problem, i.e. $TotalCost$, where $DC(l)$ and $RE(l,r)$ are booleans indicating whether to place a data center or power plant of type r at location l. $DC\_Cost(l)$, $RE\_Cost(l,r)$ and $Grid\_Cost$ represent the cost for data centers, renewable plants and the power grid system respectively.

The overall cost should be optimized under the constraints, which are listed in Figure 5. Equation 21-23 show the constraints of the provisioned capacity for data centers. Equation 24 means that the provisioned capacity for renewable energy plants are determined by the total power demand from the data centers. This constraint is added indicating that the power generation and consumption added to the grid should be balanced from the perspective of the grid system. Furthermore, Equation 26 is a strict limitation for keeping the grid out of any violations at any given time t, since we assume the grid reliability is crucial and must be guaranteed.

C. Optimizing approaches

1) Semi Brute force: A time-consuming approach is to generate all of the possible combinations for data centers and renewable energy plants. However, it is not possible to generate all kinds of capacity provisioning amounts since it's not discrete. Thus, we generate combinations of locations first, and then evenly distribute the total capacity to all of the candidate locations selected. By testing these generated configurations, the approach returns the best one with the lowest total cost. This approach still could be very exhaustive, which needs extremely long time for execution.

2) Heuristic searching: XW: TO DO SOME WORK...

IV. Evaluation Experiment

Now in this section, we study the overall cost for placing data centers and renewable energy power plants into the same grid network. By considering the grid loss or not, we can get different results of placement decisions.

A. Input data

The target area here is New England in United States, and we select 56 locations as candidates inside this area, as shown in Figure 6. We obtained the Typical Meteorological Year (TMY) information for 56 locations from US Department of Energy (2), which includes a one-year dataset of hourly weather values for a location. We simplify the problem by only considering one type of renewable energy - wind. We computed the average wind power generation using $effRE(l,t)$ derived from the specifications for the 1.5MW Series wind turbine from General Electric Company [21], TMY wind speed, TMY air pressures, and conversion losses.

Besides, we collected various data of PUEs, data center construction costs, wind farm construction costs, land costs, transmission lines and network connection costs and grid energy costs in the same way as stated in [17]. Specifically, for grid costs, we use $priceLoss$ the same as the maximum electricity price in the whole area. The input data for the grid is derived from the New England grid system, which is shown as in Section II,including all of the settings for buses, branches and generators in it. Thus, the transimission loss, $transLoss(t)$, could be computed by simulating the power flow process for timp epoch t.

B. Building one data center with one wind farm

Here, we are trying to solve a simplest case of the defined problem by placing only one data center and only one wind farm onto the grid system. The added generation (wind power) and consumption (data center load) should be balanced for the grid to keep reliability to the best extent. In this case, we can brute force all of the combinations of locations for the data center and the wind farm. Figure 7 shows the results of the total cost by using five different strategies when seeking the best locations when we are building a 100MW data center and a wind farm which can supply green energy for it.

The five different strategies are explained as follows:

(1)**DC_WF_OPT.** This strategy tries to seek for the best location for the data center where its total cost could be

$$
\begin{aligned}
totalCost &= dcCost + wfCost + energyCost & (4)\\
dcCost &= dcCAPEX + dcOPEX & (5)\\
dcCAPEX &= dcLandCost + dcBuildCost + dcITCost & (6)\\
dcOPEX &= dcNetCost + dcEnergyCost & (7)\\
dcLandCost &= pLand(d) \cdot dcArea \cdot dcCapacity & (8)\\
dcBuildCost &= dcTotalPow \cdot pBuildDC(dcTotalPow) + cLinePow(d) + cLineNet(d) & (9)\\
dcTotalPow &= dcCapacity \cdot maxPUE(d) & (10)\\
dcITCost &= nServers \cdot pServer + nSwitches \cdot pSwitch & (11)\\
nServers &= dcCapacity/(serverPow + switchPow/serversPerSwitch) & (12)\\
nSwitches &= nServers/servsPerSwitch & (13)\\
dcNetCost &= nServers \cdot pNetBWServ & (14)\\
dcEnergyCost &= \sum_{t \in T} |t| \cdot powDemand(t) \cdot pEnergy(d) & (15)\\
wfCost &= wfCAPEX - wfRev & (16)\\
wfCAPEX &= wfLandCost + wfBuildCost & (17)\\
wfLandCost &= pLand(w) \cdot wfArea \cdot wfCapacity & (18)\\
wfBuildCost &= pBuildWF \cdot wfCapicity + cLinePow(w) & (19)\\
wfRev &= revEnergy(w) \cdot \sum_{t \in T} |t| \cdot \beta(w,t) \cdot wfCapicity & (20)\\
gridCost &= pGridLoss \cdot \sum_{t \in T} |t| \cdot transLoss(t) & (21)\\
& & (22)
\end{aligned}
$$

Fig. 4: Optimization framework.

$$
\begin{aligned}
\forall_{l \in \mathcal{L}}, capDC(l) \le DC(l) \cdot CapacityDC &\Rightarrow \text{capacity of data center at } l \text{ should be zero when } DC(l) \text{ is 0} & (23)\\
\sum_{l \in \mathcal{L}} DC(l) \cdot capDC(l) = CapacityDC &\Rightarrow \text{total capacity of built data centers should meet the requirement} & (24)\\
\forall_{t \in T}, demand(l,t) \le capDC(l) &\Rightarrow \text{power demand of the data center should not exceed its capacity} & (25)\\
\sum_{l \in \mathcal{L}, r \in \mathcal{R}} RE(l,r) \cdot capRE(l,r) \cdot avgEff(l,r) = \sum_{t \in T, l \in \mathcal{L}} DC(l) \cdot demand(l,t) \cdot PUE(l,t) &\Rightarrow \text{the generated green energy should be balanced with consumption} & (26)\\
\forall_{t \in T}, 0 \le effRE(r,l,t) < 1 &\Rightarrow \text{efficiency of power plant should be between } [0,1) & (27)\\
\forall_{t \in T}, numLineVio(t) = 0, numVolVio(t) = 0 &\Rightarrow \text{No violations in each time epoch} & (28)
\end{aligned}
$$

Fig. 5: Optimization constraints.

minimized, and the best location for the wind farm where its total cost could be minimized. XW: According to Divya, today there are standard methods used to determine the best location for wind farms. So are you suggesting us to focus on the placement of datacenters only? Grid costs are not considered here.

(2)**DC+G_WF+G.** Since only putting data center or wind farm into the grid will change the power flow, this strategy tries to add the transimission loss costs into consideration when seeking for best locations separately for the data center and the wind farm.

(3)**Jointly.** All of the combinations of locations are exploited by using this strategy to seek for the jointly placement choice for both the data center and the wind farm, considering the total cost as calculated by Equation 2.

(4)**Min_Loss.** Here, this strategy seeks for the locations for data center and wind farm which can lead to the minimal grid losses cost.

(5)**Co-location.** Assuming the the data center should be build together with an on-site wind farm, this strategy seeks for one location to place both the data center and the wind farm towards miniming the total cost.

From the figure, we can see that by considering grid costs,

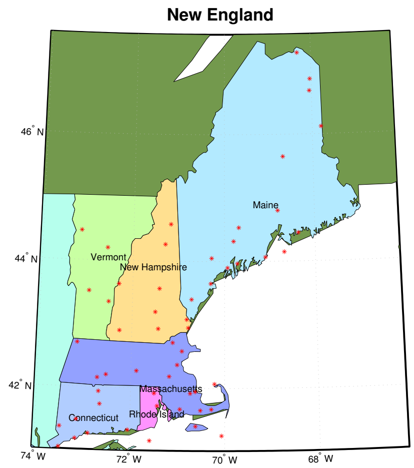

Fig. 6: Candidate locations in New England

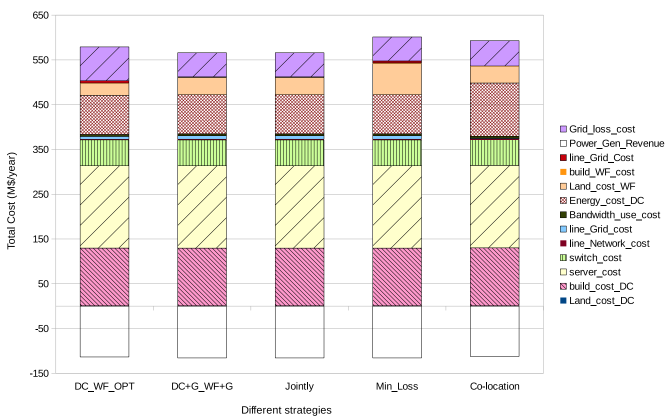

Fig. 7: Costs of building one data center (100MW) with one wind farm

the total cost will be further saved compared to best choices for the data center and the wind farm separately. Also, Min_Loss can achieve minimal losses of all possible choices, but the total cost is large mainly because it selects an expensive place for purchasing land for the wind farm. The best choice of co-location options is also nearly 7% higher than the $Jointly$ choice, and it's easy to understand since the best location for data center is not necessarily the best for wind farm and vice versa.

We calculate all of the combinations for wind farm and data center locations and use the average total cost of these combinations (which is \$667.1M per year) as the baseline for comparison. Then, the locations found and the corresponding cost savings of the five strategies are listed in Table V.

C. Multiple data centers and wind farms

XW: NEED MORE MORE WORK....

TABLE V: Detailed results of cost savings by different strategies.

Strategy	Data center location	Wind farm location	Total cost (M\$/year)	Cost saving (%)
DC_WF_OPT	Burlington,NH	Mount Washington, NH	465.6	30.2
DC+G_WF+G	Springfield Hartnes, VT	Nash Island, CO	450.3	32.5
Jointly	Springfield Hartnes, VT	Nash Island, CO	450.3	32.5
Min_Loss	Springfield Hartnes, VT	Marthas Vineyard, RI	485.6	27.2
Co-location	Nash Island, CO	Nash Island, CO	480.7	27.9

V. Related Work

This section reviews relevant work to this paper in the recent literature, which are classfied into two categories.

A. Effect of data centers to the grid

In the smart grid era, data centers began to show the advantages for demand response and facilitate ancillary services due to its great and controllable flexibility. Researchers studied the effect of datacenter demand response on power consumption reduction [22], [23]. They found that 25% of the demand savings can be done with minimal or no impact on datacenter performance. Also, 10% of the load can be shed with short response time with no operational impact. They did not consider dynamic load migration of the workload which can result in further reduction in power demand.

Mohsenian *et al.* in [24] proposed a request distribution policy among datacenters to ensure power load balancing. They tried to minimize the maximum power on any transmission line by distributing the computing requests to suitable datacenter. Their work assumes that a fairly large number of datacenters (e.g. 6) are connected to the same power distribution network. In practice, it is very rare for some company to build several datacenters connected to the same power distribution network. Aikema *et al.* in [25] studied the energy cost savings that can be achieved when datcenter participates in ancillary services. Their simulation shows that 12% cost savings can be done at the cost of 2% performance loss (i.e. increased latency).

Recently, Wierman *et al.* [26] surveyed the opportunies and challenges for data centers to ease the incorporation of renewable energy source into the grid and shaving the peak load. Further, Liu *et al.* in [27] focused on the impact of datacenter demand response on grid. They concluded that datacenter demand response can reduce the storage requirement of a grid with renewable energy source. The other key finding of their work is voltage violation frequency is lower when datacenter is placed on the same power bus with the PV solar source.

Different from these work, we quantify the data center impact on the grid by focusing on the losses brought because of penetrating additional load and generations to the grid network in a regional area. By incorporating such effects into a holistic framework, we convert such losses to grid operational costs and regard it as part of the total cost when building and planning the capacity of data centers and also the renewable energy power plants.

B. Datacenter placement and capacity planning

Some prior work has discussed about the placement issues of data centers. Alger [13] explained how to choose an optimal location for the data center several years ago, by considering hazars, accessibility and scalability factors. Stansberr [28] ranked some cities by estimating the annual operation costs of the data center. Oley [14] considered looking for a proper location for the data center establishment only by investigating the power rates of different states. Goiri *et al.* [12] focused on intelligently finding the best places for building multiple data centers to form a network for interactive Internet services. This work is to some extent close to ours, but they didn't consider the provisioning issues of renewable energy plants and the relevant costs.

Larumbe *et al.* [15] presented a mathematical problem aiming at solving the location and routing of cloud service components. Gao *et al.* [29] studied how to sit data centers near existing wind farms, and distributing load using a greedy online algorithm. Berral *et al.* [17] considered to select sites for data centers and on-site power plants aiming at follow-the renewable cloud services. Unlike our work, they didn't put insights the possible impact of data centers and distributed energy generations on the utility grid system.

Different from these work, we quantify and incorporate the impact of the site selection on the grid operation, and regard the summarized cost as the objective, which shows the importance of collaboring service providers and grid operators together to do the site and capacity planning.

VI. Conclusion

In this paper, we studied the problem of smartly placing green data centers in proper locations with considering the impact on electricity grid. Since connecting data centers and renewable energy plants to the utility grid will change the power flow of the whole grid network, and thus the grid losses should be incorporated when trying to minimize the total cost. In this context, we proposed an optimization framework with the purpose of optimizing the total cost including data center, renewable energy plant and also the grid cost. The problem is formulated with several necessary constraints according to the reliablity issues for the grid. By solving the problem, we tried to seek for best locations for sitting and provisioning the capacity of both data centers and renewable energy plants. Our optimization results show that considering the grid cost will have different impact on the placement choice for sitting data centers and renewable power plants. This means grid-aware placement stategies could further help saving the overall cost from the perspective of both IT companies and grid operators.

References

[1] R. Urgaonkar, B. Urgaonkar, M. J. Neely, and A. Sivasubramaniam, "Optimal power cost management using stored energy in data centers," in *SIGMETRICS*, 2011. 1

[2] J. Koomey, *Growth In Data Center Electricity Use 2005 To 2010*. Oakland, CA: Analytics Press, 2011. 1, 3

[3] G. Ghatikar, "Demand response opportunities and enabling technologies for data centers: Findings from field studies," 2014. 1

[4] Google, "Renewable energy Data Centers Google," 2014, http://www.google.com/about/datacenters/renewable/. 1

[5] Apple, "Apple and the Environment," 2013, http://www.apple.com/environment/renewable-energy/. 1

[6] R. Miller, "Data Centers Scale Up Their Solar Power," 2012, http://www.datacenterknowledge.com/archives/2012/05/14/data-centers-scale-up-their-solar-power/. 1

[7] I. Goiri, W. Katsak, K. Le, T. D. Nguyen, and R. Bianchini, "Parasol and GreenSwitch: Managing Datacenters Powered by Renewable Energy," in *ASPLOS*, 2013. 1

[8] Julfikar Ali Manik and Nida Najar, "Electricity Returns to Bangladesh After Power Failure," 2014, http://www.nytimes.com/2014/11/02/world/asia/power-grid-failure-puts-bangladesh-in-the-dark.html. 1

[9] H. Wang, J. Huang, X. Lin, and H. Mohsenian-Rad, "Exploring smart grid and data center interactions for electric power load balancing," *SIGMETRICS Perform. Eval. Rev.*, vol. 41, no. 3, pp. 89–94, Jan. 2014. [Online]. Available: http://doi.acm.org/10.1145/2567529.2567556 1

[10] R. de Groot, J. Morren, and J. Slootweg, "Investigation of grid loss reduction under closed-ring operation of mv distribution grids," in *PES General Meeting— Conference & Exposition, 2014 IEEE*. IEEE, 2014, pp. 1–5. 1

[11] U.S. Energy Information Administration, "How much electricity is lost in transmission and distribution in the United States?" 2014, http://www.eia.gov/tools/faqs/faq.cfm?id=105&t=3. 1

[12] I. Goiri, K. Le, J. Guitart, J. Torres, and R. Bianchini, "Intelligent Placement of Datacenters for Internet Services," in *ICDCS*, 2011. 1, 9

[13] D. Alger, "Choosing an Optimal Location for Your Data Center," in *InformIT*, 2006. 1, 9

[14] B. Oley, "Power & Approaching the Best Place to Build a Data Center," pp. 32–33, 2009. 1, 9

[15] F. Larumbe and B. Sansò, "Optimal Location of Data Centers and Software Components in Cloud Computing Network Design," in *CCGrid*, 2012. 1, 9

[16] K. Le *et al.*, "Capping the Brown Energy Consumption of Internet Services at Low Cost," in *IGCC*, 2010. 1

[17] J. L. Berral, I. Goiri, T. D. Nguyen, R. Gavalda, J. Torres, and R. Bianchini, "Building green cloud services at low cost," in *Distributed Computing Systems (ICDCS), 2014 IEEE 34th International Conference on*. IEEE, 2014, pp. 449–460. 1, 6, 9

[18] G. Bills *et al.*, "On-line stability analysis study," *Report for the Edison Electric Institute RP90-1*, 1970. 3

[19] Data Center Research, "Data Center Map," 2014, http://www.datacentermap.com/datacenters.html. 3

[20] Y. Lei, A. Mullane, G. Lightbody, and R. Yacamini, "Modeling of the wind turbine with a doubly fed induction generator for grid integration studies," *Energy Conversion, IEEE Transactions on*, vol. 21, no. 1, pp. 257–264, 2006. 3

[21] General Electric Company, "1.5 MW Wind Turbine Technical Specifications," 2013, http://site.ge-energy.com/prod_serv/products/wind_turbines/en/15mw/specs.htm. 6

[22] V. Ganti and G. Ghatikar, "Smart Grid as a Driver for Energy-Intensive Industries: A Data Center Case Study," 2012, lBNL. 8

[23] G. Ghatikar, V. Ganti, N. Matson, and M. A. Piette, "Demand Response Opportunities and Enabling Technologies for Data Centers: Findings from Field Studies," 2012, lBNL. 8

[24] A.-H. Mohsenian-Rad and A. Leon-Garcia, "Coordination of Cloud Computing and Smart Power Grids," in *2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2010, pp. 368–372. 8

[25] D. Aikema, R. Simmonds, and H. Zareipour, "Data centres in the ancillary services market," in *IGCC*, 2012, pp. 1–10. 8

[26] I. L. Adam Wierman, Zhenhua Liu and H. Mohsenian-Rad, "Opportunities and Challenges for Data Center Demand Response," in *Proceedings of IEEE IGCC, 2014.*, 2014. [Online]. Available: http://users.cms.caltech.edu/~adamw/papers/dcdrsurvey.pdf 8

[27] Z. Liu, I. Liu, S. Low, and A. Wierman, "Pricing data center demand response," in *Proc. ACM Sigmetrics*, 2014. 8

[28] M. Stansberr, “Data Center Locations Ranked by Operating Cost,” in *SearchDatacenter.com*, 2006. 9

[29] Y. Gao, Z. Zeng, X. Liu, and P. Kumar, “The Answer Is Blowing in the Wind: Analysis of Powering Internet Data Centers with Wind Energy,” in *INFOCOM*, 2013. 9