ANovelAlgorithmwithReducedMutualInformationforSmart MeterPrivacyProtection

[1]SUDEEP KUMAR SINGH, *Gandhi Institute of Excellent Technocrats, Bhubaneswar,India*

[2]SIDDHANT BARIK, Mahavir Institute of Engineering and Technology, Bhubaneswar, Odisha, India

***Abstract*—Fine-grained power utilization information from savvy meters might build the gamble of clients uncovering protection. This paper plans a technique to understand the stowing away of continuous power interest through family energy capacity gear (for example vehicle battery). By planning the rate contortion work issue and tackling it utilizing Blahut-Arimoto calculation, the accompanying beneficial impacts are accomplished: 1) the procedure is attainable; 2) the shared data between continuous power interest and preset meter information can be limited; 3) the information displayed in the meter are normal. Additionally, to confirm the viability of the proposed calculation, examinations with the conventional best-exertion and it are directed to step calculation. The outcomes show the proposed calculation has preferred execution over the over two calculations concerning the common data measurements. At last, the impact of most extreme battery power on the procedure is additionally mimicked.**

Keywords-smart metering; privacy protection; load monitor;dataanalysis; ratedistortiontheory

I. INTRODUCTION

Due to its self-healing function, safety and compatibility,the smart grid has been developed by countries all over theworld[1]-[3].Oneofthefoundationsforefficientoperationof smart grid is the Advanced Metering Infrastructure (AMI),whichreliesonsmartmeters.However,fine-grainedelectricity consumption data recorded by smart meter poses ariskofprivacyleakageforusers.Forexample,byusingnonintrusive load monitoring data analysis techniques, users'usage patterns and personal habits can be obtained [4], [5].Since the possibility of privacyleaks in smartmeters, partsoftheUnitedStatesandEuropeareprohibitedfrominstallation.

Manyalgorithms have been proposed for privacy-preservingofusers’electricityconsumptiondatarecordedby smart meters [6]-[9]. Kalogridis et al. introduce a best-effort(BE) algorithm [10], which tries to keep the meter data fixed.Unfortunately,fortheconstraintsofbatterypowerandcapacity, BE algorithm will definitely expose user’s actualelectricity data and thus leak privacy.Stephen proposes anon-intrusiveload leveling (NILL)algorithm that controlsthebatterytocharge/dischargewhenthecapacityistoolow/high to maintain a fixed value [11]. Given insights fromthefailureofthealgorithmsabovetohandleloadpeaks,a differentapproachbasedonquantizingthedemandloadintoastepfunctionisproposed by[12].

However, none of these works have considered optimalcontrol algorithms to protect the privacy of smart meter data.The optimal control strategy minimizes the impact of batteryhardwareconstraintsonthestrategywhileensuringlowmutualinformation.

Goodprotectioncanbeachievedonlyifthedatapresentedbythemeteris presetand has alowcorrelationwith the actual power demand. In this paper, an algorithmbasedonratedistortiontheorythatminimizes themutualinformation between real-time power demand and the pre-setmeterdataisproposed.

The rest of the paper is organized as follows. InSectionII, we introduce the system model. Section III defines theproblem to be solved. A classical algorithm for solving ratedistortion function is proposed in Section IV. PerformanceevaluationisproposedinSectionV.ThepaperisconcludedinSectionVl.

II. SYSTEMMODEL

Theconsideredinput/outputdiscretetimemodelisshown in Figure 1. Input X_t is the power demand of the homedevice at time *t*. Output $Y_t$is the energy obtained from theutilityprovider(UP).$B_t$isthepowerofthebattery.Thepositivevalueindicatesthatthebatteryischarging,otherwisethebatteryisdischarging.Asdescribedinthemodel, the real-time power demand $X_t$is jointly supplied bythe battery power and the grid power, and EMU’s (EnergyManagement Unit) adjustment of $B_t$can realize the change ofY_t.Intheend,peoplecanpartiallycoveruptheactualelectricityconsumptiontoprotectprivacy.Thefollowingpart introduces the mathematical models of input and outputpower.

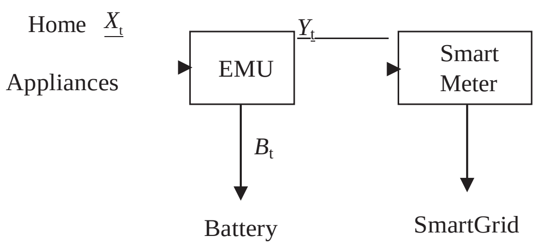

Figure1.Thesystemmodel

A. InputPowerModel

Definition1(HistoricalelectricitydatasetA):Thehistorical electricitydatasetof the family, denoted by$A$.

Definition2:(UniqueConsumptionSet$G$):Grepresentstheset ofunique consumptionrates in$A$.

Definition3(Function$h(\cdot)$):$h(\cdot)$isdefinedtocalculate thenumberoftimesacertainpowerappearsinthefamilyhistoricalelectricity data.

Definition4(Function$g(\cdot)$):Thenumberofelementsin asetcan becalculatedby$g(\cdot)$.

Foreaseofanalysis,theinputpowerdemandisassumed tobeindependentandidenticallydistributed(i.i.d.)with

$$p(X=\omega)=\frac{h(\omega)}{g(A)},\omega\in G.$$

B. OutputPowerModel

Toprotectprivacy,theelectricityconsumption$Y_t$recordedbythemetershouldbewhattheuserwantstopresent. Therefore, a data collection y is customized for $Y_t$toselect. For the feasibility of customized set of meter data,someconstraintsneed tobeconsidered:

- Constraint1:$\max(y)\leq\max(A)+B_{max}$
- Constraint2:$\min(y)\geq\min(A)+B_{max}$
- Constraint3:Arrangeyinorder,aandbareanytwo adjacent numbersiny,then: $|a-b|\leq 2B_{max}$
- Constraint4:$g(y)\geq\frac{\max(A)-\min(A)}{2B_{max}}$

If constraint 1 and 2 are not met, the maximum/minimumvalue of y can be directly deleted. Because it is impossible toachievebyadjusting thebatterypower.

Constraint 3 and 4 are to ensure thatfor any X_t, at leastone element in y can be reached by charging or dischargingthebattery.

III. PROBLEMDEFINITION

In this section, an optimization framework is proposed tosearchtheoptimalstrategythatminimizesthemutual informationbetweeninputpowerdemandandpresetmeterdata.Ratedistortiontheoryhasbeenproventobevery effectiveincalculatingtheminimummutualinformationbetweentwoprobabilitydistributionsetsundergiven constraints.Tothisend,aratedistortionfunctionproblemisformulatedtocapturethebestfeasiblestrategy.

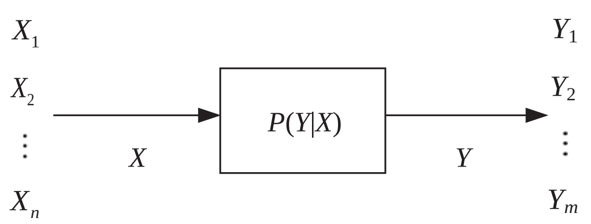

Figure2.Informationtransmissionmodel

A. SystemDynamics

Since the uncertainty of real-time electricity consumptionmakes the quantization problem of solving a single randomvariableverycomplicated,ajointdescriptionismoreeffective than a single description. Based on rate distortiontheory, an information transmission model is developed todescribepowerdemandandmeterdatajointly,whichisshownin Figure2.XinFigure2standsforreal-timepower demandand$Y$isthecustomizedmeterdataset.$p(Y|X)$in thefigurerepresentsthetransmissionprobabilityofthechannel,whichisrepresented bya matrixas:

$$[p(Y|X)]=\begin{bmatrix} p(Y_1|X_1) & p(Y_2|X_1) & \cdots & p(Y_m|X_1) \\ p(Y_1|X_2) & p(Y_2|X_2) & \cdots & p(Y_m|X_2) \\ \cdots & \cdots & \cdots & \cdots \\ p(Y_1|X_n) & p(Y_2|X_n) & \cdots & p(Y_m|X_n) \end{bmatrix}$$

Each element of the matrix represents the probability thatthe real-time power demand is $X_i$and the meter data is Y_i.After setting the meter data, all strategies can be presented intheformofthismatrix.

B. DistortionFunction

After fully considering the role of battery power in thestrategy, absolute distortion is used as the distortion functionofthispaper:

$$d(X,Y)=\begin{cases} |X_t-Y_t|, & |X_t-Y_t|\leq B_{max} \\ d_{tt}, & \text{otherwise} \end{cases} \quad (1)$$

where d_{tt}ismuchlargerthanB_{max}.

Forthedatathatcannotbeachievedbychargingordischarging the battery, the distortion should be large enoughso that there will be no unreachable meter data be selectedgiventheexpected distortion.

C. ObjectiveFunction

Theorem1[13].Forthesource$x$ofi.i.d.,ifthedistributionis$p(x)$ and thebounded distortionfunctionis$d(x,\hat{x})$,thentheratedistortionfunctionisequal tothe correspondinginformationratedistortionfunction,whichis:

$$R(D)=R^{(I)}(D)=\min_{p(\hat{x}|x):\sum_{(x,\hat{x})}p(x)p(\hat{x}|x)d(x,\hat{x})\leq D} I(X,\hat{X}) \quad (2)$$

Accordingtotheorem1,thefollowingobjectivefunction canbeestablished:

$$R(D)=\min_{p(Y|X):\sum_{(X,Y)}p(X)p(Y|X)d(X,Y)\leq D} I(X,Y)$$

$$s.t.\begin{cases} \sum_Y p(Y|X)=1 \\ \sum p(Y|X)p(X)d(X,Y)\leq D \\ p(Y|X)\geq 0 \end{cases} \quad (3)$$

TheintermediatevariableDcanbeshownas:

$$\begin{aligned} D&=\sum_{i=1}^{n}\sum_{j=1}^{m}p(X_iY_j)d(X_i,Y_j) \\ &=\sum_{i=1}^{n}\sum_{j=1}^{m}p(X_i)p(Y_j|X_i)d(X_i,Y_j) \end{aligned} \quad (4)$$

D isnotonlytheexpectederror,butalsoequalsinvaluetotheexpectedpowerofthebattery.Theminimumvalueis takenfromalltransitionprobabilitymatrices $p(Y|X)$ that makethe joint distribution $p(X)p(Y|X)$ satisfy the expectationofdistortionconstraint D.Forthisproblem,various toolsof rate distortiontheorycanbe utilized.

IV. SOLUTIONMETHODOLOGY

Thissectionintroducesaclassicalmethod,Blahut-Arimotoalgorithm[13],forcalculatingtheratedistortionfunction. A description of the Blahut-Arimoto algorithm ispresented in Algorithm 1. It is executed by the EMU to learntheoptimalstrategybasedonhistoricalhouseholdconsumption data.

TheBlahut-Arimotoalgorithm uses an alternating minimizationalgorithmtocalculatetheminimummutual informationbetweenpowerdemand(X)andcustomizedmeterdata (Y).Ineachiteration,foragiveninput distribution $p(X)$ and output distribution $p(Y)$, first calculate $p(Y|X)$ thatminimizesmutualinformationunder distortionconstraints:

$$p(Y|X)=\frac{p(Y)e^{-\lambda d(X,Y)}}{\sum p(Y)e^{-\lambda d(X,Y)}} \quad (5)$$

where λ isthelagrangeparameter.

Lemma1[13]:Let $p(X)p(Y|X)$ bethegivenjointdistribution. Then the distribution $r^*(Y)$ which minimizes therelativeentropy

$$D(p(X)p(Y|X)||p(X)r^*(Y)) =\min_{r(Y)}(D(p(X)p(Y|X)||p(X)r(Y))) \quad (6)$$

isthemarginaldistribution $r^*(Y)$ correspondingto $p(Y|X)$, where $r^*(Y)=\sum_X p(X)p(Y|X)$.

Accordingtothelemma1,theoutputdistribution $p(Y)$, whichminimizesmutualinformation,canbecalculatedasfollows:

$$p(Y)=\sum_X p(X)p(Y|X) \quad (7)$$

Theentropyandjointentropyoftherandomvariableinstep 8 canbe obtainedbythefollowingtwoformulas:

$$H(X)=-\sum_{i=1}^{n} p(X_i)\log_2 p(X_i) \quad (8)$$

$$H(X,Y)=-\sum_{X_i\in X}\sum_{Y_i\in Y} p(X_i,Y_i)\log_2 p(X_i,Y_i) \quad (9)$$

$I_1(X;Y)$ and $I_2(X;Y)$ inStep10representthemutual informationobtainedbytwoadjacentcomputations.

As the number of iterations increases, this minimizationprocessconvergesto $R(D)$.

V. PERFORMANCEEVALUATION

In this section, the proposed Battery-based Load Hiding(BLH) strategy is compared with two other common BLHalgorithms, BE andStepping, to evaluate its performance.Moreover,theinfluenceofthemaximumpowerofthebatteryonthisstrategyisalsostudiedandshownintheFigure4.

Algorithm1.Blahut-AritomoAlgorithm

Step1:Inputthedistributionofelectricitydemand $p(X)$.
Step2:Inputiterationprecision *prec*,maximumnumber ofiterations t andLagrangeparameter λ.
Step3:Inputthepresetmeterdataanditsdistribution $p(Y)$.
Step4: Repeat(for each λ):
Step5: $c=1$
Step6:Calculatethetransitionprobabilitymatrix $p(Y|X)$ from (5).
Step7:Obtaintheoutput distribution $p(Y)$ from(7).
Step8:Calculatemutualinformationbetween X,Y:
$$I(X;Y)=H(X)+H(Y)-H(X;Y)$$
Step9: $c=c+1$
Step10:Until $c>t$ or $|I_1(X;Y)-I_2(X;Y)|\le prec$
Step11:Output $p(Y|X)$

BEalgorithm

BEalgorithmtriestoresist(tothedegreepossible) againstpowerloadchanges.Whentherequiredloadiseitherlargeror smallerrespectivelythanthepreviouslymetered load,EMUwillchargeordischargethebatterytomakeupthedifference.

Steppingalgorithm

Steppingalgorithmaimstoquantifythedemandloadinto astepfunction,whichisdeterminedbythebattery's maximumcharge/dischargerate.Inthesteppingalgorithm, theEMUforcestheexternalloadtobemultiplesof £,which isdeterminedbythebattery's parameters.

A. InfluenceofDifferentAlgorithms

Weconsidera24-hourplanningperiodinthesimulation. Thesmartmetercurrentlydeployedinprivatedwellingsin theEastMidlands,UK.Tounderstandtheeffectsofthesealgorithmsintuitively,theoriginalelectricitydataisalso shown.Themaximumpowerofthebatteryusedforthe simulationis1000Wandcapacityis0.5kWh.Themeterdata presetbytheproposedalgorithmis(0,200,400,800,1000, 1200,1600,2000,2400,2800,3200,3600, 4000,4800).

Themutualinformationbetweenthecurrentmeterdataandthereal-timepowerdemandispresentedbyusinthe upperrightcorneroftheexternalloadfigureforeachalgorithm.

SimulationresultsshowthattheBEalgorithmhasagood effect on the power demand less than the maximum power ofthe battery. But, due to the constraint of maximum power ofthe battery, it has limited ability to hide the large real-timeelectricity consumptionrate. Itcanalso beseenfromthefigure that since the stepping algorithm maximizes the errorbetween the power demand and the meter data, the battery isvery easy to overcharge or over discharge. Eventually lead totheleakageofreal-timeelectricityinformation.Ouralgorithmnotonlyhasagoodeffectonthepowerdemandlessthanthemaximumpowerofthebattery, butalsocanfind a good representation of the power demand greater thanthe maximumpowerofthebatteryfromthe preset powerset.

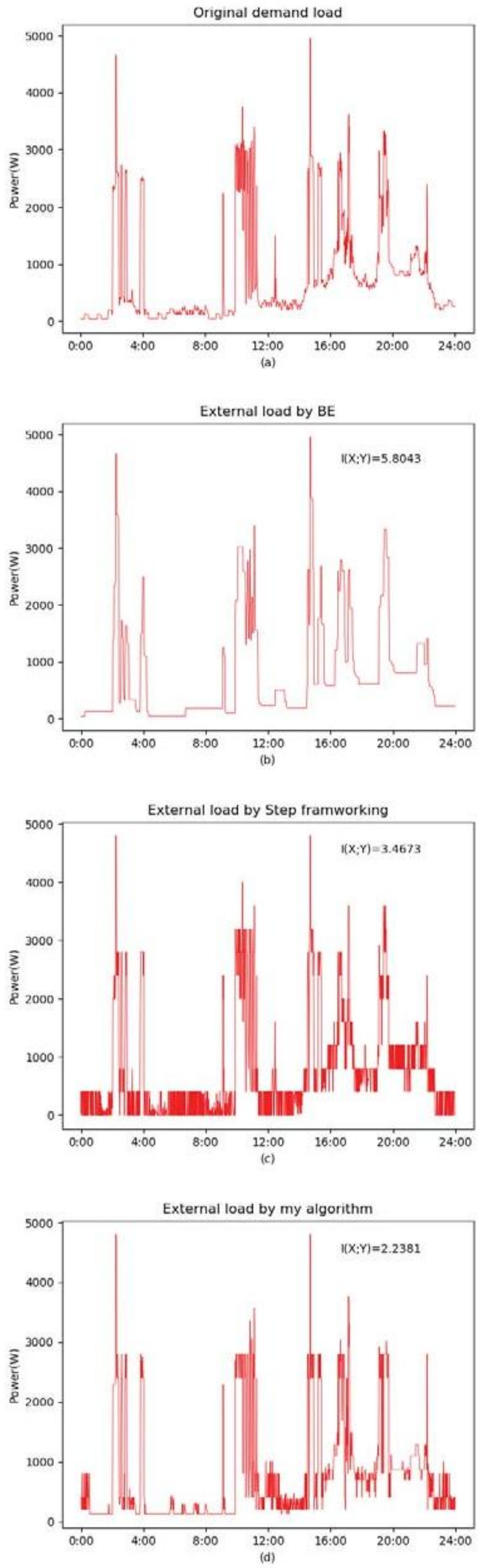

Figure3.Influenceofdifferentalgorithms

B. InfluenceofDifferentBatteries

Figure 4 shows the impact of different maximum powerbatteries on this strategy. The abscissa is the expected error(which is numerically equal to the expected power of thebattery) and the ordinate is the minimum mutual information.It is a non-increase convex function on *D*, indicating that theminimummutualinformationcanbeachievedisgettingsmaller andsmallerastheexpectedpowerofthebatteryincreases.

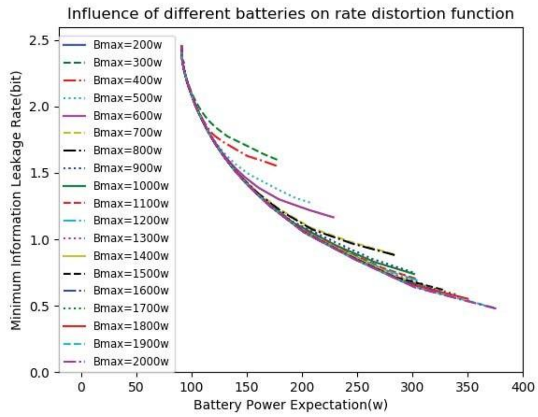

Figure4.Influenceofdifferentbatteriesonratedistortionfunction

VI. CONCLUSION

TheexistingBLHstrategiesdonotconsideroptimalcontrol algorithms to protect smart meter data. We preset thedatapresentedbythemeterandproposeanewBLHalgorithmwi threducedmutualinformationbasedonthetheory of rate distortion. Compared with other strategies, thealgorithm proposed in this paper is an online optimizationalgorithmdesignedforthefamily'shistoricalpowerc onsumptionsituationanduserneeds,soitcanachievebetter protection effects. The transition probability matrix iscalculated in advance, so the smart meter does not need tocarry out a lot of calculations in real time, and the hardwarerequirements of the smart meter are not high.The mutualinformation is used as a measure to compare the proposedalgorithmwithBEandSteppingalgorithmandtheexperi mentalresultsdemonstratetheeffectivenessoftheproposed algorithm. In the future, the impact of other factors,suchas batterycapacity,onthisstrategywillbestudied.

REFERENCES

[1] L. Yang, X. Chen, J. Zhang, and H. V. Poor, "Cost-effective andprivacypreservingenergymanagementforsmartmeters,"IEEETrans .SmartGrid,vol.6,no.1,pp.486–495,Jan.2015.

[2] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid—The new andimproved power grid: A survey," IEEE Commun. Surveys Tuts., vol.14,no.4,pp.944–980,Oct.2012.

[3] C. Zhao, J. He, P. Cheng, and J. Chen, "Analysis of consensus-baseddistributed economic dispatch under stealthy attacks," IEEE Trans.Ind.Electron.,vol.64,no.6,pp.5107–5117,Jun.2017.

[4] J. Kelly and W. Knottenbelt, "Neural NILM: Deep neural networksappliedtoenergydisaggregation," inProc. ACMInt.Conf.EmbeddedSyst. EnergyEf烄cientBuiltEnviron.,Seoul, SouthKorea,Nov. 2015,pp. 5564.

[5] M.Baker, G. Hicks, S. Rodriguez, andM. Fuller, "A test of commercially available products for estimating end uses from smartmeter data," in Proc. Int. Workshop Non Intrusive Load Monitor.,Vancouver,BC,Canada,May2016.[Online].Available:http://nilmworkshop.org/2016/proceedings/Paper_ID22.pdf

[6] Y.Sun,L.Lampe,andV.W.S.Wong, "Smartmeterprivacy:Exploiting the potential of household energy storage units," IEEEInternetThingsJ.,vol.5,no.1,pp.69–78,Feb.2018.

[7] Y. Hong, W. M. Liu, and L. Wang, "Privacy preserving smart meterstreaming against information leakage of appliance status," IEEETrans. Inf. Forensics Security, vol. 12, no. 9, pp. 2227–2241, Sep.2017.

[8] G. Giaconi, D. Gündüz, and H. V. Poor, "Smart meter privacy withrenewable energy and an energy storage device," IEEE Trans. Inf.ForensicsSecurity,vol.13,no.1,pp.129–142,Jan.2018.

[9] E. Liu and P. Cheng, "Achieving privacy protection using distributedload scheduling: A randomized approach," IEEE Trans. Smart Grid,vol.8,no.5,pp.2460–2473,Sep.2017.

[10] G. Kalogridis, C. Efthymiou, S. Denic, T. Lewis, and R. Cepeda.Privacyforsmartmeters:Towardsundetectableapplianceloadsi gnatures. In Smart Grid Communications (SmartGridComm), 2010FirstIEEEInternationalConferenceon,pages232–237,oct.2010.

[11] S. McLaughlin, P. McDaniel, and W. Aiello. Protecting consumerprivacy from electric load monitoring. In Proceedings of the 18thACM conferenceon Computer and communications security,CCS'11,pages87–98,NewYork,NY,USA,2011.ACM.

[12] W. Yang et al., "Minimizing private data disclosures in the smartgrid,"in Proc.ACMCCS,2012,pp.415–427.

[13] M Cover Thomas, A Thomas Joy, Elements of Information Theroy,2ndedn.(JohnWileyandSons,Hoboken,NewJersey,2006.