Manipulating Residents' Behavior to Attack the Urban Power Distribution System

Gururaghav Raman, *Student Member, IEEE*, Jimmy Chih-Hsien Peng, *Member, IEEE*, and Talal Rahwan

***Abstract*—The reliable operation of the power distribution system is a matter of national security. Increasingly, urban distribution systems rely on communications between customers and the utility to implement consumer-centric programs such as demand response that enhance the grid resilience. This paper reports an unconventional and previously-unexamined mode of malicious attack on the power distribution infrastructure of cities. It demonstrates that consumer behaviors in such a system could be manipulated by an attacker using false communications, which could significantly impact the system reliability. Using a novel decision-making model for consumer response, possible network impacts of such an attack are examined, which include reduction in system reserves, increase in peak demand, lower voltage profiles, and potential system blackouts. These detrimental effects are shown to worsen in the future as more consumers join such programs and adopt flexible high-power loads. Further, though the system is resilient to random errors or failures, it remains highly vulnerable to strategic attacks like those demonstrated here. These results recommend urgency in developing solutions to detect and tackle possible injection of fake information into such critical systems, which, as shown here, can have a very real impact on the energy infrastructure reliability.**

***Index Terms*—Consumer behavior, behavioral demand response, power system security, vulnerability identification.**

NOMENCLATURE

Variables

$\mathbf{A} = \{A_i\}$	Set of home appliances.
$\mathbf{A}_{dr}$	Set of DR-capable appliances.
$\mathbf{P}_i = \{P_{t,i}\}$	Set of probabilities that appliance A_i starts at each time step $t \in \{1, 2, \ldots, T\}$.
$P^*_{t,i}$	Updated probability that A_i starts at time step t after following through a DR event.
$x_{t,i}$	State of appliance A_i at time step t, which is 1 if A_i is turned on, and 0 otherwise.
$\mathcal{P}_i$	Real power rating of appliance A_i.
$[t_{start}, t_{end}]$	Interval of the DR event.
β	Fraction of consumers who believe a DR message.
β_a	Fraction of consumers who believe the attacker's message.
χ	Fraction of consumers that receive the defender's counter-message.
β_d	Fraction of consumers who believe the defender's counter-message.
θ_b	Belief status of a consumer, which is 1 if she believes a DR message, and 0 otherwise.
θ_a	Propensity of a consumer to accept a DR event.
θ_{ft}	Degree of follow-through of a consumer.
θ_{util}	Fraction of load reduction requested by the utility.
θ	Net degree of compliance of a consumer to a DR event.
$\mathbf{S^A} = \{s^A_j\}$	Set of the attacker's strategies.
$\mathbf{S^D} = \{s^D_k\}$	Set of the defender's strategies.
$P^A(s^A_j)$	Probability that the attacker will play strategy s^A_j.
$P^D(s^D_k)$	Probability that the defender will play strategy s^D_k.
ϕ_a	Attacker's payoff.
ϕ_d	Defender's payoff.
K	Cost of carrying out an attack.

This work was supported in part by the National Research Foundation, Singapore under Grant NRF2018-SR2001-018.

G. Raman and J. C.-H. Peng are with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore, 117583 (email: gururaghav.raman@u.nus.edu, jpeng@nus.edu.sg).

T. Rahwan is with the Department of Computer Science, New York University Abu Dhabi, UAE (email: trahwan@gmail.com).

I. INTRODUCTION

RENEWABLE sources of energy such as solar photovoltaic and wind generation are essential to achieving a sustainable economy. However, as the penetration levels of such intermittent sources achieve high values, the need for reserve generation to maintain sufficient power system reliability increases. Such reserve generation is comprised conventionally of high-emission gas-turbine based power plants. To this, consumer-centric smart grid services such as demand response (DR) are valuable alternates that could be exploited by system operators to alleviate the variability in generation, while offering other high-level benefits [1], [2]. In a deregulated scenario, multiple utilities manage the distribution sector in a geographical region, with some primarily serving residential loads. In such systems, the residents' demand flexibility is of high interest to the utilities, and is exploited using behavioral or incentive-based DR programs. While other types of DR implementations such as price-based or direct load control are possible, these have historically experienced several challenges that have inhibited their widespread acceptance [3]–[5]. Therefore, behavioral DR programs are likely to have a high penetration in the future grid [6].

One result of increasing consumers involvement in the system operation is that more and more communication streams between the utility and consumers affect grid control in real-time [7]. In most modern systems, the communication medium is predominantly textual [8]–[11]. Once a DR task is generated

by the utility, information about that task is sent via a DR messaging service to all the participating consumers. The participants then react to the message by changing their load consumption according to the task specified. From a systems perspective, in part due to this ever-permeating information communication, the risk of cyber-attacks increases [7], [12]. In particular, attackers can potentially influence the distribution system by manipulating the behavioral patterns of the consumers using this, or another external communication medium. This paper demonstrates using simulations that communications sent to residents through media such as SMS, email, or other social media platforms could lead to suboptimal demand patterns that compromise the system reliability.

Though proper use of DR could benefit the system [13], improper scheduling of DR events could lead to unexpected changes in the system load. As a result, significant effects could be felt on the system in the form of voltage limit violations, thermal loading of distribution lines, and higher stress on control devices such as tap-changing transformers [14]. Overall, the system reliability would be jeopardized, and therefore, such an attack should be of great concern to the utility. This concern about the grid reliability being affected by residential demand patterns is not without precedent. A recent study [15] found that uncoordinated charging of even small concentrations of plug-in electric vehicles (EVs) could result in a significantly altered system peak demand. Though flexible home appliances have a magnitude difference in their rating when compared to EVs, this study demonstrates that an attack through information injection considering only home appliances still produces a significant threat. Further, the fact remains that unlike EVs, utilities cannot explicitly regulate the use of common home appliances and hope to optimize the demand patterns. Additionally, although EV penetration may not reach very high levels anytime soon [16], DR penetration in homes could potentially reach high levels due to the certain availability of appliances such as air conditioners, water heaters, and clothes washers in most homes. Utilities, therefore, should be very concerned about potential manipulations of its consumers, particularly as more and more participants are inducted into demand flexibility programs. Illustration of this vulnerability is one of the contributions of this paper.

To quantify the impacts of an attack manipulating consumer behavior, an accurate model of the consumer response is essential. Though residential load modeling has been dealt with extensively in literature such as in [17], [18], there is a dearth of research regarding DR modeling. Additionally, to the best of the authors' knowledge, those DR models that do exist only target price-based [14] or market-based DR implementations [19]. To this end, a novel high-resolution bottom-up residential load model is developed in this paper, which explicitly models the residents' decision-making process after being notified of a DR event. The proposed model is therefore novel in its very scope of modeling consumer response to an event-based DR implementation, which is done based on factors such as the residents' varying degrees of participation, acceptance and follow-through rates for DR tasks received by them.

Finally, with the developed behavioral model, several hypothetical attacks are simulated wherein sections of the participants in a DR program are prevented from receiving legitimate messages, or fed false messages with the intent to distort their appliance usage patterns. With conservative reaction rates of consumers to DR events, it is demonstrated that an attacker could, unbeknownst to the utility, increase the daily system peak demand significantly. Utilities would therefore incur an excessive cost in the spot market to mitigate such an attack. Though the numerical results presented in this paper for the attack scenarios are specific to the case study considered here, the trends and vulnerabilities demonstrated hold good for stressed distribution systems. In the least, the attacker could nullify the economic benefits of the utility; at worst, the system may lose stability if the reserves become dangerously low, and result in cascading blackouts. The impact on reliability would be even worse in case of low-inertia microgrids or virtual power plants, whose penetrations are expected to increase in the future [20]. A strategic Stackelberg game is formulated to identify the equilibrium under various system parameters, simulations of which suggest that effective detection and countermeasures could possibly avoid, or nullify such attacks to a significant extent. Further, it is found that the system is resilient to most random attacks and errors, but remains highly vulnerable to strategic attacks.

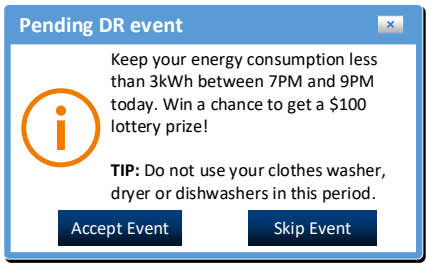
In summary, the contributions of this work are as follows: (i) it elicits the vulnerability of the distribution system to attacks that target residential customer behavior using communication such as texts (SMS), email, social media, etc., (ii) presents a novel customer behavioral model for event-based DR, (iii) develops probable strategic attack scenarios highlighting the above vulnerability, (iv) proposes a game-theoretic formulation of the above problem to analyze countermeasures and the resultant equilibrium, and finally, (v) analyses the error vs. attack tolerance of the distribution system under behavioral DR implementations. Importantly, the proposed modeling and impact analysis framework can easily incorporate more complex behavioral models if necessary.

This paper unfolds as follows. Section II describes the vulnerability of the consumers to external manipulations. The proposed consumer behavioral model for event-based DR is presented in Section III. Section IV describes probable strategic attack mechanisms, and presents simulation results for the same. It also discusses a game theoretic setting of this problem, along with an error vs. attack tolerance analysis. Section V concludes the work.

II. The Behavioral DR Infrastructure

The preferred [8], [9] and most effective [10], [11] mode of communication between the utility and consumer is through a messaging service, with the display device at the consumer end being a mobile phone, personal computer, or a pad. Text messaging could be implemented either using a dedicated communication network, or through the internet, with the consumer-end device running a utility-provided energy management application. The latter implementation has gained popularity with the internet-of-things revolution. Further, it moots the necessity for the utility to procure highly customized hardware and communication media. The utility thereby saves

(a)



(b)

Fig. 1. The DR messaging service. (a) Overview, and (b) A sample DR message generated by the utility, as seen by the participant. The reward for task completion in this case is a lottery system, which has been found to be more effective than a fixed reward scheme [9].

on installation and maintenance costs, and at the same time bolsters the willingness to and ease of participation for customers who may otherwise be averse to buying new display and communication devices. Such an architecture has been implemented in numerous smart-city projects (for e.g. in Mannheim, Germany [21]), and by utilities like Baltimore Gas and Electric Company, and Commonwealth Edison, USA, to name a few. A Swedish field trial, detailed in [9], presents a similar DR implementation for residential consumers using a lottery reward scheme. The latter study is used as basis for the application interface and message design shown in Fig. 1, which will be employed in this study.

The general sequence of actions that happen during a DR event are shown in Fig. 1(a). These are explained below:

1) Consumers receive a DR message (see Fig. 1(b)) from the utility that invites them to participate in an upcoming DR event, with the time, duration, and task specified.
2) The consumers decide whether to accept or reject the DR event request.
3) If the event is accepted, consumers take steps, either manually, or using an automated Home Energy Management System (HEMS) if one exists, to reduce consumption during the specified event period.

A. External manipulation of consumer behavior

The DR communication system, though simple and cost-effective for the utility, could potentially allow a malicious entity to hijack the user application and/or communication channels, or steal user credentials. Despite security measures for establishing confidentiality, integrity and authenticity of messages such as encryption and digital signatures being routinely employed, time and again attackers have successfully managed to impersonate legitimate information senders using mechanisms such as spear phishing and spoofing campaigns [22]–[25], and carry out man-in-the-middle attacks [26], [27]. Applications produced by prominent companies have been found to be vulnerable to the latter form of attack [28], and companies today continue to use such mechanisms to insert advertisements into third-party web pages [29].

Thus, the attacker could provide false information to the DR participants, or block all communications from the utility. Since power systems are a national security matter and any attack on the infrastructure could result in heavy losses to the utility and the nation, it is essential to analyze the possible consequences of such attacks. Sophisticated attacks would likely not be obvious to users who have no reason to suspect DR messages to be fake. This expectation is well supported by research in behavioral psychology [30] which predicts that people, normally lulled into a sense of cognitive ease, do not question the validity of information unless it is significantly different than those from previous events. Moreover, residents do not perceive a direct threat to themselves in accepting/rejecting DR events. Therefore, unlike other forms of cyber-attacks, where previous experience or history might make people cautious about the information they receive (e.g. phishing emails), in a DR scenario, residents normally do not receive any warnings regarding the possibility of an impostor injecting false messages into the system.

It is noted here that the above message delivery mechanism is only one possible implementation of a communication paradigm. Other implementations could be possible, and these still harbor the same vulnerability that is discussed here. Further, as will be explained later, an attack may be possible without hacking the DR messaging system at all. Therefore, the impact analysis presented in this paper is agnostic towards the design of the messaging interface, reward, and delivery mechanism.

III. The Proposed Event-Based Demand Response Model

The inputs of the proposed behavioral model are consumer attributes which will be detailed in the sequel, and its output the load profile for each residence in the system. In essence, this is a bottom-up model wherein the probabilities of use of individual home appliances at different time periods are determined based on time-use-surveys. These appliance-start probabilities are however altered if a DR event is requested by the utility and the consumer decides to participate in it. The proposed model is presented here as two parts: bottom-up load-curve generation when no DR exists, and the behavioral sub-model of the residents' response to the DR message.

A. Residential load profile generation sub-model with no DR

This sub-model, adopted from [17], generates the load profile for each residence based on a statistically accurate list of the home appliances present. Let the electrical appliances in a home be $\mathbf{A} = \{A_1, A_2, \cdots, A_n\}$. For any appliance A_i, let $\mathbf{P}_i = \{P_{1,i}, P_{2,i}, \cdots, P_{T,i}\}$ represent the respective starting probability for time steps $\{1, 2, \cdots, T\}$ in a day. Variations in the appliance list in different homes have already been incorporated into the starting probabilities [17]. To generate the daily load profile for this residence, a random number $rand$ is generated for each time step t and appliance A_i.

That appliance is then turned on if the generated number $rand \leq P_{t,i}$. This is represented mathematically below, where $x_{t,i}$ represents the state of the appliance.

$$x_{t,i} = \begin{cases} 1, & \text{if } rand \leq P_{t,i}. \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

This procedure is repeated for each appliance and time step to generate the daily load profile for that residence. The procedure for modification of the appliance-use probabilities during a DR event is now explained.

B. BAFT sub-model of consumer behavior for event-based DR

Consider a DR event requested in the interval $[t_{start}, t_{end}]$. As detailed in Section II, the consumer now has to make the following decisions sequentially: (i) believe that the received message is a legitimate DR event request sent by the utility, (ii) accept or reject the DR task, (iii) if accepted, follow through to complete the assigned task. As the reader will note, the first decision relating to the belief of the DR message is only relevant when the likelihood of false information injection is considered. The proposed DR model takes into cognizance each of the above decisions: *Believe*, *Accept* and *Follow-Through*, thereby lending the name BAFT to this model. Each of the above steps are now defined formally.

Belief: The propensity of the set of residents in a community to believe a message is defined as β, a fraction in the range $[0, 1]$. If $\beta = 1$, all DR participants believe the authenticity of a received DR notification, and if zero, none believe the message. Given this value for the community, the belief status of each individual participant, θ_b, is defined as:

$$\theta_b = \begin{cases} 1, & \text{if resident believes the DR message} \\ & \text{to be authentic.} \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Belief levels are generally high if the attacker is sophisticated enough to replicate the structure of the utility's DR messages with high fidelity.

Acceptance: It is natural that each resident or family has their own willingness to accept a DR request. This propensity to accept (θ_a) is defined for each consumer as the average number of times they would accept a DR task, if generated by the utility, to the total number of tasks generated by the utility. It is expressed as a fraction between 0 and 100%. Note that this parameter does not model the magnitude of demand reduction as a result of a DR task.

Follow-through: The degree of follow-through (θ_{ft}) of a participant is defined as the percentage of curtailment of their flexible load during a DR event. The value of this parameter varies from 0 to 100% and represents the efforts taken by that consumer to complete the DR task to the specified extent.

It is noted here that the propensities to accept and follow-through are considered as properties of each resident, while the degree of enrollment into the DR program (the DR penetration level) and the propensity to believe a DR message β are the attributes of the system as a whole.

Modeling the magnitude of an expected DR reduction: The proposed bottom-up model is accurate at the system level, meaning that the probabilities of usage of the various home appliances, combined with their average on-cycles already accounts for the variation in the set of appliances and stochastic usage behaviors across the various households (for further details, the reader may refer to [17]). This also means that, for DR implementation, each home has the same flexible load. Combining this observation with the fact that the utility requests the same reduction (usually in kWh) from each household in a locality, any DR request can be translated from 0% (meaning no reduction required) to 100% (full reduction of flexible load) for each home. This fraction of reduction, called for by the utility through the DR message, is defined as:

$$\theta_{util} = \frac{\text{Reduction requested (kWh)}}{\text{Event duration (h)} \times \text{Total flexible load (kW)}}. \quad (3)$$

For simplicity, in the above equation, it is assumed that the energy reduction throughout the DR interval is uniform. More complexity could be introduced into this parameter if required.

Having defined the above behavioral attributes of the residents, the net degree of compliance of a particular DR participant to an event is given by:

$$\theta = \theta_b \times \theta_a \times \theta_{ft}. \quad (4)$$

Algorithm 1 Generating residential load profiles using the proposed BAFT model for event-based DR.

Input: Appliance sets $\mathbf{A_{dr}}$ and $\mathbf{A}$, probabilities of use $\mathbf{P}_i$ for each appliance, DR penetration level, propensity to believe β for the system, propensity to accept and degree of follow-through θ_a and θ_{ft} for each resident.

1) Translate the DR message into θ_{util} using (3).
2) Based on the DR penetration level and β, choose DR participants and assign values for θ_b for each participant.
3) Assign values of θ_a and θ_{ft} for each participant based on their respective distributions.
4) Determine final probabilities of use of home appliances:
   a) For residents participating in the DR program, calculate new probability vectors $P^*_{t,i}$ for flexible appliances $\mathbf{A_{dr}} \subset \mathbf{A}$ using (4)-(7).
   b) For non-participants, retain the original probability vectors $\{\mathbf{P}_i \mid i = 1 : n\}$.
5) Based on the final starting probabilities, generate net load profile of each home using (1).

Output: The daily load profile for each residence.

C. Combining the two sub-models

Consider the set of DR-capable appliances $\mathbf{A_{dr}} \subset \mathbf{A}$. During the duration of a DR event, the resident, on accepting the DR task, would either defer (or advance) the usage of such flexible appliances out of the event period. Accordingly, their probabilities of use reduce during the event period:

$$P^*_{t,i} = P_{t,i}\ (1 - \theta\ \theta_{util}), \forall t \in [t_{start}, t_{end}], A_i \in \mathbf{A_{dr}}, \quad (5)$$

$$P_{deferred,i} = \sum_{t=t_{start}}^{t_{end}} (P_{t,i} - P^*_{t,i}). \quad (6)$$

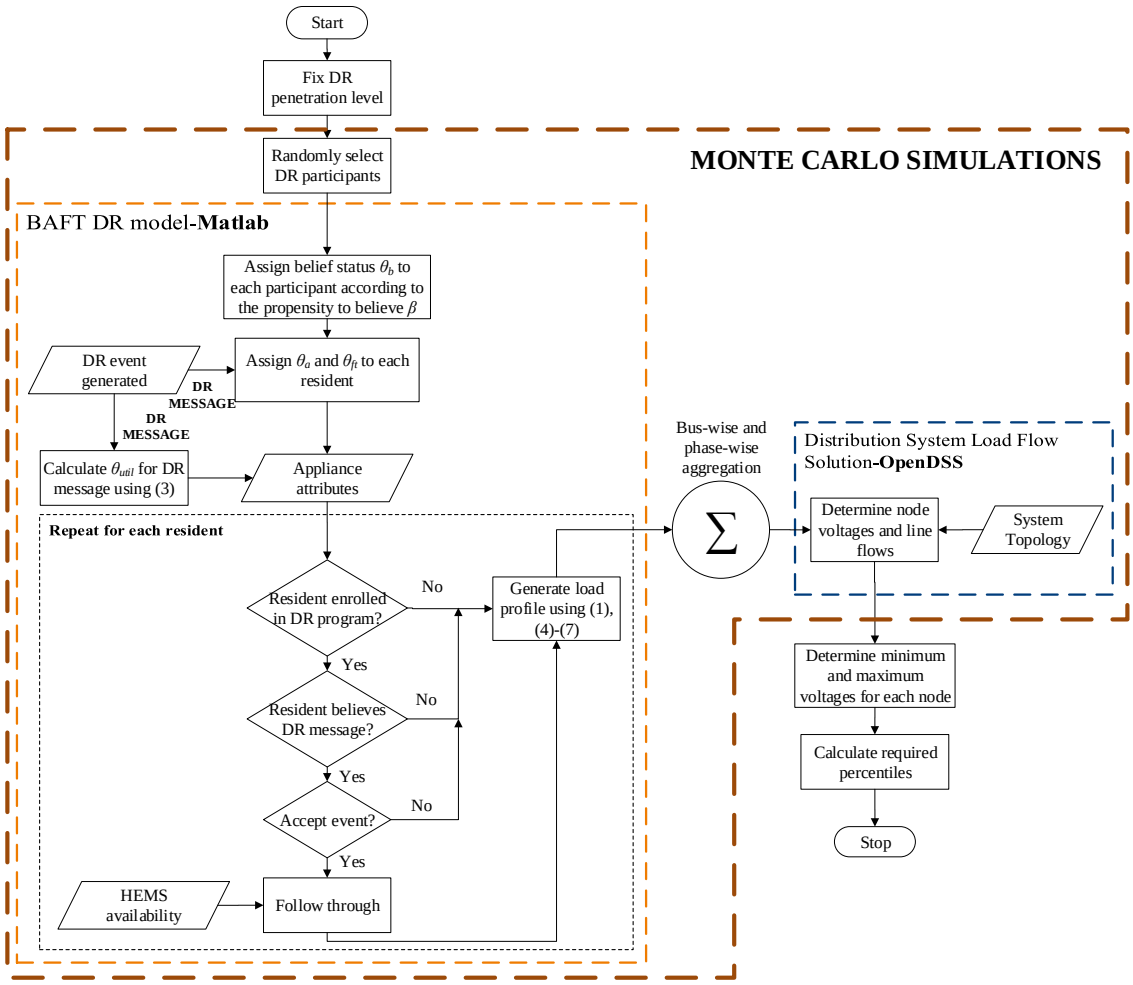


Fig. 2. The proposed co-simulation approach for assessing the impact of an attack. The orange (inner) box shows the BAFT behavioral model.

Here, $P^*_{t,i}$ is the reduced probability of usage of an appliance A_i during the DR event period. Note that θ is scaled by θ_{util} to reflect the magnitude of the called-for response on the actual consumer demand. $P_{deferred,i}$ refers to the deferred probability of usage, which is then added randomly to the original probability of use at an hour $t_r \notin [t_{start}, t_{end}]$ outside the event period:

$$P^*_{t_r,i} = P_{t_r,i} + P_{deferred,i}. \tag{7}$$

This hour t_r is chosen from the reschedulable period with a probability that is proportional to the original probability of use of the appliance at that time, $P_{t_r,i}$. Realistically, the starting limit t_{start} could be set at, say, 6AM. Note that the usage probabilities of the inflexible appliances $A_i \notin \mathbf{A_{dr}}$, i.e., those that are not DR-capable, are unaffected. The above model is summarized in Algorithm 1. The parameters of this model may depend on several important factors such as the compensation offered by the utility for the DR event, and the presence of HEMS that automatically schedule flexible loads according to consumer preferences. The first could be modeled by scaling the values of θ_a and θ_{ft} appropriately using existing economic models. Secondly, if an HEMS system is present, the proposed model is modified by assuming that the propensity to follow-through θ_{ft} for that home is a constant. It is sensible to assume a high value for this parameter, say 95%.

IV. Attack Mechanisms-A Case Study

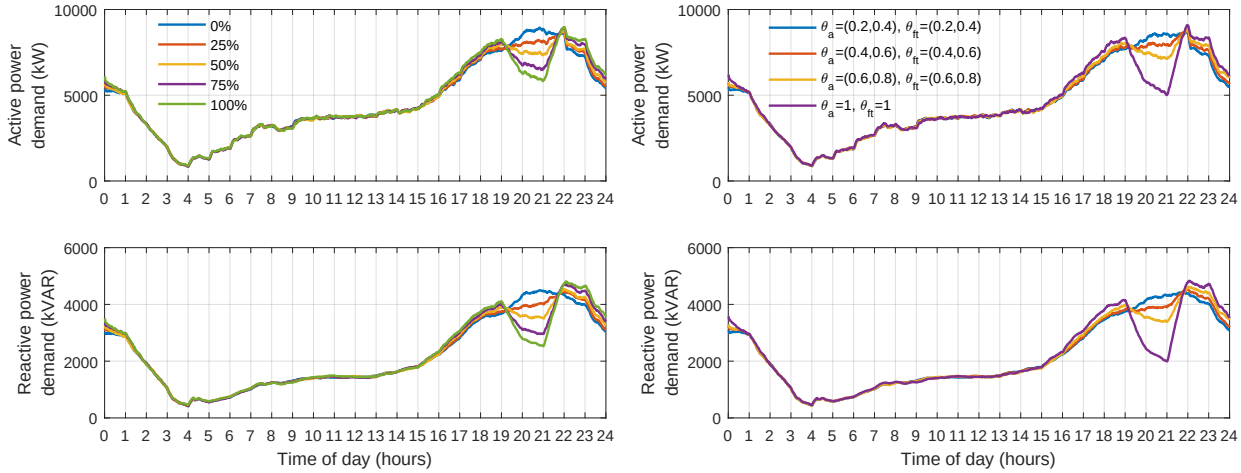
The overall flowchart for determining the network impact of an attack is shown in Fig. 2. The penetration level of DR in the distribution system is defined as:

$$\text{DR Penetration Level} = \frac{\text{Residents enrolled in DR program}}{\text{Total number of residents}}. \tag{8}$$

The behavioral model developed in Section III is simulated using MATLAB, and a co-simulation approach is adopted here with the power flow solution being obtained by OpenDSS.

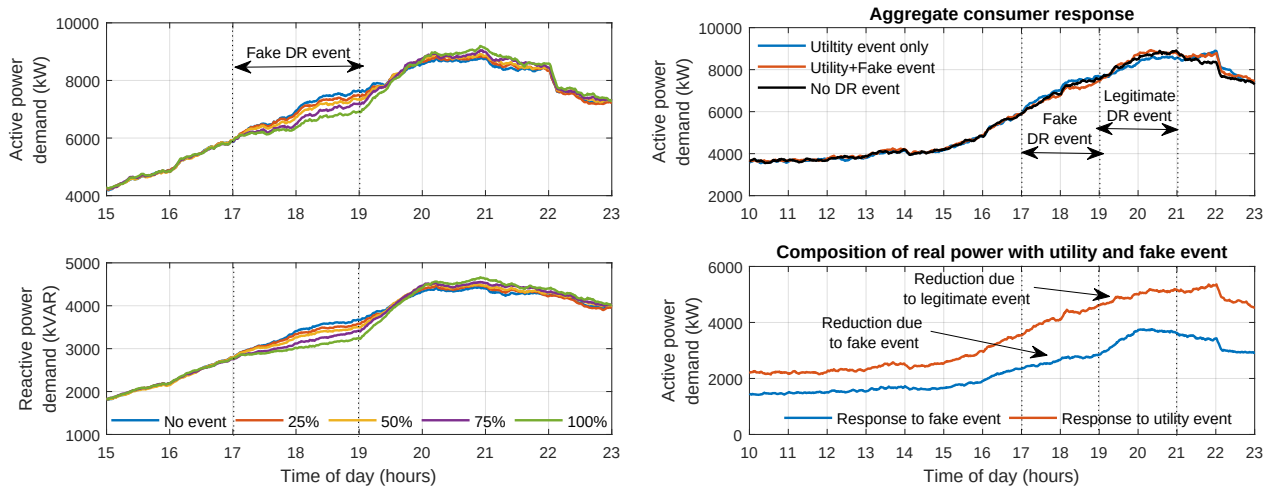
A. System Description

The standard IEEE 123 node test feeder [31] is considered as the topology on which the test residential system is based. This system is heavily loaded to mimic stressed legacy systems [2], such as those in North America and Europe that typically employ DR to improve their reliability. It is assumed that each of the spot loads on this network is comprised of multiple residences, with 2094 homes in total distributed over the various nodes. Load curves for this system generated for various DR penetration levels, acceptance and follow-through rates are provided in Fig. 3 as an illustration. Appliance specifications are presented in the supplementary file.

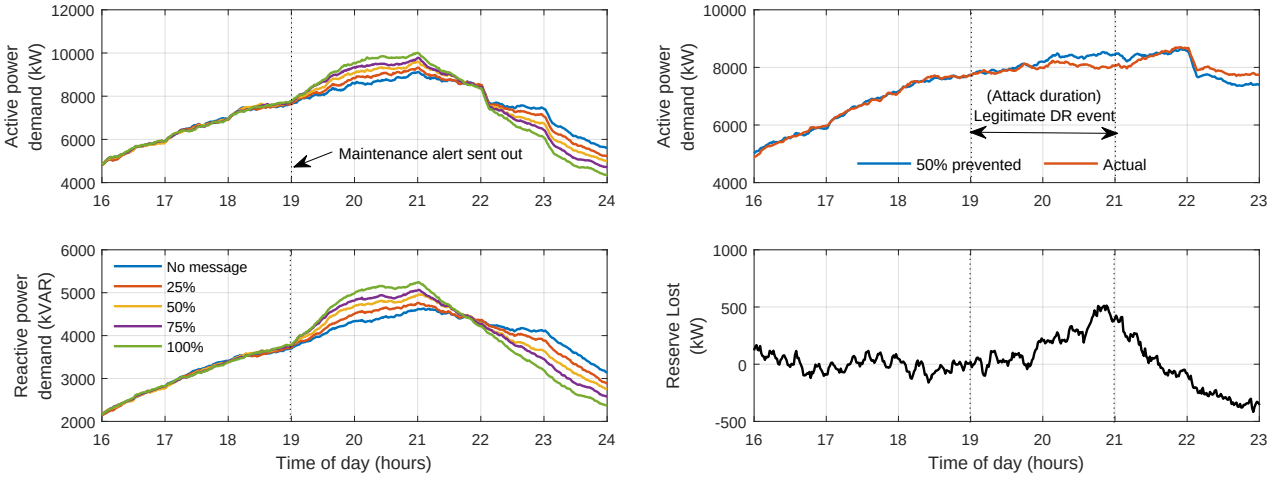


(a) (b)

Fig. 3. Illustrating the BAFT DR model: (a) Aggregated real and reactive power profiles for varying DR penetration. The DR event lasts from 7-9PM. $\theta_b = 1$, with θ_a and θ_{ft} uniformly distributed from 0.8 to 1.0. (b) Effect of θ_a and θ_{ft} uniformly distributed between various ranges, with DR penetration level = 100%.



(a) (b)



(c) (d)

Fig. 4. Simulations showing an instance of the system load profiles for (a) Attack-1: False DR event scheduled just before the peak period, for various percentages of total participants receiving the false DR message. (b) Attack-2: False DR event scheduled from 5-7PM for 40% of the participants just before a legitimate DR event scheduled by the utility 7-9PM. (c) Attack-3: False maintenance shutdown alert sent to all participants at 7PM, with various percentages acting on the message. (d) Attack-4: 50% of the participants are prevented from receiving the utility DR request between 7-9PM.

B. Strategic attack scenarios

Several possible strategically-planned attacks are now simulated, and their effects on the distribution system analyzed. These attacks target the main properties of a DR event: the event timing and duration, and the set of consumers that respond to the task. The list of attacks given here, by no means exhaustive, is selected to represent most attacks that may be possible in reality.

Consumer response parameters: According to field-trials conducted previously for behavioral DR [9], typical event acceptance and follow-through rates range from 45-66% and 30-64% respectively. This paper considers the same limits, with the actual values of θ_a and θ_{ft} for the residents uniformly distributed in their respective ranges.

1) Attack-1: Creating a fake DR event just before, or after the daily peak period: In this scenario, the strategic attacker schedules a fake DR event involving some or all enrolled participants, with the purported event ending just before, or starting just after the peak loading period during the day. The utility itself does not send a (legitimate) DR signal to anyone.

The peak period occurs typically in the evening, from about 8PM and lasts until about 10PM. A typical false message could be: "*Please do not consume more than 3kWh tomorrow between 5-7PM. Doing so, earn a FREE lottery ticket! TIP: For best results, just schedule your washing machines, dryers and dishwashers to work between 7-9PM.*". This message encourages residents to defer their flexible appliances away from the stated event period of 5-7PM, right into the period when the system demand is normally at its peak. The resulting demand for our case study is shown in Fig. 4(a). The overshoots after 8PM are clearly visible. This would deteriorate the system reserves, may cause overloading, and also affect the voltage profile, as shown in Fig. 5. The boxplots shown correspond to 100 trials, while the minimum voltages were obtained as a result of 1000 Monte-Carlo iterations for one set of load profiles, using the 5th percentile of the minimum node voltages.

This attack would also impact the performance of tap-changing regulators in the system, which would have to respond to the sudden change in the system voltage. Prolonged operation of such transformers under such high loading conditions would increase the number of switching cycles, and thereby deteriorating their life. Note here that the system under consideration is a radial network with a circular path, and hence a load change has a smaller effect on the node voltages. The effect would be worse downstream had there been only one path feeding these nodes, which is true in many legacy systems.

Fig. 6 shows the effect of the residents propensity to believe a message and the DR penetration level on the system maximum demand (averaged over 100 trials), assuming all the participants receive the fake message. As one might expect, when more and more participants believe in the attacker's message, the more the effect is on the network. For the system at hand, with just 50% of recipients believing the fake message and conservative estimates of acceptance and follow-through rates, the attacker could alter the system daily peak demand by more than 2% when the DR penetration level is 70%. Note that this sudden change in peak demand is comparable to its normal growth over several months, and is quite significant to the utility.

2) Attack-2: Tailgating a real DR event: When the utility schedules a legitimate DR event, the attacker sets up a fake event just before the event period for a section of the participants. The resulting overshoot in demand due to the fake event would nullify or reduce the effect of the consumer response to the actual event. Such a system load curve is shown in Fig. 4(b) for a legitimate DR event scheduled between 7-9PM, and 40% of the residents receiving a fake DR message indicating an event from 5-7PM.

3) Attack-3: Faking maintenance shutdown alerts: Another possible mode of attack could be disseminating fake messages urging users to consume their loads at periods of high stresses by suggesting that a maintenance shutdown, or load-shedding event was imminent. An instance of such a message could be: *"Maintenance activities could affect electricity supply between 9PM and 12AM; try to use any appliances before the maintenance period."*. Spreading of such alerts could be easily possible through diverse media such as SMS, email, social media, etc. This would trigger a panic as users turn on essential equipment such as domestic water pumps, clothes washers and dryers, or pre-cool their homes. Fig. 4(c) shows simulation results where all participants receive the above alert at 7PM, with varying percentages of residents acting on it. This simulation assumes $\theta_a = 1$ and $\theta_{ft} = 50\%$ for all residents who believe the alert to be true and act on it. The resulting sudden demand increases could lead to voltage sags and overloading of feeders that may potentially trigger cascaded blackouts, especially in already-stressed systems [32].

4) Attack-4: Intercepting legitimate DR messages: The attacker does not allow a section of the participants to receive a legitimate DR message sent out by the utility. Alternatively, the attacker may send a fake message declaring that the said event was canceled by the utility. The attacker may further send false acceptance notifications on behalf of multiple participants to the utility to prevent detection. Consequently, the affected participants do not perform the required load reductions.

This attack, if executed during high-demand periods, could result in low reserve levels, thereby requiring the utility to spend large amounts of money to procure emergency reserves in the real-time market. An example of this type of attack is shown in Fig. 4(d), where 50% of the users were prevented from receiving the DR message. The DR event generated by the utility lasts from 7-9PM. It is assumed that DR penetration level is unity and all residents could potentially react to a DR message under normal circumstances (no attack). Note that in reality, reserve levels do vary continuously within a short range to compensate for changing demands and generation. However, such an attack would cause a sudden and large reduction, which requires immediate action by the utility. A summary of the impact of this attack on the reserves is presented in Fig. 7 for various sizes of the affected consumer group. Due to the randomness involved in the simulation process, box-plots are shown over 100 trials.

In these simulations, the assumption is that there is no use of home automation systems by any residents. This assumption

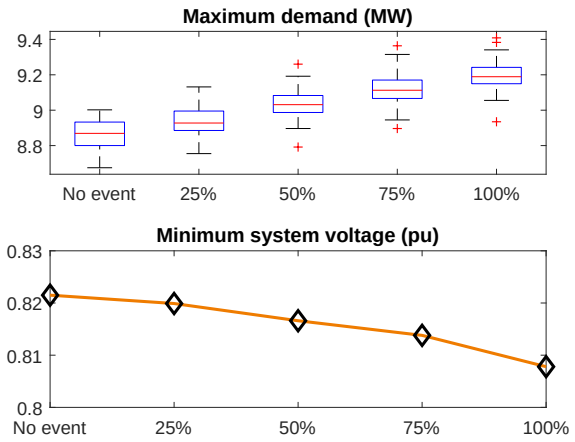


Fig. 5. Impact of the extent of attack-1 on the daily peak system demand and minimum voltage for 100% DR penetration.

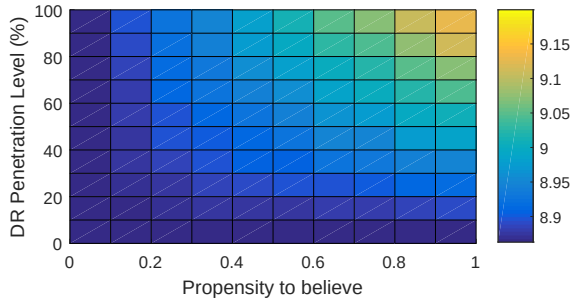


Fig. 6. Variation of peak aggregate system demand (MW) with the community's propensity to believe a fake message, and DR penetration level for attack-1.

would only make the shown results more conservative; if more and more such devices are implemented in residences, the magnitude of demand reduction would only increase for those receiving the message, and hence increase the disparity between the curves for the two cases shown in Fig. 4(d).

C. Strategic utility response to an attack: a game-theoretic analysis

It has been heretofore assumed that the attacker is strategic, but the power utility is unaware of the attack under progress and consequently does not take any defensive actions. A game theoretic model is now developed to analyze utility countermeasures and the resulting equilibrium. For brevity, only attack scenario-1 is considered for this analysis; this model can be easily extended for the other scenarios described in the subsection IV-B. It is noted here that the term 'power utility'

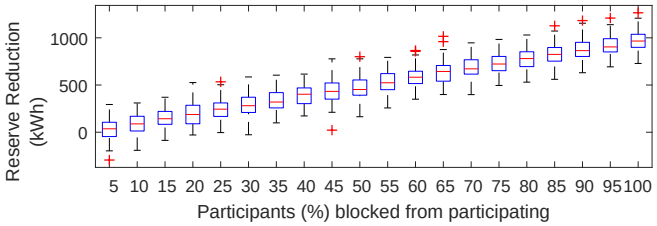


Fig. 7. Box-plot showing the impact of attack-4 with varying number of users prevented from receiving DR messages, over 100 trials. Reserve reduction is calculated for the event period.

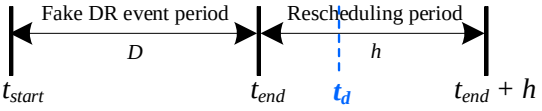


Fig. 8. False DR event specifications.

is substituted hereon within this subsection by 'defender' to avoid any conflict with the traditional use of the term utility to mean 'payoff' in a game-theoretic setting.

Assume that the defender becomes aware of an attack (details of which include the duration and suggested rescheduling period referring to Fig. 8) after a time of detection t_d such that $t_{end} \leq t_d \leq t_{end} + h$. If $t_d < t_{end}$, it is assumed that $t_d = t_{end}$. This assumption refers to the fact that the attack impact on the peak demand is only felt during the rescheduling period; detection before this period starts (i.e., $t = t_{end}$) is sufficient to potentially nullify this effect. Any earlier detection, therefore, does not add extra value. The exact mechanism of attack detection is not considered to be in scope of this discussion.

The defender would, after detection, respond to nullify the attack impact with the knowledge of the attackers strategy, and therefore, this is modeled as a two-player non-cooperative Stackelberg game with the attacker being the leader and the defender the follower.

1) Strategy spaces: The possible strategy spaces $\mathbf{S^A} = \{s_j^A\}$ and $\mathbf{S^D} = \{s_k^D\}$ respectively for the attacker and the defender are as follows.

Attacker: All possible times of the purported DR event, considered for simplicity in hourly intervals. Let the event duration be D and the suggested rescheduling period for any load deferred be the h-hour period immediately after the purported DR event (see Fig. 8), to make it easier for consumers to act. Therefore, each attack strategy would be $s_j^A = \{t_{start}, D, h\}$. It is important to note that another possible attacker strategy is to not attack at all. Finally, the set $\mathbf{S^A}$ is given by $\{\text{'No Attack'}, 1, 2, \cdots, 24\}$, assuming fixed values of D and h as per Fig. 8. As the false event's duration increases, the potential attack impact increases; however, the believability of the DR message decreases (it would be impossible to have all participants turn off their appliances for extended periods of time). It is assumed that all residents receive this message.

Defender: The distribution company broadcasts a countering message urging participants to not consume according to the suggestions of the malicious message received by them earlier. It is necessary to not confine this response to only those consumers receiving the fake message because first, identification of this consumer set may not be possible, and second, some consumers may only respond to the first message and not be willing or able to respond to the defender's counter-message. That said, any countermeasures must be designed so as to cause minimal disruption to the customers, and hence the defender only sends the counter-message to a fraction χ of the entire population. Therefore, the defender strategy is defined as $s_k^D = \{\chi\}$, and $\mathbf{S^D} = \{s_k^D\} = \{0, 0.05, \cdots, 1.0\}$.

2) Payoffs: If 'No Attack' is the strategy chosen by the attacker, the payoffs to both the defender and attacker are assumed to be zero. Payoffs for the other strategy combinations are now detailed. If an attacker succeeds in carrying

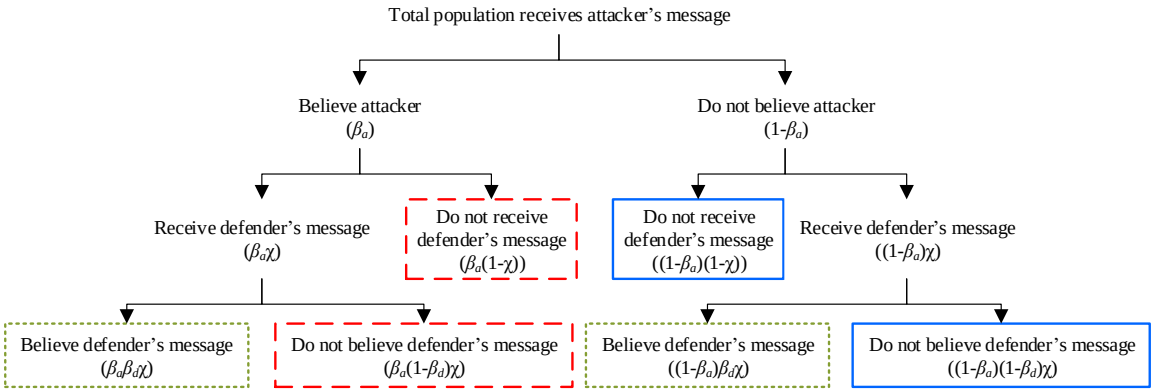


Fig. 9. Classification of consumer responses to the attacker's and defender's messages. Dashed/red, dotted/green, and solid/blue boxes respectively refer to response to attacker's message, response to defender's message, and zero response. The value within the parentheses for each category represents the ratio of the total consumers within that category.

out an undetected attack (alternatively $t_d \geq t_{end} + h$), no defender message is sent out. The attacker's payoff ϕ_a depends on the system impact, which in this case, is the increase in the peak system demand due to the deferred load. This could be approximately transformed into a function of the time of attack, its duration, the number of consumers that respond to it, and the believability (β_a) of the attacker's message:

$$\text{System Impact} \approx \sum_{\substack{\beta_a \\ consumers}} \theta_a \theta_{ft} \sum_{A_i \in \mathbf{A_{dr}}} \mathcal{P}_i \sum_{t \in D} p_{t,i}. \tag{9}$$

Note that the believability of the message reflects the fraction of the population that responds to it, which is β_a here. $\mathcal{P}_i$ refers to the real power rating of appliance A_i.

If however, the defender detects the attack at a time $t_d < t_{end} + h$, the defender receives a reward, while the attacker accrues a negative payoff. The payoff to the defender depends on t_d, the consumer response to its counter message (depending in part on its believability β_d), as well as the potential impact of the attack if it were undetected. The payoffs to the defender and attacker are respectively defined as

$$\phi_d = L - M - K\chi \text{ and } \phi_a = M - L - K, \tag{10}$$

where L and M are defined in (11), displayed at the bottom of this page. Referring to (11), note that the term L increases if the attack is detected early on. $\eta > 1$ models the fatigue and the resultant reduction in the consumers' acceptance and follow through rates on receiving multiple communications. Referring to Fig. 9, in total, a $\beta_d\chi$ fraction of residents believe and respond to the defender's message (this response is quantified as L), while a $\beta_a(1-\beta_d\chi)$ fraction respond to the attacker's message and do not receive/believe the defender's message (this response is quantified as M). K is the cost incurred by the attacker in sending the messages to the total population. The term $K\chi$ in (10) is used to incentivize the defender to disturb as small a consumer group as possible to achieve a counter-response to the attack.

3) Stackelberg equilibrium: Say the attacker plays a strategy s_j^A with a probability $P^A(s_j^A)$. In response, the defender plays a strategy s_k^D with a probability $P^D(s_k^D)$. The Stackelberg equilibrium is obtained by solving the following mixed-integer optimization problem which maximizes the attacker's expected payoff:

$$\max \sum_{s_k^D \in \mathbf{S^D}} \sum_{s_j^A \in \mathbf{S^A}} P^A(s_j^A)\ P^D(s_k^D)\ \phi_a(s_j^A, s_k^D) \tag{12a}$$

$$\text{s.t.} \sum_{s_j^A \in \mathbf{S^A}} P^A(s_j^A) = 1, \tag{12b}$$

$$\sum_{s_k^D \in \mathbf{S^D}} P^D(s_k^D) = 1, \tag{12c}$$

$$P^A(s_j^A) \in [0,1],\ \forall s_j^A \in \mathbf{S^A},\ \text{and} \tag{12d}$$

$$P^D(s_k^D) \in \{0,1\},\ \forall s_k^D \in \mathbf{S^D}. \tag{12e}$$

The attacker's payoff ϕ_a is obtained from (10) and (11). Constraints (12b) and (12d) refer to the attacker's (who is the leader in this game) mixed strategy, while the defender (follower) only plays her best pure strategy after observing the actions of the attacker, as indicated by (12c) and (12e). The above quadratic problem could be simplified by using the DOBSS approach [33], or using the simpler, albeit more computationally expensive approach presented in [34]. This problem is sufficiently simple for the latter to solve in a few seconds, and is hence adopted here. The game is now solved for the system under study with D=h=2 hours, and η = 1.1 for all consumers. For the reader's convenience, we repeat here the set of attacker's strategies: $\mathbf{S^A} = \{\text{'No Attack'}, 1, 2, \cdots, 24\}$, which consists of 25 possible pure strategies.

Simulations indicate that as long as an attack is detected before the start of the rescheduling period ($t_d < t_{end} + h$) and if β_d=1, it is possible to deter even the most sophisticated attacker (with β_a=1) from attacking at all (pure strategy 1),

$$L = \sum_{\substack{\beta_d\chi \\ consumers}} \frac{\theta_a \theta_{ft}}{\eta} \sum_{A_i \in \mathbf{A_{dr}}} \mathcal{P}_i \left(\frac{\sum_{t=t_d}^{t_{end}+h} p_{t,i} - \sum_{t=t_{end}}^{t_d} p_{t,i}}{\sum_{t=t_{end}}^{t_{end}+h} p_{t,i}} \sum_{t \in D} p_{t,i} \right),\ M = \sum_{\substack{\beta_a(1-\beta_d\chi) \\ consumers}} \theta_a \theta_{ft} \sum_{A_i \in \mathbf{A_{dr}}} \mathcal{P}_i \sum_{t \in D} p_{t,i}. \tag{11}$$

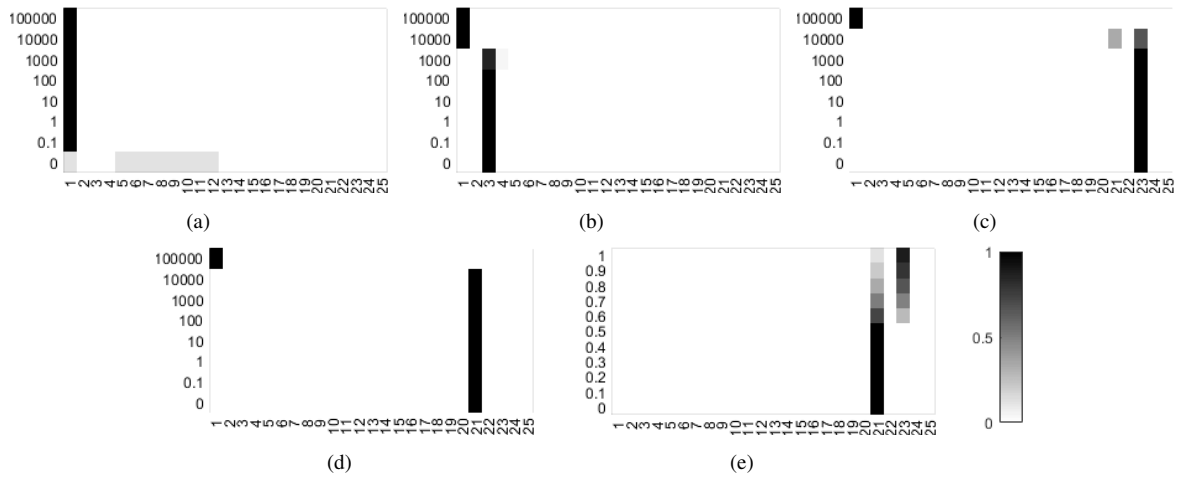


Fig. 10. Attacker's mixed strategy under Stackelberg equilibrium: Immediate detection with (a) β_d=1, β_a=1 and (b) β_d=0.8, β_a=0.7. (c) Late detection with β_d=0.8, β_a=0.7. (d) No detection with β_d=0.8, β_a=0.7. (e) Impact of β_d on the attacker's mixed strategy, given β_a=0.7 and $K = 1 \times 10^4$. X-axis represents the various attacker's pure strategies for (a)-(e). Y-axis represents K for (a)-(d), and β_d for (e).

whatever be the cost K. This is shown in Fig. 10(a), which depicts a heatmap of the attackers equilibrium mixed strategy for varying values of K. Note that when K=0, the attacker, with equal likelihood, attacks between 4AM-11AM (when the available deferrable load and therefore system impact are zero), or does not attack at all. Correspondingly, the defender under equilibrium does not send any counter messages ($\chi = 0$); the very threat of detection and full counter-measures is an effective deterrent to the attacker. However, achieving unity β_d and β_a may not be practically possible. Fig. 10(b) illustrates the equilibrium when β_d=0.8 and β_a=0.7. In this case, an attack does indeed occur for lower values of K, but away from the peak period and therefore, the impact will be low.

Importantly, note that immediate attack detection may not always be possible. The equilibrium significantly changes when the attack is detected, say, 1 hour into the rescheduling period; the heat map for this case with $\beta_d = 0.8$ and $\beta_a = 0.7$ is presented in Fig. 10(c). The equilibrium defender response for the cases with $K \leq 1000$ is $\chi = 1$, and $\chi = 0$ for the higher values of K. Clearly, this result is not favorable from the defenders perspective: until the cost of messaging becomes prohibitively high, the attacker always attacks around the peak demand period. Further, referring to Fig. 10(d), when the attack is not detected at all ($t_d \geq t_{end} + h$), as long as the cost of the attack is reasonable, the attacker always attacks at 8PM (strategy 21) so as to inflict maximum damage.

To summarize, analysis of the Stackelberg equilibrium underscores the need for early detection of an attack. If the defender possesses fast detection capability, and effective counter-messages can be broadcast to the population, any potential attackers can be dissuaded. However, the impact of the attack, when detected late, depends on the response of the utility. Referring to Fig. 10(e) which corresponds to a cost of $K = 1 \times 10^4$ and $\beta_a = 0.7$, highly effective counter-measures can succeed in at least deterring the attack away from periods of peak demand in such cases.

D. Error and attack tolerance of behavioral DR programs

The previous subsections dealt solely with strategically planned attacks that manipulate consumer behavior. It is also possible that random failures and/or human error could change or prevent the delivery of DR messages. This subsection compares the effects of such events vis-à-vis strategic attacks.

In the context of this paper, an error refers to the modification of the contents of a DR message due to typographical/human errors or its non-delivery due to random technical failures. Strategic attacks, to the contrary, manipulate the message so as to maximize the detrimental impact on the electrical network. Specific definitions for these are presented in Table I. Attacks-1 and 2 are both based on the premise of creating false DR events near peak demand periods. In the interest of brevity, this analysis only considers the former, which can be extended for the latter scenario as well. Further, for simplicity, it is assumed that all affected participants, if any, receive the same erroneous/maliciously-drafted message in these scenarios.

Simulation results describing the effects of these errors and strategic attacks are presented in Fig. 11. Clearly, the timing of attacks 1-3 is of high importance in determining the resultant system effects, and therefore, the system is quite resilient to random errors of these kinds. To the contrary, for the last scenario, there is no distinguishing the system-level effects of an attack and random failures. However, in scenario-4, random attacks could potentially be detected earlier if a large section of the participants do not respond to (accept/reject) the utility-created DR task.

These observations can be explained as follows. Under normal circumstances, the appliance-use patterns of the various households are random, and this demand staggering is essential for system planning and operation. However, problems arise when this diversity is lost, or reduced by communications sent to the community. On the one hand, the system is quite resilient to random mistakes in the messages (except in scenario-4) which only affect a small section of the population and at arbitrary times, thereby preserving the original diversity of the electricity consumption. But on the other hand, strategic attacks impact the load consumption diversity of a large number of households, particularly at times when the system is already under stress, thereby impacting the system reliability significantly. It is also noteworthy that the impact of attacks 1

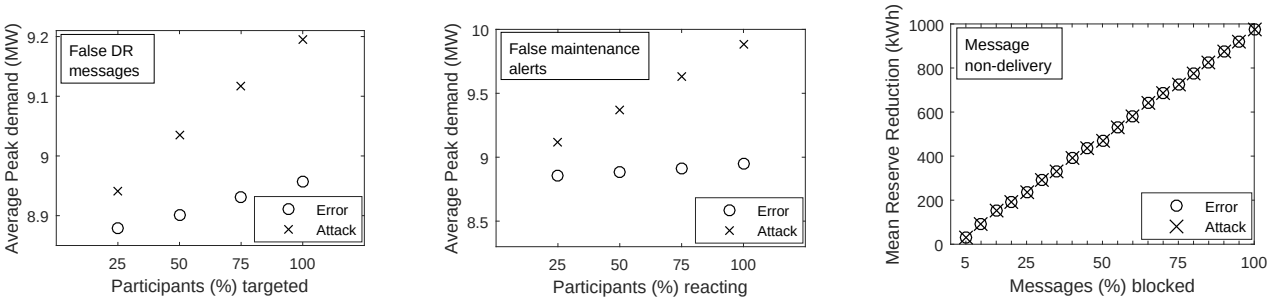


Fig. 11. Error and attack impacts of behavioral DR programs on the distribution system.

TABLE I
ERROR AND ATTACK DEFINITIONS

Attack scenario	Error Definition	Attack Definition
1 and 2	False DR event created at a randomly-chosen time in the day for each consumer.	False event created immediately before the peak demand period.
3	Alerts sent out indicating random times for maintenance and possible consumption.	Alert urges consumption during the peak demand period.
4	Non-delivery of a message to some participants due to random technical failures.	Deliberate interception of a message while possibly returning false acceptance notifications to the utility on behalf of the consumer.

and 2 (false DR messages) is expected to be lower than that of attack-3 (false maintenance alerts); this reflects the difference in the behavior of the consumers during what they perceive is business-as-usual for a DR event vs. a possible blackout, which is an abnormal and possibly more impactful occurrence.

Further, regarding scenarios 1-3, acting on a DR message requires residents to be at home to respond. If random errors result in DR tasks requiring action during working hours, many residents may not be at home, thereby reducing the response significantly. To the contrary, a strategic attacker, as shown in the previous section, targets the evening hours when most residents are likely to be able to respond to her signal. The results shown in Fig. 11 assume that residents respond at all times during the random event periods generated. Even so, the impact of a random failure/attack is very small when compared to that of a strategic attack.

E. Discussion

Though the exact magnitudes of the network effects presented in this paper are system specific, the trends illustrated are valid for stressed distribution systems. The goal here is to demonstrate that adverse system impacts could be caused by an attacker manipulating consumer behaviors through text messages. The most benign of the potential impacts of an attack would be a reduction in the reserve and loss of economic benefit for the utility, and in the worst case, changes in the peak demand could result in voltage instability, controlled load shedding, or rolling blackouts [32]. Moreover, residential systems see less utility investment into inverter-based or other voltage-control equipment as their commercial and industrial counterparts, and therefore, voltage problems cannot be mitigated as easily in these systems. Also note that with the advent of new business models, more and more distribution systems are being operated as a collection of low-inertia microgrids or virtual power plants. Sudden loss of reserves would have a worse effect in these cases and could easily lead to the loss of voltage and frequency stability.

The analysis presented in Subsection IV-B shows that the effect of increasing the users propensity to accept and follow-through for tasks at a given DR penetration level is similar to increasing the DR penetration levels given constant acceptance and follow-through rates. For simplicity, in the results presented in this paper, only the latter case is shown, with conservative but constant values for acceptance and follow-through. In the near future, as targeted enrollment drives for behavioral DR programs are undertaken by utilities, these values will move closer to 100%, and the effects of an attack would only become worse. Note that the values considered in this work represent the status quo in a real residential community [9] and are therefore grounded in reality.

It is noteworthy that air-conditioning loads and EVs have not been considered to be a part of the flexible load in these simulations. While these can be incorporated easily, it has been shown here that system performance deteriorates significantly even without these heavy loads. Addition of these will only increase the impact of an attack. Results presented in [15] demonstrating the increase in the system peak demand due to EV charging, support these conclusions.

It is important to highlight the effect of the time-scale of event scheduling on the vulnerability described in this paper. A real-time DR implementation, with a few hours notice to the users, may be easier for the attacker to cause sudden and undetectable attacks, whereas, in a day-ahead scenario, the utility has more time to detect and foil them.

Following the analysis in Section IV-C, possible utility actions to tackle this vulnerability are as follows. Firstly, consumer-centric services must be analyzed to determine hardware and software vulnerabilities that may provide an opening for an information-injection attack. Recognizing the possibility and potential impacts of an attack, in case an attacker does manage to inject false messages or hijack the messaging system, utilities require a real-time situational awareness tool that monitors load patterns and identifies potential attacks underway. This tool would need to incorporate accept/reject feedback from users to identify suspicious activity, and development of this tool is our current focus.

V. Conclusion

Residential demand response programs are expected to achieve high penetrations in the future distribution grid. In such a scenario, hitherto unforeseen vulnerabilities with consumers as the focal point arise. Threat identification is the first step to securing a system, and in that spirit, this paper posits that risk assessment of the grid must analyze the impact of manipulation of consumer behaviors on the system reliability.

This study has analyzed several possible mechanisms of a malicious attack targeted at residential consumers through messaging. Having modeled the response of consumers to a DR event using a bottom-up probabilistic technique, the impact of such an attack on the system load, reserves, and voltage is ascertained. It is shown that manipulating consumers with high enthusiasm levels to respond fully to DR signals could cause adverse network effects at high DR penetration levels. While utilities today strive to safeguard their communication infrastructure against cyber attacks, as explained in this paper, direct external manipulation of the consumers with misinformation is easily possible. Therefore, appropriate countermeasures such as immediate reporting and feedback mechanisms, and anomaly detection systems monitoring the consumer demand should be incorporated in their demand response management systems.

This study could be extended in several possible directions. Firstly, the present work does not model social interactions between multiple individuals in the community, which could affect the information flow and hence the attack impact. Second, the behavioral model could be made more sophisticated to account for multiple social-economic and demographic classes in the community, and the differences in their behaviors. Finally, from the utility perspective, machine learning techniques could be developed to enable them to detect anomalies in consumer behaviors to identify and counteract any attacks underway.

References

[1] P. D. Lund, J. Lindgren, J. Mikkola, and J. Salpakari, "Review of energy system flexibility measures to enable high levels of variable renewable electricity," *Ren. Sust. Energy Reviews*, vol. 45, pp. 785–807, 2015.

[2] P. Cappers, J. MacDonald, J. Page, J. Potter, and E. Stewart, "Future opportunities and challenges with using demand response as a resource in distribution system operation and planning activities," Lawrence Berkeley National Lab, CA (USA), Tech. Rep., 2016.

[3] W. Yang, T. C. T. Ho, L. Xiang, C. C. Chai, and R. Yu, "An overview and evaluation on demand response program in singapore electricity market," in *IEEE Conf. Energy Convers. (CENCON)*, 2014, pp. 61–66.

[4] H. T. Haider, O. H. See, and W. Elmenreich, "A review of residential demand response of smart grid," *Ren. Sust. Energy Reviews*, vol. 59, pp. 166–178, 2016.

[5] R. Deng, Z. Yang, M.-Y. Chow, and J. Chen, "A survey on demand response in smart grids: Mathematical models and approaches," *IEEE Trans. Ind. Inf.*, vol. 11, no. 3, pp. 570–582, 2015.

[6] G. Schuitema, L. Ryan, and C. Aravena, "The consumer's role in flexible energy systems: An interdisciplinary approach to changing consumers' behavior," *IEEE Power Energy Mag.*, vol. 15, no. 1, pp. 53–60, 2017.

[7] J. Lopez, J. E. Rubio, and C. Alcaraz, "A resilient architecture for the smart grid," *IEEE Trans. Ind. Inf.*, vol. 14, no. 8, pp. 3745–3753, Aug. 2018.

[8] "[online] opower-transform every customer into a demand response resource," http://www.opower.com/bdrpotential/index.html#us, accessed on: 2017-11-29.

[9] M. Jain, V. Chandan, M. Minou, G. Thanos, T. K. Wijaya, A. Lindt, and A. Gylling, "Methodologies for effective demand response messaging," in *IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, 2015, pp. 453–458.

[10] M. Nicolson, G. M. Huebner, D. Shipworth, and S. Elam, "Tailored emails prompt electric vehicle owners to engage with tariff switching information," *Nature Energy*, vol. 2, no. 6, p. 17073, 2017.

[11] L. C. Haynes, D. P. Green, R. Gallagher, P. John, and D. J. Torgerson, "Collection of delinquent fines: An adaptive randomized trial to assess the effectiveness of alternative text messages," *J. Policy Analysis and Management*, vol. 32, no. 4, pp. 718–730, 2013.

[12] Editorial, "The many faces of resilience," *Nature Energy*, vol. 3, no. 83, pp. 1–1, 2018.

[13] M. M. Rahman, G. Shafiullah, A. Arefi, H. Pezeshki, and S. Hettiwatte, "Improvement of voltage magnitude and unbalance in lv network by implementing residential demand response," in *IEEE PES General Meeting*, 2017, pp. 1–5.

[14] K. McKenna and A. Keane, "Residential load modeling of price-based demand response for network impact studies," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2285–2294, 2016.

[15] M. Muratori, "Impact of uncoordinated plug-in electric vehicle charging on residential power demand," *Nature Energy*, vol. 3, pp. 193–201, 2018.

[16] M. Wolinetz, J. Axsen, J. Peters, and C. Crawford, "Simulating the value of electric-vehicle–grid integration using a behaviourally realistic model," *Nature Energy*, vol. 3, no. 2, p. 132, 2018.

[17] L. Chuan and A. Ukil, "Modeling and validation of electrical load profiling in residential buildings in singapore," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2800–2809, 2015.

[18] I. Richardson, M. Thomson, and D. Infield, "A high-resolution domestic building occupancy model for energy demand simulations," *Energy and buildings*, vol. 40, no. 8, pp. 1560–1566, 2008.

[19] S. R. Konda, L. K. Panwar, B. K. Panigrahi, and R. Kumar, "Investigating the impact of load profile attributes on demand response exchange," *IEEE Trans. Ind. Inf.*, vol. 14, no. 4, pp. 1382–1391, 2018.

[20] M. Smith and D. Ton, "Key connections: The us department of energy? s microgrid initiative," *IEEE Power Energy Mag.*, vol. 11, no. 4, pp. 22–27, 2013.

[21] "[online] the model city mannheim beacon project," https://www.mvv.de/en/mvv_energie_gruppe/nachhaltigkeit_2/nachhaltig_wirtschaften_1/innovationen_1/modellstadt_mannheim_1/moma.jsp, accessed on: 2018-04-11.

[22] "[online] phishing scams cost american businesses half a billion dollars a year," https://www.forbes.com/sites/leemathews/2017/05/05/phishing-scams-cost-american-businesses-half-a-billion-dollars-a-year/#3bdad2243fa1.

[23] C. Gutierrez, T. Kim, R. Della Corte, J. Avery, M. Cinque, D. Goldwasser, and S. Bagchi, "Learning from the ones that got away: Detecting new forms of phishing attacks," *IEEE Trans. Dependable Secure Comput.*, vol. 15, no. 6, pp. 988–1001, 2018.

[24] M. Jakobsson, "Two-factor inauthentication–the rise in sms phishing attacks," *Computer Fraud & Security*, vol. 2018, no. 6, pp. 6–8, 2018.

[25] K. L. Chiew, K. S. C. Yong, and C. L. Tan, "A survey of phishing attacks: their types, vectors and technical approaches," *Expert Systems with Applications*, vol. 106, no. 1, pp. 1–20, 2018.

[26] J. S. Warner and R. G. Johnston, "Gps spoofing countermeasures," *Homeland Security Journal*, vol. 25, no. 2, pp. 19–27, 2003.

[27] "[online] cybercriminals use man-in-the-middle attacks to steal 6 million euros," https://www.esecurityplanet.com/hackers/cybercriminals-use-man-in-the-middle-attacks-to-steal-6-million-euros.html, accessed on: 2018-04-11.

[28] "[online] trust your apps-equifax breach revisited," https://www.appthority.com/mobile-threat-center/blog/trust-apps-equifax-breach-revisited/, accessed on: 2018-04-11.

[29] "[online] comcast continues to inject its own code into websites you visit," https://thenextweb.com/insights/2017/12/11/comcast-continues-to-inject-its-own-code-into-websites-you-visit/, accessed on: 2018-04-11.

[30] D. Kahneman, *Thinking, fast and slow*. Macmillan, 2011.

[31] W. H. Kersting, "Radial distribution test feeders," in *IEEE PES Winter Meeting. Conf. Proc. (Cat. No.01CH37194)*, vol. 2, 2001, pp. 908–912.

[32] "[online] report on the grid disturbance on 30th july 2012 and grid disturbance on 31st july 2012," http://www.cercind.gov.in/2012/orders/Final_Report_Grid_Disturbance.pdf.

[33] P. Paruchuri, J. P. Pearce, J. Marecki, M. Tambe, F. Ordonez, and S. Kraus, "Playing games for security: An efficient exact algorithm for solving bayesian stackelberg games," in *Proc. 7th Intl. joint Conf. Autonomous agents and multiagent systems-Vol. 2*, 2008, pp. 895–902.

[34] V. Conitzer and T. Sandholm, "Computing the optimal strategy to commit to," in *Proc. 7th ACM Conf. Electronic Commerce*. ACM, 2006, pp. 82–90.