In the midst of these difficulties faced by utilities, growing fuel costs, environmental awareness and government directives have increased the push to deploy Electric Vehicles (EVs). These EVs also can provide substantial amounts of local storage - a single EV can store and consume the equivalent of 3 to 5 homes straining the grid [10]. While EVs have environmental benefits [11], they will have substantial impacts on the economics [14], design, operation, and stability of the grid [20]. At the same time distributed storage, either in the form of EVs or batteries have been suggested for utility and consumer cost reduction [33, 36, 32].

Fortunately, the introduction of new communication [30] and control [12] infrastructure into the electricity grid is expected to allow increased prosumer participation in the smart grid. This participation will in fact be necessary to reduce costs. However, the infeasibility of continuous human intervention and consumption control has led to the model of autonomous software agents, representing the consumers, that intelligently optimize and schedule energy usage [33, 24]. Our focus in this paper is on the *design of negotiation protocols* between electricity distribution utilities and smart consumption agents that exploit this advanced communication and control infrastructure to allow coordination of consumption and help maintain the stability of the grid. Our work lies at the intersection of game theory [25], mechanism design [26] and the design of distributed algorithms [23].

1.1 Electricity Congestion Control Protocols

While distributed algorithms can serve a number of purposes in the smart grid, ranging from frequency control and maintaining voltage stability to providing various ancillary services [21], we focus on the problem of congestion control as an illustrative example. However, the protocols we suggest should be applicable in many other scenarios.

As mentioned in the introduction, shifts in demand and supply continue to strain a grid already on the brink. Thus, in order to limit peaks, minimize losses and reduce costs, utilities need to exploit the patience and flexibility of consumers and any available local storage. For example, when EVs are not in use, they can play the role of a battery allowing users to time shift their electricity purchases. When consumers have access to storage, they can potentially store energy during off-peak hours and use this energy to satisfy demand during the peak. Peak limitation through storage adoption may be more acceptable to consumers than consumption limitation since it does not require behaviour modification. However, increased storage adoption can lead to 'herding' effects where co-incident charging of batteries creates new and larger peaks, increasing costs [8] and losses. Thus, utilities need to encourage users to not only shift but also coordinate consumption, through what are often called congestion control mechanisms.

Congestion control schemes in the power systems literature (often based on successful protocols developed for internet congestion control [19]) that discourage consumption when the grid is loaded, fall into four basic categories: day ahead or time of use (TOU) pricing, dynamic pricing, signaling and back-off strategies and tâtonnement (or negotiation).

While many of these methods have been proposed in the literature they all come with their own limitations. TOU pricing (particularly in situations where users have significant storage, patience or flexibility) leads to the herding problem described above, where large amounts of consumption are shifted to low price regions creating new peaks [8]. Dynamic pricing can lead to customer confusion, and "consumers generally shy away from markets when products are complicated, supply is uncertain, prices are volatile, and information is lacking" [9]. Signaling based approaches, where consumers respond directly to signals from the utility, require that consumers are willing to give up control of their consumption which may be unlikely, or require that users reveal their preferences, which in turn leads to privacy and security concerns [2]. Randomized back-off type approaches where users back-off on consumption when the grid is loaded (motivated by CSMA protocols developed for wireless communication) require trust that the agent will respond as expected to signals. This is not incentive compatible as the response strategies are unverifiable and it is in the agents' interest to choose small back-off windows which allow them to consume electricity with minimum inconvenience.

Tâtonnement [34], where consumption agents and the utility exchange prices and consumption profiles until convergence, is then perhaps the most promising protocol for user consumption scheduling and pre-commitment. While they were not explicitly categorized as such, tâtonnement forms the basis of much work in this area including [24, 27, 33, 22, 35, 17]. Unfortunately, the strategies proposed in the literature either do not converge with parallel or require synchronous serial updates (where no two agents can update consumption at the same time) to ensure convergence. This implies substantial communication overhead and a number of iterations linear in the number of agents, as explicitly noted in [24, 27], which rapidly becomes infeasible as the number of users in the grid increases.

Since typical distribution grids could have millions of users and tens of millions of devices, *developing protocols for multi-agent consumption coordination that overcome these synchronization, convergence, and incentive compatibility issues is the primary focus of our work.* In particular we develop constrained tâtonnement protocols for fast and incentive compatible distributed demand management in smart grids

1.2 Our Contributions

In this paper, we consider a game theoretic resource allocation setting with a central distribution company that wishes to allocate a limited quantity of an expensive electricity resource to a set of autonomous consumption agents, at minimum cost. Typically, because the cost of generation is quadratic in the quantity generated and because of peak and congestion effects, the distribution company would prefer balanced consumption across the slots, subject to physical capacity constraints and the requirements of the users. The allocation to each agent is decided in an iterative manner. In each iteration, the center signals information to the agents (such as prices for consumption in each time slot) and the agents respond with new consumption levels. The center then updates its signals and the process is iterated until convergence. These kinds of protocols have been widely studied in the game theoretic settings, where they are called tâtonnement protocols [34] and more recently in the power systems literature [19, 24, 16, 17, 27, 33].

In this paper, we show how a protocol that adds verifiable constraints on the user consumption changes during the tâtonnement process, can be used to non-trivially improve the iterative strategies proposed earlier for demand response [24, 27], EV charging [19] and distributed storage management [17, 33]. We show that when the proposed protocols are used, user selfish best response strategies converge within a poly-log (in an approximation parameter that depends on the desired accuracy) number of rounds *independent of the number of agents*. In this direction, our contributions are to develop the following.

- **Tâtonnement game model:** We describe the problem of peak minimizing or minimum cost electricity resource allocation in smart grids as a tâtonnement game between a single seller and multiple buyers. We explain how solutions presented in the literature, such as [24, 27] require a linear number of sequential updates to converge and others, such as [3, 19], may not be incentive compatible.
- **Fast, incentive compatible distributed protocols:** We map the energy demand management problem to a server allocation problem studied in [4] and develop a new protocol where, in addition to signals and prices, *centrally verifiable constraints* are also communicated to the users. We demonstrate with theory and through extensive simulations, that this leads to incentive compatible, distributed protocols that converge in a *constant* number of iterations, *independent* of the number of agents which represents a substantial speed-up for the situations studied in [24, 27].
- **Distributed demand management protocols incorporating capacity constraints and micro storage:** We show how *constrained* tâtonnement is useful in different resource allocation problems that are practically relevant in electricity grids, including load scheduling when users have local storage (the problem studied in [33]) or the distribution utility has capacity constraints in different time slots. Here, we demonstrate that these extensions require a *rethinking of how prices are to be set when new constraints need to be satisfied*.

A highlight of the protocol is its generality as it can be used on top of the distributed optimization techniques suggested in the literature and at the same time achieves faster convergence and ensures incentive compatibility.

The rest of this paper is organized as follows. In Section 2, we describe the problems facing energy utility companies in more detail, review some of the recent literature on agent based models for electricity consumption scheduling and their limitations with respect to convergence and incentive compatibility. In Section 3, we explain how a parallel to theoretical work on a server load balancing can be applied to the problem of electricity resource allocation to design new distributed protocols for energy consumption scheduling. We also show how extending these techniques to the problem of load balancing with capacity constraints in each time slot (a problem of importance in electricity grids with day ahead markets) requires a novel structure of the prices announced to the users. Finally, we show how these protocols also can be used for demand management when users have access to local storage - either from batteries or EVs - allowing fast convergence.

2. DISTRIBUTED STORAGE AND DEMAND SIDE MANAGEMENT

The modern electricity grid is expected to provide a reliable supply to consumers (the US grid operates at an average of three nines 99.9% reliability) at any cost. "In order to supply demand that varies daily and seasonally, and given that demand is largely uncontrollable and interruptions very costly, installed generation capacity must be able to meet maximum (peak) demand" [31]. As a result, "the average utilisation of the generation capacity is below 55% and the lowest marginal cost plant would operate at about 85% load factor, while a plant with the highest fuel cost would operate only a few hours per year". Reducing this peak to average ratio is one of the primary goals of energy utilities, and can be achieved through customer demand response programs.

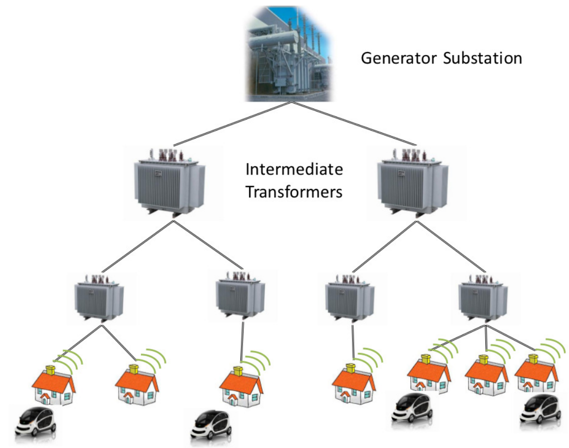

Figure 1: Smart Grid Structure

Demand response in electricity markets can be defined as the changes in electric consumption by users from their normal consumption patterns in response to signals (e.g. pricing) from the distribution authority [1]. In response to these signals, users can defer or reduce loads in coordination with other consumption agents to reduce the overall load on the grid. However, this response could lead to temporary user inconvenience and this inconvenience should be limited as far as possible to encourage participation in demand response programs.

One avenue to shift consumption or shave peaks with minimal user inconvenience is through the use of energy storage. Local storage would allow users to coordinate charging and consumption and then use energy from the storage when the grid is over loaded. While storage is expensive, distributed storage is already prevalent. For example, due to a chronic shortage of electricity of more than 10% [13], outages are common throughout India, causing customers to purchase storage and local generation. At the same time EVs are expected to become more common throughout the developed world [14]. When these EVs are parked, they can be plugged in to absorb energy and store it for later consumption, acting as large local storage.

2.1 Electricity distribution system model

Traditionally, most electricity distribution grids are radial in nature [5] and hence can be represented in the form of a tree with the nodes representing the buses and the undirected edges representing the electrical line connections, as in Figure 1. This tree can further be modeled as a rooted tree with the root node representing the generator sub-station and the leaf nodes representing the loads (the customers buying electricity). This rooted tree structure is used as a model of the distribution grid in this paper.

The loads are parametrized by a *power requirement* which is the total amount of power they want to consume across some time duration called a *consumption period*, such as a day. The consumption period is further divided into smaller duration called *slots*, corresponding to hours, minutes or even seconds. With each load, a *strategy vector* is associated which is a vector of powers consumed by the load across all slots in a consumption period. Each load is free to consume any amount of power in the individual slots as long as the total power consumed across all slots in a consumption period equals the specified requirement.

User preference can also captured by further allowing the loads to be active (consume power) only in some subset of slots during any consumption period. That is, for every load, we can further associate a binary 'indicator' vector indicating which subset of slots in any consumption period the load can consume power. An example where such a model can be applied is in the case of EV (Electric Vehicle) charging by individual customers at their homes. Some people may have to use their cars for work and can charge only during off-work hours. However, such users will be in-different as to 'how' the car gets charged during the off-work hours as long as it is charged by beginning of the work hour next day. In this example, the consumption period could be one day and the slots' duration can refer to one hour. The power requirement of the EV is specified by its storage capacity. Hence any scheduler has complete freedom in choosing an appropriate strategy vector so long as the the total power requirement is met and the user's preference (indicator) is respected. Note that in general, as we will discuss later, this model is also general enough to model arbitrary energy consumption when users have access to local storage and is not limited to the EV charging described here.

2.2 Existing Demand Response Protocols and their Limitations

As described in Section 1.1, demand response protocols fall into four basic categories: day ahead or time of use (TOU) pricing, dynamic pricing, signaling, and tâtonnement. In TOU and dynamic pricing [1], consumers are charged based on the state of the grid, either predicted day ahead (TOU pricing) or real time (dynamic pricing). This requires a good estimate of user requirements and preferences at the central utility, which may be unrealistic, and have been seen to lead to large swings in prices and consumption, which discourage user participation and lead to grid instability.

The third approach to this problem is motivated by parallels to the development of the internet [19] and in particular internet protocols such as TCP [3], which assume agents respond to signals from the central authority (essentially by fiat) and will curtail load as required. Deciding on which users should or could reduce consumption requires them to reveal their preferences and consumption patterns which raises privacy concerns [2]. Back-off strategies borrowed from the wireless domain, where users delay consumption when the grid is overloaded, have also similarly been suggested for peak power reduction. However, these are not incentive compatible since it is often not in the users (selfish) interest to delay consumption and since the back-off strategies are not verifiable by the utility. That is, selfish users may be incentivized to choose small back-off windows resulting in a breakdown of the protocol. In this paper we develop incentive compatible protocols by introducing centrally verifiable constraints that actually speed up convergence.

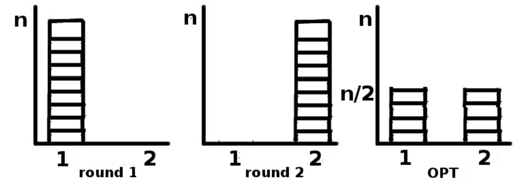

Figure 2: Example demonstrating that parallel updates may not converge and that sequential updates could require a linear number of iterations

In tâtonnement [34], the price and consumption information is updated iteratively until the aggregate user demand matches the supply. Tâtonnement in smart grids has been proposed both from an optimization viewpoint and from a game theoretic perspective. Primal dual algorithms allow (primal) consumption to be set iteratively in response to (dual) prices [16]. Similar approaches have also been suggested [17] for the grid level supply demand matching problem, (known as Economic Dispatch in the power systems community). The game theoretic approach to these resource allocation problems assumes that consumers are selfish utility maximizing agents who maximize welfare (utility-price) in response to pricing signals [22, 1, 24].

Even with electricity pricing and distributed storage, herding effects [8] can actually cause an increase in the size of electricity peaks. That is, users with storage will purchase large amounts of electricity when prices are low, creating new peaks. In addition, storage and demand response can increase the unpredictability of demand. This particularly problematic in electricity markets, where, because of physical power system constraints distribution companies have to commit one day ahead to purchase levels [15].

Thus, an important question is how to incentivize users to coordinate their use of storage to flatten power consumption peaks while allowing distribution companies to gain estimates of the overall demand profile. Many variants of tâtonnement, where the central authority and consumption agents exchange consumption information and pricing signals until convergence, have been proposed in the multi-agent and power systems literature. This game theoretic perspective has been advanced both with storage in [33, 27] and without in [24] (an excellent survey covering these and other game theoretic approaches can be found in [29]). A similar approach, from an primal-dual optimization perspective (without storage) was suggested in [16].

However, all these protocols have some convergence issues that we seek to address here. For example, the protocols in [24, 16, 33, 27] do not converge if users update their consumption profiles in parallel. To see why this is the case, consider a simple example where we have two slots and large number n of users as shown in Figure 1. Suppose initially all users consume in the first slot. With parallel updates the users will alternate between State 1 and State 2. The optimal solution is the one where load on each slot is $n/2$ which will take $n/2 = O(n)$ iterations with serial updates. Thus, the protocol requires a linear number (in the number of consumers) of synchronous, serial updates, which is infeasible in large grids with millions of users and adds substantial signaling overhead. In this paper, we develop a parallel to the problem of server load balancing and show that using an algorithmic technique called bounded best response we can develop protocols that converge to near-optimum solution in a number of rounds independent of the number of users.

2.3 The Grid Model

User Agents and Preferences: Let $\mathcal{N} = \{1, 2, \ldots, n\}$ represent the set of autonomous agents or users, each of which controls its own load. We consider the model with discrete time intervals (i.e. time is divided into hourly slots or in slots of 15 minutes in a day). We divide a day into H time slots $\mathcal{H} = \{1, 2, 3, \ldots, H\}$. An agent i is available in time slots $\mathcal{S}_i$. This setting can represent, for example, the case where an agent controls an EV and has a preferable time slot to charge its vehicle. If the agent wishes to go to the office at $8:00$ A.M. in the morning and return at $7:00$ P.M. then the agent would charge the vehicle between slots $7:00$ P.M. to $8:00$ A.M. For simplicity, we assume that an agent has the flexibility to consume any amount of electricity in a time slot, but has a certain number of total units required for a particular day. The model is also general enough to be extended to the situation where multiple appliances are controlled by a single agent, or when per slot limits on appliance consumption are placed.
Consumption profiles: Based on the signals received from the central utility, each user i decides on an amount of electricity to be consumed in slot h denoted by l_i^h, which we call the user's consumption profile. This results in a total consumption of $L^h = \sum_{i\in\mathcal{N}} l_i^h$ in each time slot which is the overall consumption profile.
Distribution company optimization objectives: There are a number of different objectives that a central electricity distribution utility may be interested in. One important goal is to minimize the Peak to Average Ratio (PAR),

$$\text{PAR} = \frac{H \max_h L_h}{\sum_{i=1}^{H} L_h} \tag{1}$$

subject to the user consumption constraints, namely that (i) For any time $h \notin \mathcal{S}_i$, we have $l_i^h = 0$ and (ii) For each user i, there is a fixed requirement for charging, denoted by ξ_i such that $\sum_{h\in\mathcal{S}_i} l_i^h = \xi_i$ where ξ_i is the unit of electricity a user wants to consume which is fixed. Note that the denominator in calculating PAR is a constant in our model, as the energy requirement of each load agent is considered to be fixed and the problem reduces to that of minimizing $\max_{h\in H} L_h$, which is a peak load minimization problem.

There are number of possible (convex) cost functions that the central distribution company may wish to minimize. For example, the utility may be interested in minimizing a quadratic function of the total load $\sum_h aL_h^2 + bL_h + c$, since the generation cost of many generators can be modeled as a quadratic function [15]. Alternatively, peak minimization and I^2R loss minimization are often important goals [5]. The common feature of all these cost functions is that moving consumption from a heavily loaded time slot to a less loaded slot decreases the total cost. Constrained tâtonnement protocols can be shown to work for any such cost function, though we focus on the peak load minimization problem here.
Note that since cost function that utility company consider is convex monotonic increasing function in total load L_h, minimizing the maximum peak is identical to minimizing the total cost which is given by $\sum_h C(L_h)$ where C is the cost function that the utility company is paying.

2.4 Smart Grid Settings Investigated

The primary problem we are interested in in this work is of user, appliance or storage load sifting to meet system constraints and minimize costs and losses. Our aim is to show that by choosing constraints and prices appropriately we can improve the convergence of protocols suggested for a number of different smart grid scenarios [24, 16, 22, 33, 27] demonstrating the wide applicability of this idea. In this paper, we consider three settings that illustrate different aspects of the constrained tâtonnement protocol. Three of these were previously studied in the literature for which linear time algorithms were suggested and we demonstrate the improvements possible here.
Setting 1: Basic tâtonnement game: Here, we model autonomous loads and no local storage or per slot capacity constraints, motivated by [24, 27]. This allows us to highlight the parallel to the problem of load balancing and also demonstrate the proofs of fast incentive compatible, convergence for the basic peak minimization or convex cost minimization problem.
Setting 2: Demand management with capacity constraints: This scenario is unique to our work (from a distributed multi-agent perspective) and we demonstrate that it is possible to ensure distributed constraint satisfaction while maintaining fast incentive compatibility. In particular, we consider a utility that participates in the day ahead energy market [15]. In electricity markets, due to large generator start up times and slow ramping rates of the most efficient generators, electricity distributors and generators commit day ahead to levels of consumption and supply in each time slot over the next 24 hours. These are essentially contracts between generation and distribution companies and come with (large) penalties for any deviation from the committed levels. After considering these day ahead commitments and the predicted consumption of users who do not participate in demand management programs, a distribution company is then left with the problem of incentivizing users to schedule consumption subject to capacity constraints. We show, using the intuition from Setting 1, that we can develop a protocol that guarantees fast convergence for this important practical problem. An interesting point we make here is that for fast convergence with capacity limits, the form of the prices announced to each user have to be defined carefully : a direct extension of the previous methods does not work correctly. In particular we show that constrained tâtonnement guarantees convergence to a solution which satisfies capacity constraints, provided one exists, in a number of iterations independent of the number of users *if prices are structured appropriately.*
Setting 3: Distributed micro-storage management: This setting is motivated by [33], where each agent has access to limited local storage, possibly in the form of a parked EV or battery backup. The agent then wishes to use this storage to minimize costs, while the central utility would like to exploit user storage to minimize peaks. The fundamental question then is how the central utility should update prices in the presence of distributed storage. For this class of problems again a distributed linear time algorithm was presented in [33]. We show that constrained tâtonnement

allows fast convergence, even in the presence of selfish users who use storage to minimize their local costs.

Settings 2 and 3 serve to highlight the generality of the protocol and how it can be used for a wide variety of user and distribution company settings. We show through simulations on real consumption traces how to extend the protocol to the situation of a large number of distributed demand management agents with local storage and evaluate the practicality of our protocols.

3. CONSTRAINED TÂTONNEMENT FOR DISTRIBUTED ENERGY RESOURCE ALLOCATION

We now formalize our game theoretic model of the energy resource allocation problem between a central distribution utility and autonomous consumption agents in a setting we refer to as a tâtonnement game. This setting allows us to analyze the situation where a set of autonomous consumption agents selfishly try to maximize their utilities independently with co-ordination enforced by the utility through pricing and constraints.

3.1 Game Theoretic Model

Game theory is the study of strategic decision making. More formally, it is "the study of mathematical models of conflict and cooperation between intelligent rational decision-makers." [25]. The players in our game are the autonomous agents who are trying to optimize their consumption profile. Strategy for a player i is the load vector $l_i = \{l_i^1, l_i^2, \ldots, l_i^T\}$. We also denote l_{-i} as the load vectors of all other players except i. The payoff to each player i can then be defined as

$$u_i(l_i, l_{-i}) = -\frac{\xi_i}{\sum_j \xi_j} \sum_t c_t \qquad (2)$$

The first term ($\frac{\xi_i}{\sum_j \xi_j}$) in the payoff denotes the fraction of load agent i is using and c_t is the price the electricity company charges at time slot t. Thus $\sum_t c_t$ denotes the total amount that the electricity company receives. In our model, we simply set $c_t = C(L_t)$ where $L_t = \sum_i l_i^t$ (total amount of load consumed at time t), $C(L_t)$ is the cost, electricity company is paying for L_t unit of load at time t. We assume that $C(L_t)$ is a strictly convex increasing function. Thus the objective of each player i, would be to consume the load l_i so as to maximize their utility defined in (2)

We will be interested in two properties of a final consumption solution. A load profile $(l_1^*, l_2^*, \ldots, l_n^*)$ is a *Nash equilibrium* if:

$$u_i(l_i^*, l_{-i}^*) \geq u_i(l_i, l_{-i}^*) \;\forall\; l_i \;\forall\; i \qquad (3)$$

Nash equilibrium in some sense denotes the stability of the result of our protocol where no user has an incentive to deviate unilaterally. The other important property of a solution is Pareto optimality. A load profile $(l_1^*, l_2^*, \ldots, l_n^*)$ is Pareto optimal solution if $\nexists(l_1, l_2, .., l_n)$ such that:

$$\begin{aligned} u_i(l_i^*, l_{-i}^*) &\leq u_i(l_i, l_{-i}) \;\forall\; i \text{ and} \qquad (4)\\ u_i(l_j^*, l_{-j}^*) &< u_i(l_j, l_{-j}) \text{ for some } j \qquad (5) \end{aligned}$$

We then have the following theorems from [24]:

THEOREM 3.1. *A load profile $(l_1^*, l_2^*, \ldots, l_n^*)$ is a Nash equilibrium iff it is the solution of optimization problem* $min \sum_h C(L_h)$.

THEOREM 3.2. *If load profile $(l_1^*, l_2^*, \ldots, l_n^*)$ is a solution of optimization problem* $\min \sum_h C(L_h)$*, then it is a Pareto optimal profile.*

From the two theorems it can be seen that any algorithm which minimizes the cost function $\sum_h C(L_h)$ will also result in Nash equilibrium and Pareto optimal solution. In the following sections we develop a distributed algorithm in which player play selfishly subject to centrally verifiable constraints to reach to the optimal solution.

3.2 Constrained Tâtonnement

In the basic Setting 1, autonomous loads are co-ordinated by price updates. We first discuss how the parallel can be established to the problem of server load balancing studied [4] and then prove fast convergence. We then extend the results to Setting 2 and 3 as how to impose capacity constraints and how to include storage in the model. Interestingly, we show that imposing capacity constraints directly does not work and one needs to design the cost functions with care.

Recall that we consider a setting where agents have appliance that have certain consumption requirements and are only available to be run in a certain sub-set of the time slots. This models a number of appliance like EVs and even appliances like heaters and air conditioners that have explicit or implicit storage capabilities. The tâtonnement process we study is primarily a model for investigating stability of equilibrium [34]. In each iteration, prices for each slot are announced (by the central utility), and agents state how much electricity they would like to purchase (demand) in each time slot at that price. No transactions take place at non-equilibrium prices. Instead, prices are lowered for slots with positive prices and excess supply and raised for slots with excess demand. Even though they were not explicitly named as such, the protocols suggested in [24, 27, 33] and many other are essentially tâtonnement processes, which require a linear number of sequential updates to ensure convergence. We show that, with suitable modifications to this basic protocol, it is possible to get bounds on the number of updates required for convergence that depend only on the approximation factor, independent of the number of agents. This also highlights the fact that even settling for an approximate solution can often lead to a substantial speed-up in tâtonnement protocols.

There is a strong parallel between our problem of energy resource allocation and the work on server scheduling [4]. In [4], they consider the problem of allocating jobs to servers to minimize peak loads on servers. In our model, jobs correspond to the loads or appliances controlled by the agents and each machine corresponds to one of the time slots. The total load on a machine corresponds to the amount of energy consumed in each time slot. To emphasize the parallel, we will also use load to refer to the energy consumed.

3.3 Setting 1 : Basic Setting

Let S_i denotes the set of time slots in which player i prefers to consume some load and p_{ih} be the fraction of load which the i^{th} user consumes at h^{th} hour and d denotes the iteration

of the algorithm. We have $\sum_{h \in S_i} p_{ih} = 1, l_i^h = \xi_i p_{ih}\ \forall h \in S_i$. Each user in this setting will try to balance the load subject to $\sum_{t \in S_i} l_i^t = \xi_i\ \forall i$. In the basic setting, the two constraints for the tâtonnement protocol that the distribution company broadcasts and a user must satisfy in each update are the
(i) **Inertia constraints** whereby an agent i may move a fraction of load from hour j to hour r where $j, r \in S_i$ only if $L_j \geq (1+\epsilon)^3 L_r$ and the
(ii) **Bounded Step constraints**, $\eta \leq p_{ih} \leq 1$ for all i where $\eta = \frac{\epsilon}{H^2}$ and $p_{ih}^d/(1+\epsilon) \leq p_{ih}^{d+1} \leq (1+\epsilon)p_{ih}^d$.
Here ϵ is a parameter that we select later depending on the accuracy required. Essentially, the inertia rule allows users to make a move only if there is significant difference in cost (ensuring convergence) and the bounded step rule prevents large swings and cycles in the dynamics of the protocol. The formal outline of the protocol is in 1. The protocol does not requires communication between the agents and thus required less overhead.

Protocol 1 The constrained tâtonnement protocol for distributed demand management

```
1: while Any change in user consumption profiles do
2:     Distribution utility broadcasts total load in each time
       slot and corresponding cost functions for each slot
3:     Users update consumption profile based on the cost
       function, internal preferences, Bounded Step constraints
       and Inertia constraints.
4:     Users transmit new consumption profiles to the dis-
       tribution company
5:     The distribution utility checks that inertia con-
       straints and bounded step constraints are not violated
6: end while
```

Distributed Incentive Compatibility: In Protocol 1, every user is in fact unaware of other users' loads, leading to a distributed protocol. The user only sees the total load present on the slots in which she is consuming energy and tries to balance consumption according to her preferences, the cost functions announced by the central utility for those slots and the constraints making the protocol incentive compatible. Privacy problems do not arise since the load profiles of users are not required to be public knowledge.

The central authority announces prices of $(L_h)^2$, for each $h \in \mathcal{S}_i$ to each agent i. Thus, in iteration $d+1$, user i is faced with the following optimization problem,

$$\underset{\gamma_1,\ldots,\gamma_H}{\text{minimize}} \quad \sum_{t \in \mathcal{S}_i} (L_t^d - \gamma_t)^2 \tag{6}$$

$$\begin{aligned} \text{s.t.} \quad & \gamma_t = \sum_{t' \in R_t} x_{t,t'}\ \forall\ t,\ \sum_{t \in S_i} \gamma_t = 0 \\ & \gamma_t \leq p_{it}^d\ \xi_i\ \frac{\epsilon}{1+\epsilon},\ -\gamma_t \leq p_{it}^d\ \xi_i\ \epsilon \end{aligned} \tag{7}$$

where γ_t represents the amount of load agent i moves from (or to) time t. $R_t = \{t' : L_t \geq (1+\epsilon)^3 L_{t'} \text{ or } L_{t'} \geq (1+\epsilon)^3 L_t\}$ denotes the set of time slots t' for time t such that user can either shift his load from t to t' or from t' to t and finally $x_{t,t'}$ is a design variable which is positive when the load is shifted from t to t' and negative when load is shifted from t' to t. R_t and $x_{t,t'}$ arise from the inertia constraints and the bound on γ_t comes from the bounded step constraints.

Let OPT be the the maximum load on any slot in the optimal assignment that minimizes the peak load over all $h \in H$. To allow the protocol to work we assume that each user starts with at least an η fraction of its total load in each time slot $t \in \mathcal{S}_i$. Since, $OPT \geq \frac{1}{H}\sum_i \xi_i$, the effect of this assumption on the final solution is small. Two lemmas follow immediately,

LEMMA 3.1. *For all hours j and iterations d, we have* $\frac{L_j^d}{1+\epsilon} \leq L_j^{d+1} \leq (1+\epsilon)L_j^d$

LEMMA 3.2. *The dynamics are acyclic.*

PROOF. By the inertia constraint, load can move from hour j to hour r only if $\frac{L_j}{L_r} \geq (1+\epsilon)^3$. Since agents are maximally greedy, the load on slot j will decrease at least by $\frac{L_j}{1+\epsilon}$ and due to Lemma 3.1, the load of slot r will increase to at most $L_r(1+\epsilon)$, resulting in a total decrease of load by at least a factor of $(1+\epsilon)$. The sum of squares of loads decreases over time and thus the dynamics are acyclic. □

Thus, if both inertia and bounded constraints are satisfied at each iteration then situations like in Figure 2 can be avoided. Using these lemmas, it is possible to derive the following result,

THEOREM 3.3. *The constrained tâtonnement protocol converges to a $(1+\delta)$ approximation in $O(\frac{1}{\delta^6} log^6 \frac{H}{\delta})$ iterations with $\epsilon = O(\frac{\delta^2}{log H})$.*

Proof follows from the results in [4]. Thus, the protocol converges in a number of iterations which is independent of the number of agents unlike the prior approaches discussed in Section 2.2.

3.3.1 Simulations for Setting 1

To obtain the load profiles of users in all our experiments we used the model and data collected as part of an extensive sensing and survey based study of U.K. homes, where electricity demand was recorded over the period of a year within 22 homes in the UK [28]. For each agent i, the time slots are chosen i.i.d. following the average demand distribution in Figure 3. The number of slots for each agent i, $|\mathcal{S}_i|$, is drawn from a multinomial distribution with mean 4 (except in Figures 7 and 8 with varying mean).

We first investigate the effect of various parameters on the convergence and optimality of the constrained tâtonnement protocol. For studying the effect of δ and expected number of slots, we fixed the number of users to be 50. We typically ran 500 experiments to obtain 95% confidence bounds. Let COST be the cost of the final solution obtained by the protocol and OPT be the optimal solution achieved by running the optimal problem centrally. From Theorem 3.3, we see that the required approximation factor δ trades off accuracy and rate of convergence. Figures 4 and 5 show the change in the number of rounds and the COST/OPT ratio of the protocol with different δ's. As δ increases, the protocol converges faster though the final COST increases, however the increase is not substantial. In particular, we see that to achieve a cost within 1% of the optimal, required only about 60 iterations. *Note that from Figure 6 this number does not depend on the number of agents.* In addition, Figure 4 shows the speed-ups possible with just 2000 agents, hinting at the promise of when large smart grids with millions of users.

Figure 6 shows that the number of iterations required for convergence with the proposed protocol is independent of

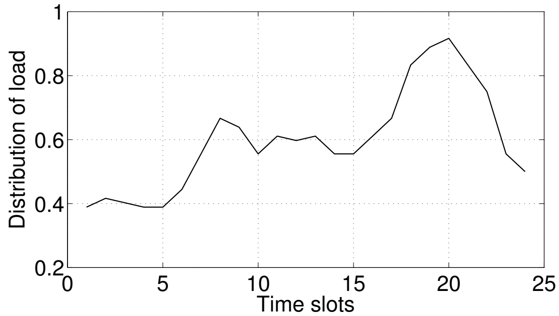

Figure 3: Distribution of the load on a hourly basis

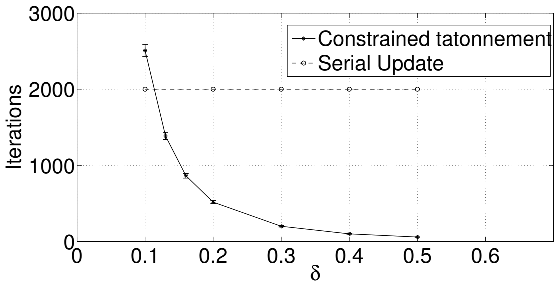

Figure 4: Decrease in the number of iterations with increase in the approximation factor δ

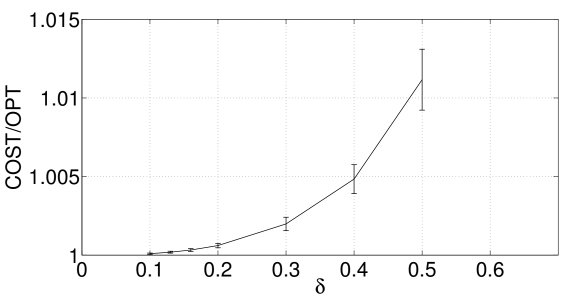

Figure 5: Sub optimality of final solution with desired approximation factor δ

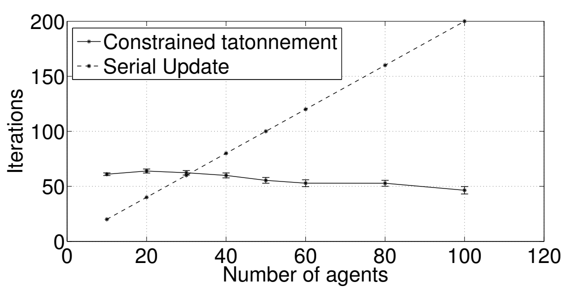

Figure 6: Comparison of sequential updates and the constrained tâtonnement protocol

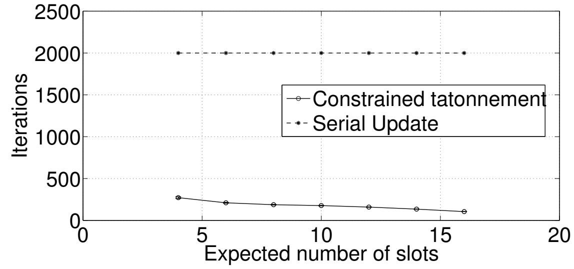

Figure 7: Faster convergence of constrained tâtonnement with increased user flexibility $E[|\mathcal{S}_i|]$

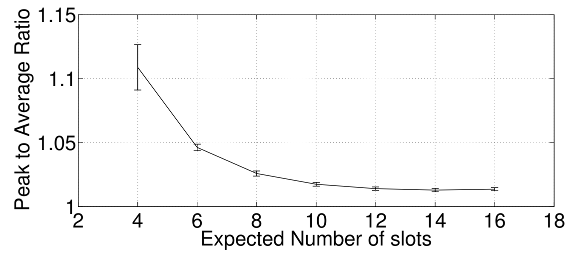

Figure 8: Decrease in peak to average ratio with increased user flexibility $E[|\mathcal{S}_i|]$

the number of users, in agreement with Theorem 3.3. This is in contrast to the linear increase in number of iterations for the sequential protocols in [24, 33]. For comparison, we kept the value of $\delta = 0.5$, thus $\epsilon = 0.0787$. Figures 7 and 8 show how the number of rounds and peak average ratio vary with the increase in expected number of slots, $E[|\mathcal{S}_i|]$, a representation of the patience of the users. We see that when users are more flexible, the protocol converges much faster as they have more choices to shift, also resulting in decrease in peak load. However, we note that the standard protocols do not benefit from increased user flexibility.

3.4 Setting 2 : Incorporating Capacity Constraints for Each Hour

As described in Sections 1 and 2, distribution companies make day ahead commitments and so want users to balance load subject to per slot capacity constraints of the form $L_t = \sum_{i=1}^{n} l_i^t \leq C_t, \forall t$. Now, depending on the requirements of the distribution utility, various types of cost functions could be communicated to the users. However, incorporating the capacity constraint does not follow as a direct extension to the protocol described in the previous section.

Non-Triviality: Suppose our goal is to minimize $\max_t L_t^2$ subject to the capacity limits and we directly apply Protocol 1 developed in the previous section, adding an additional constraint that the users are allowed to move a fraction of load from slot j to slot r if r is not at its capacity. Consider a scenario where $L_r < C_r,\ L_j > C_j,\ L_j < L_r$ then the user would shift load from a time j to time r because it has unfilled capacity, but that results in violating the inertia constraint.

Our solution to this problem is to use *scaling*, whereby we use constrained tâtonnement to balance $L_t/C_t,\ \forall\ t$. Thus, the cost function announced to the users is a scaled quadratic cost, This scaling can be done by defining a 'scaled world' where, $l'_{ih} = l_{ih}/C_h, l'_{ih} = \xi'_{ih}p_{ih}, \xi'_{ih} = \xi_i/C_h \forall h, i,\ C'_h = 1, L'_h = L_h/C_h\ \forall h$ and $, \sum_h p_{ih} = 1\ \forall i$. Now, shifting γ unit of load by user i from hour j to hour r, in round $d+1$ implies:

$$l'^{d+1}_{ij} = l'^{d}_{ih} - \gamma \text{ and } l'^{d+1}_{ir} = l'^{d}_{ir} + \gamma\frac{C_j}{C_r} \tag{8}$$

Note that since the end user is unaware of the scaling, utility company should be providing the simple rules to the end customer for simplicity. So that users won't have to be worried about the capacity constraints. (8) ensures that if a user shifts his load from low capacity to high capacity then his utility increases more compared to shifting load from high capacity to low capacity. Note that we can enforce (8) by designing the cost function in a proper way so that user the does not have to take into account complication shifting. The two constraints can then be modified as (i) **Inertia constraints:** A fraction of load of user i may move from hour j to hour r where $j, r \in S_i$ only if $L_j \geq (1+\epsilon)^3 L_r \frac{C_j}{C_r}$ and (ii) **Bounded step constraints:** $\eta \leq p_{ih} \leq 1$ for all i where $\eta = \frac{\epsilon}{H^2}$ and $p^d_{ih}/(1+\epsilon) \leq p^{d+1}_{ih} \leq (1+\epsilon)p^d_{ih}$. The bounded step rule remains the same as it imposes constraints only on the fraction of load and not the absolute load, but the inertia constraints change because of the scaling. We will now see that after imposing the modified constraints and following the rules given in (8) the dynamics will remain acyclic.

LEMMA 3.3. *By applying modified inertia rule and bounded step constraint, dynamics are acyclic. Moreover if there exist a feasible solution(i.e. exist a load profile such that all the slots are within capacity limit satisfying the constraints) then at the end of the algorithm we have $L_t < (1+\epsilon)^3 C_t\ \forall t$ i.e. the load on each slots are within some approximate fraction of its capacities.*

PROOF. First we will prove that the dynamics are acyclic. If a user i shifts ρ fraction of load from j^{th} hour to r^{th} hour at iteration $d+1$, then he will be charged according to the following updation rules: $p^{d+1}_{ij} = p^d_{ij} - \rho p^d_{ij},\ p^{d+1}_{ir} = p^d_{ir} + \rho p^d_{ij}\frac{C_j}{C_r}$
But by bounded step rule we have

$$p^d_{ij} - \rho p^d_{ij} \geq \frac{p^d_{ij}}{1+\epsilon},\ p^d_{ir} + \rho p^d_{ij}\frac{C_j}{C_r} \leq p^d_{ir}(1+\epsilon) \tag{9}$$

By equation (9) we have:

$$\rho p^d_{ij} \leq p^d_{ij}\frac{\epsilon}{1+\epsilon}, \rho p^d_{ij} \leq \frac{C_r}{C_j}p^d_{ir}\epsilon \tag{10}$$

Note that in order to shift ρp^d_{ij} amount of load from time j to r, user i has to choose

$$\rho p^d_{ij} = min\{p^d_{ij}\frac{\epsilon}{1+\epsilon}, \frac{C_r}{C_j}p^d_{ir}\epsilon\}$$

Consider 2 cases:

- Case 1: if $p^d_{ij}\frac{\epsilon}{1+\epsilon} > \frac{C_r}{C_j}p^d_{ir}\epsilon$ then
  $p^{d+1}_{ir} = p^d_{ir} + p^d_{ir}\epsilon$
  Thus load on machine r increases by at most $(1+\epsilon)$ fraction.
- Case 2: $p^d_{ij}\frac{\epsilon}{1+\epsilon} < \frac{C_r}{C_j}p^d_{ir}\epsilon$ then
  $p^{d+1}_{ij} = \frac{p^d_{ij}}{1+\epsilon}$
  Thus load on machine j decreases at least by $\frac{1}{1+\epsilon}$ factor.

Next we will show that once the algorithm converges all the slots are within approximate capacity limits. Suppose when the algorithm converges $\exists\, t$ such that $L_t > C_t(1+\epsilon)^3$. Since there exists a feasible solution
$\Rightarrow \exists i, t_j$ s.t. $L_{t_j} < C_{t_j}$ and $t, t_j \in S_i$. Since $\frac{L_t}{C_t} > \frac{L_{t_j}}{C_{t_j}}$, it would be beneficial for agent i to shift the load from t to t_j. Thus we can assume that the inertia rule fails
$\Rightarrow \frac{L_t}{C_t} < \frac{L_{t_j}}{C_{t_j}}(1+\epsilon)^3 < (1+\epsilon)^3$ which leads to contradiction. $\square$

Thus, the central utility announces quadratic scaled prices such that agent i is faced with the following optimization problem in iteration $d+1$,

$$\begin{aligned} \underset{\gamma_1,\dots,\gamma_H}{\text{minimize}} \quad & \sum_{t\in S_i}(L_t - \gamma_t - \frac{(\sum_{t'\in S_i} L_{t'})C_t}{\sum_{t'\in s_i} C_{t'}})^2 \\ \text{s.t.} \quad & \gamma_t = \sum_{t'\in R'_t} x_{t,t'}\forall t, \sum_{t\in S_i}\gamma_t = 0 \\ & \gamma_t \leq p^d_{it}\ \xi_i\ \frac{\epsilon}{1+\epsilon}, -\gamma_t \leq p^d_{it}\ \xi_i\ \epsilon \end{aligned} \tag{11}$$

where $R'_t = \{t' : L_t \geq (1+\epsilon)^3 L_{t'}\ C_t/C_{t'}$ or $L_{t'} \geq (1+\epsilon)^3 L_t\ C_{t'}/C_t\}$

This quadratic cost function ensures that it is in best interest of every player i to balance the load $L_t/C_t\ \forall t \in S_i$. If such prices and constraints are used then we have the following theorem,

THEOREM 3.4. *The constrained tâtonnement dynamics in the 'scaled world' converges to $(1+\delta)$ approximation in $O(\frac{1}{\delta^6}log^6\frac{H}{\delta})$ iterations for $\epsilon = O(\frac{\delta^2}{logH})$.*

We then have $L'_{max} \leq (1+\delta)OPT'$ where OPT' is the optimal load in the scaled world. In the original world, let C'_* and C''_* be the capacity of slot having the OPT' and L'_{max} load respectively. Then we have

$$\frac{L_{max}}{C_{max}} \leq \frac{L_{max}}{C'_*} \leq (1+\delta)\frac{OPT}{C''_*} \leq (1+\delta)\frac{OPT}{C_{min}} \tag{12}$$

where C_{max} and C_{min} are the largest and smallest capacities, respectively. Hence in the original (unscaled) world we have the following corollary:

COROLLARY 3.1. *The scaled constrained tâtonnement protocol results in a $(1+\delta)*(\frac{C_{max}}{C_{min}})$ approximation in $O(\frac{1}{\delta^6}log^6\frac{H}{\delta})$ iterations for $\epsilon = O(\frac{\delta^2}{logH})$.*

Here again we see that an approximately optimal solution *that satisfies the capacity constraints* is reached in a number of iterations independent of the number of agents.

3.4.1 Simulations for Setting 2

With updated rules we now illustrate the dynamics of the scaled constrained tâtonnement protocol where users solve (11). An important point to emphasize is that whenever there exists a solution that satisfies the capacity constraints, $L_t \leq C_t \forall t$, the protocol will output a feasible solution. For simulations in this section we fixed number of agents to be 50, $\delta = 0.2$. Expected number of slots for every users are selected by uniform distribution between 2 to 24 and capacities of each slot are also chosen uniformly between C_{min} and C_{max}. We define cost of the solution as $\sum_t L_t^2$ whereas OPT is defined as minimum value of $\sum_t L_t^2$ s.t. $L_t \leq C_t$.

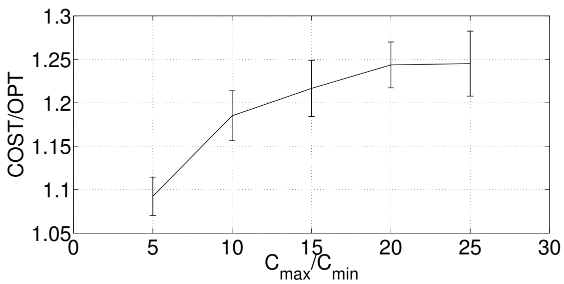

Figure 9: Increased sub-optimality (COST/OPT) with increasing bandwidth (C_{max}/C_{min})

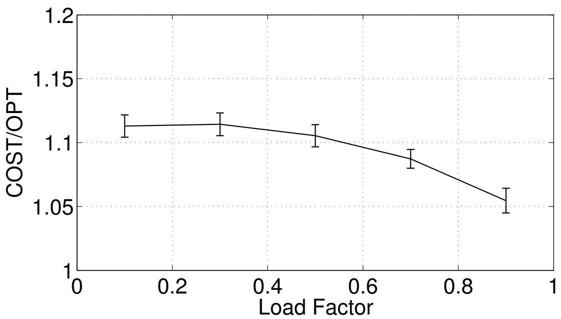

Figure 10: Decreasing of sub-optimality (COST/OPT) with increased load factor ($\frac{\sum_h C_h}{\sum_{i,h} l_i^h}$)

Figure 9 shows that as the ratio of the maximum capacity and minimum capacity, also called the bandwidth ($\frac{C_{max}}{C_{min}}$) increases, the ratio between protocol cost and optimal cost also increases, in agreement with Corollary 3.1. The results also show that even though we are balancing L_t/C_t we are not that far from the true optimal solution which balances L_t over all t.

We next try to understand the performance of the protocol under different loads. We define the 'load factor' of the system to be the fraction $\frac{\sum_h C_h}{\sum_{i,h} l_i^h}$. Thus, a load factor of 1 means that the system is completely loaded to capacity. For the purpose of simulation in this setting we try to find the values of ξ_i by fixing the load factor on every slot. After finding the $\xi_i's$ according to the load factor we then run our algorithm after placing random load from each user on each time slots. Figure 10 shows that when the system is operating at or near capacity, the protocol performs particularly well. The reason is that when the load factor is smaller it results in larger scaling of the total load available for the slots which has higher capacity.

3.5 Setting 3 : Distributed Micro-Storage Management

In the storage problem, each household has storage of capacity B_{max} (We considered here that each user has same storage capacity but the algorithm can be easily extended for different storage capacity as users are running optimization problem at their end independent of other users) and a consumption profile dm such that dm_t amount of load is required at each time slot t. Each user i then has t additional constraints, corresponding to the requirement that the

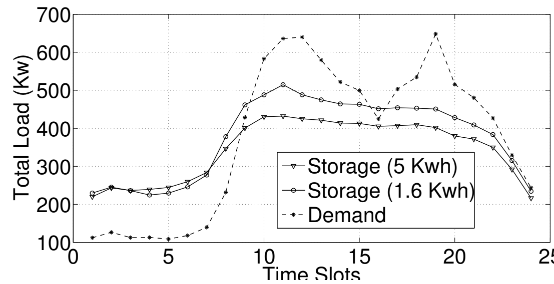

Figure 11: Flattening of overall load profile though constrained tâtonnement with micro-storage

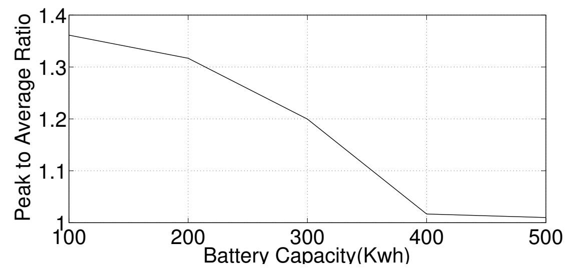

Figure 12: Decrease in peak to average ratio with increased per user storage capacity

electricity consumed in each time slot must be drawn from the grid prior to t. For this problem we directly apply constrained tâtonnement with the center announcing quadratic costs, $(L_h)^2$ where in addition to the inertia and bounded step constraints each user has to satisfy their own internal storage constraints. The optimization problem faced by user i in iteration $d+1$ is,

$$\underset{\gamma_1,\dots,\gamma_H}{\text{minimize}} \quad \sum_{t\in\mathcal{H}} (L_t^d - \gamma_t)^2 \tag{13}$$

$$\text{s.t.} \quad 0 \le \sum_{t=1}^{\tau} (l_i^t - \gamma_t - dm_t) \le B_{max} \ \forall\tau, \gamma_t = \sum_{t'\in R_t} x_{t,t'} \ \forall t$$

$$\sum_{t\in\mathcal{H}} \gamma_t = 0, \ \gamma_t \le p_{it}^d \xi_i \frac{\epsilon}{1+\epsilon}, \ -\gamma_t \le p_{it}^d \xi_i \epsilon$$

Note that the goal here is to find the storage profile given the load profile. One interesting direction can be to combine setting 1 and setting 3 so that a user can update the load and storage profiles in parallel.

3.5.1 Simulations for Setting 3

We have simulated daily energy consumption profiles based on the models in [28] to obtain dm_t. This model results in realistic load curves as seen in Figures 3 and 11. The broken line in Figure 11 represents the initial total load profile without storage and the solid lines represent the load profiles with storage at the end of the tâtonnement protocol(with battery capacity 1.6 KWh and 5 KWh). We can see from Figure 11 that the peak load decreases substantially with constrained tâtonnement. Figure 12 shows that as the storage capacity each user has increases, the peak load decreases. Thus, constrained tâtonnement prevents the herding that was observed in the agent models of [8, 33]. As shown in

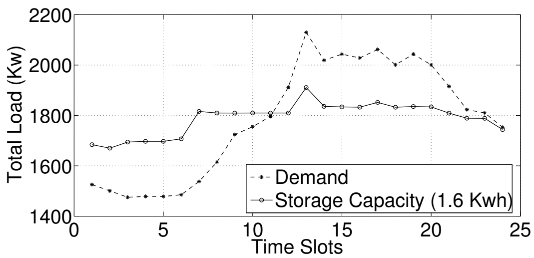

Figure 13: Flattening of overall load profile with storage

figure 11, using storage capacity of 1.6 Kwh results in 27.6% peak reduction whereas increasing storage capacity to 5 Kwh results in 33.6% reduction in peak. To demonstrate the generality of our results, we also conduct experiments on the microgrid data set from [6]. This data set consists of electrical data over a single 24-hour period from 443 unique homes. In Figure 13 we see that substantial reductions in peaks are possible with co-ordinated charging of local storage.

4. RELATIONSHIP TO PRIOR WORK

This work builds on a few different research directions that we review here for clarity. The first is the line of research pioneered in [19] where they "argue that the concepts and technologies pioneered by the Internet - the fruits of the past four decades of Internet research - can make fundamental contributions to the architecture and operations of the future grid." They identify the similarities between the cyber-physical systems that constitute the internet and the next generation smart grid and propose that many of the techniques studied, developed and proven effective in the internet could be used to design protocols for the smart grid. In particular they outline the possible benefits of both proactive and reactive congestion control, made possible by the large scale deployment of smart meters throughout the grid. [3] suggest the use of these techniques for real time congestion management. Extending this line of work we introduce game theoretic principles, modelling users as self interested agents who may lie to maximize their own welfare. In addition, we conduct a rigorous analysis of the number of iterations and sub-optimality of these algorithms.

The second line of work is based on agent based models of storage [33] and demand [27, 24] management. In particular, [33, 24] consider game theoretic concepts such as individual rationality and solve for the Nash equilibrium. However, the protocols suggested require a linear number of iterations to converge making them infeasible in large grids or even for large simulations as mentioned in [33]. Work such as [22, 16] show how how to capture the utilities of a large number of different appliances using simple constraints and create a distributed optimization problems which can then be solved by tâtonnement. These approaches form the motivation for our Setting 3. *Our primary contribution to this line of research is to show that it is possible to add simple, centrally verifiable constraints to all of these distributed demand management problems to obtain convergence in a number of iterations independent of the number of users, a substantial speed-up over the linear number of iterations required in [24, 33].* This improvement is particularly important, as mentioned earlier, in modern distribution grids with millions of users and possibly tens of millions of smart appliances.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a tâtonnement framework for resource allocation among intelligent agents in the smart grid, that non-trivially generalizes past work in this area. Using a parallel to work in server load balancing we developed fast, incentive compatible, distributed protocols by communicating *constraints* along with pricing signals. We adapt the theoretical work to a variety of practical scenarios and show that incorporating novel constraints in practical settings requires a rethinking of how prices and constraints are set. Extensive simulations validate the improvements possible with constrained tâtonnement. A highlight of our simulations is that in some scenarios performance was substantially better than expected from the (worst case) theoretical bounds. In particular, fewer iterations were required to achieve a smaller optimality gap than expected in Figures 4 and 7. Improving the bounds with better models of agents and their consumption is an interesting direction of future work. In addition, special tâtonnement strategies can be (and in some cases have been) applied to other important resource allocation problems in next generation cyber-physical systems where our work may find application. We would also like to extend our work to some more complicated setting where agents have the preferences over available slots and to include maximum charging rates of the battery in setting 3. In setting 2, we assumed that there exist a load profile consumption which satisfies the capacity constraints. In future, we would like to present the solution concept where such consumption profile does not exist. For example, extending these protocols to handle deadlines for the completion of certain loads which an agent may have. Finally, a better understanding of the practical overheads and comparison of different protocols is an important direction of future work.

6. REFERENCES

[1] M. Albadi and E. El-Saadany. Demand response in electricity markets: An overview. In *IEEE Power Engineering Society General Meeting*, 2007.

[2] R. Anderson and S. Fuloria. On the security economics of electricity metering. *Proceedings of the WEIS*, 2010.

[3] O. Ardakanian, C. Rosenberg, and S. Keshav. Real-time distributed congestion control for electrical vehicle charging. *Greenmetrics*, 2012.

[4] B. Awerbuch, Y. Azar, and R. Khandekar. Fast load balancing via bounded best response. In *SODA*, pages 314–322, 2008.

[5] P. Barker and R. De Mello. Determining the impact of distributed generation on power systems. i. radial distribution systems. In *IEEE Power Engineering Society Summer Meeting*, volume 3, pages 1645 –1656 vol. 3, 2000.

[6] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht. Smart*: An open data set and tools for enabling research in sustainable homes. 2012.

[7] S. Borenstein. The trouble with electricity markets: Understanding California's restructuring disaster. *The Journal of Economic Perspectives*, 16(1):191–211, 2002.

[8] T. Carpenter, S. Singla, P. Azimzadeh, and S. Keshav. The impact of electricity pricing schemes on storage adoption in ontario. In *e-Eenergy*, 2012.

[9] H. Chao. Competitive electricity markets with consumer subscription service in a smart grid. *Journal of Regulatory Economics*, pages 1–26, 2012.

[10] K. Clement-Nyns, E. Haesen, and J. Driesen. The impact of charging plug-in hybrid electric vehicles on a residential distribution grid. *Power Systems, IEEE Transactions on*, 25(1):371–380, 2010.

[11] EPRI and NRDC. Environmental assessment of plug-in hybrid electric vehicles. volume 1: Nationwide greenhouse gas emissions. Technical Report 1015 325, July 2007.

[12] H. Farhangi. The path of the smart grid. *IEEE Power and Energy Magazine*, 8(1):18–28, 2010.

[13] S. Ghosh. Electricity consumption and economic growth in india. *Energy Policy*, 30(2):125–129, 2002.

[14] S. Hadley and A. Tsvetkova. Potential impacts of plug-in hybrid electric vehicles on regional power generation. *The Electricity Journal*, 22(10), 2009.

[15] C. Harris. *Electricity Markets: Pricing, Structures and Economics*. Wiley, Sussex, England, 2006.

[16] L. Huang, J. Walrand, and K. Ramchandran. Optimal smart grid tariffs. In *Information Theory and Applications Workshop (ITA)*, pages 212 –220, 2012.

[17] M. Ilic, L. Xie, and J. Joo. Efficient coordination of wind power and price-responsive demand Part I: Theoretical foundations. *IEEE Transactions on Power Systems*, 26(4):1875–1884, 2011.

[18] International Energy Agency. The power to choose - enhancing demand response in liberalised electricity markets findings of IEA demand response project. 2003.

[19] S. Keshav and C. Rosenberg. How internet concepts and technologies can help green and smarten the electrical grid. *ACM SIGCOMM Computer Communication Review*, 41(1):109–114, 2011.

[20] M. Kintner-Meyer, K. Schneider, and R. Pratt. Impacts assessment of plug-in hybrid vehicles on electric utilities and regional us power grids, Part 1: Technical analysis. *Pacific Northwest National Laboratory*, 2007.

[21] P. Kundur, N. Balu, and M. Lauby. *Power system stability and control*, volume 4. McGraw-hill New York, 1994.

[22] N. Li, L. Chen, and S. Low. Optimal demand response based on utility maximization in power networks. In *IEEE Power and Energy Society General Meeting*, pages 1 –8, july 2011.

[23] N. A. Lynch. *Distributed algorithms*. Morgan Kaufmann, 1996.

[24] A. Mohsenian-Rad, V. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia. Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid. *IEEE Transactions on Smart Grid*, 1(3), 2010.

[25] R. Myerson. *Game theory: analysis of conflict*. Harvard University Press, 1997.

[26] Y. Narahari, D. Garg, R. Narayanam, and H. Prakash. *Game theoretic problems in network economics and mechanism design solutions*. Springer, 2009.

[27] S. Ramchurn, P. Vytelingum, A. Rogers, and N. Jennings. Agent-based control for decentralised demand side management in the smart grid. In *AAMAS*, pages 5–12, February 2011.

[28] I. Richardson, M. Thomson, D. Infield, and C. Clifford. Domestic electricity use: A high-resolution energy demand model. *Energy and Buildings*, 42(10):1878–1887, 2010.

[29] W. Saad, Z. Han, H. V. Poor, and T. Basar. Game theoretic methods for the smart grid. *CoRR*, abs/1202.0452, 2012.

[30] V. Sood, D. Fischer, J. Eklund, and T. Brown. Developing a communication infrastructure for the smart grid. In *Electrical Power & Energy Conference (EPEC), 2009 IEEE*, pages 1–7. Ieee, 2009.

[31] G. Strbac. Demand side management: Benefits and challenges. *Energy Policy*, 36(12):4419 – 4426, 2008.

[32] R. Urgaonkar, B. Urgaonkar, M. Neely, and A. Sivasubramaniam. Optimal power cost management using stored energy in data centers. In *Proceedings of the ACM SIGMETRICS*, pages 221–232. ACM, 2011.

[33] P. Vytelingum, T. Voice, S. Ramchurn, A. Rogers, and N. Jennings. Agent-based micro-storage management for the smart grid. In *AAMAS*, pages 39–46, 2010.

[34] D. Walker. Walras's theories of tatonnement. *The Journal of Political Economy*, 95(4):758–774, 1987.

[35] L. Xie and M. Ilic. Model predictive dispatch in electric energy systems with intermittent resources. In *IEEE International Conference on Systems, Man and Cybernetics*, pages 42 –47, oct. 2008.

[36] T. Zhu, A. Mishra, D. Irwin, N. Sharma, P. Shenoy, and D. Towsley. The case for efficient renewable energy management for smart homes. *Proceedings of the Third Workshop on Embedded Sensing Systems for Energy-efficiency in Buildings (BuildSys)*, November 2011.