2020

On Masking and Releasing Smart Meter Data at Micro-level: the Multiplicative Noise Approach

John Brackenbury

P. Y. O'Shaughnessy
University of Wollongong, poshaugh@uow.edu.au

Yan-Xia Lin
University of Wollongong, yanxia@uow.edu.au

Recommended Citation

Brackenbury, John; O'Shaughnessy, P. Y.; and Lin, Yan-Xia, "On Masking and Releasing Smart Meter Data at Micro-level: the Multiplicative Noise Approach" (2020). *Faculty of Engineering and Information Sciences - Papers: Part B*. 4519.
https://ro.uow.edu.au/eispapers1/4519

On Masking and Releasing Smart Meter Data at Micro-level: the Multiplicative Noise Approach

Abstract

Smart meter electricity data presents privacy risks when malicious agents gain insights of private information, including residents’ lifestyle and daily habits. When allowing access to record-level data, we apply the multiplicative noise method to mask individual smart meter data, which simultaneously aims to minimise disclosure of a dwelling’s consumption signal to any third party and to enable accurate estimation of the sum of a cluster of households. Three testing criteria are introduced to measure the performance of multiplicative noise masking approach relevant to the smart meter data. We propose a novel ‘Twin Uniform’ noise distribution and derive relevant theoretical results. We then implement the multiplicative noise approach in the real smart meter data from ESSnet Big Data. Results are assessed based on privacy, utility and practicality. We conclude that the multiplicative noise method has outstanding practical values. It preforms reasonably well in term of individual value protection and estimation accuracy of the sum when noise distribution is carefully selected .

Keywords

multiplicative, noise, releasing, masking, smart, approach, meter, data, micro-level:

Disciplines

Engineering | Science and Technology Studies

Publication Details

Brackenbury, J., O'Shaughnessy, P. Y. & Lin, Y. (2020). On Masking and Releasing Smart Meter Data at Micro-level: the Multiplicative Noise Approach. PSD2020 - Privacy in Statistical Databases (pp. 1-13).

On Masking and Releasing Smart Meter Data at Micro-level: the Multiplicative Noise Approach

John Brackenbury[1,2], P.Y. O'Shaughnessy[1,3], and Yan-Xia Lin[1]

[1] School of Mathematics and Applied Statistics, University of Wollongong, Australia
[2] jbrakenbury996@gmail.com
[3] poshaugh@uow.edu.au

Abstract. Smart meter electricity data presents privacy risks when malicious agents gain insights of private information, including residents' lifestyle and daily habits. When allowing access to record-level data, we apply the multiplicative noise method to mask individual smart meter data, which simultaneously aims to minimise disclosure of a dwelling's consumption signal to any third party and to enable accurate estimation of the sum of a cluster of households. Three testing criteria are introduced to measure the performance of multiplicative noise masking approach relevant to the smart meter data. We propose a novel 'Twin Uniform' noise distribution and derive relevant theoretical results. We then implement the multiplicative noise approach in the real smart meter data from ESSnet Big Data. Results are assessed based on privacy, utility and practicality. We conclude that the multiplicative noise method has outstanding practical values. It preforms reasonably well in term of individual value protection and estimation accuracy of the sum when noise distribution is carefully selected .

Keywords: Data protection · Statistical disclosure · Multiplicative noise · Utility loss · Smart Meter data.

1 Introduction

Whilst the rapid development in technology has created an entirely new data industry in the past few decades, handling of data presents a privacy concern for individuals. Notwithstanding the legal nuances of privacy policies, there is an expectation among data owners that their sensitive data would be protected, including a level of anonymisation when transferring control of data to third parties. The question of how to measure protection procedures and assess data privacy is broad. Given the diversity of data types and purposes, privacy metrics are often chosen according to application, making comparison across applications problematic. Thoughts must always be given to the nature of what is being protected, and to the means by which unauthorised agents may access such information [7].

One of the areas of application is the ‘smart meter’, which is an emerging technology in the energy industry. It records time-series electricity consumption data in contrast to the conventional cumulation methodology. The meter transmits consumption figures automatically at regular intervals, which has potential to revile residents living habit and poses serious security concerns. In addition, the needs of energy providers must also be considered for smart meter. For accurate billing, the *total consumption* over the billing period calculated from the altered time-series must not deviate beyond a certain tolerance from the true total. When the company intends to analyse network demand throughout the day, such analysis will not be meaningful unless the deviation on *each time point* is limited.

In 2010, Bohli, Sorge and Ugus published a paper specifically addressing the privacy concerns of smart meters [2]. Whilst recognising the needs for individual billing and network demand, the authors emphasised that the electricity suppliers operating the meters should also be kept from possessing sensitive data pertaining to individuals’ lifestyles. DiffeRentially privatE smArt Metering (DREAM) was proposed [1] to protect data privacy for smart meters. Based on the concept of differential privacy, a perturbative scheme is proposed that ensures the privacy of individual time-series transmitted to an electricity supplier for aggregate demand analysis at each time point. Each meter applies a Gamma noise to its own signal, which becomes a Laplace noise when cumulated across the cluster. Decryption keys is then used for protecting data transfer, and the electricity supplier recieves a noisy sum from each cluster, with each individual household differentially protected.

The existing DREAM method limits to query systems and is difficult to offer detailed information on the household-level of electricity consumption. This means that electronic suppliers are unable to conduct further analyses to gain statistical other insights as many advanced statistical analyses require recorded-level data. Here we propose to use the multiplicative noise masking method, which allows releasing the record-level information while protecting the privacy of smart meter data. Noise multiplication is a standard data masking method, but to the best of our knowledge, it has not been implemented in the smart meter data, which has its unique privacy requirements.

In this paper, we first give a brief description of the smart meter dataset from European Statistical System and discuss the relevant testing criteria specific for the smart meter data in Section 2. In Section 3, a novel distribution is introduced for the multiplicative noise and an algorithm specific for masking the smart meter data is proposed. Section 4 evaluates the performance of the multiplicative noise techniques in three aspects using the smart meter dataset and the conclusion is given in Section 5.

2 Data Description and Testing Criteria

In this section, we describe a real dataset of smart electricity meter data, and introduce the testing criteria for measuring the performance of the masking technique. Smart Meter data is obtained from ESSnet Big Data[4], which is a current project within the European Statistical System (ESS). The dataset consists of smart meter readings with various time interval lengths between 2013 and 2015 recorded in Denmark and Estonia. We select 4-day long hourly electricity consumption readings ($T = 96$ time points) with entries measured in kilowatt-hours (kWh) from $N = 3371$ households for illustration in this paper.

Let X_t denote the smart meter readings at time t for $t = 1, \ldots, T$. More specifically, we assume that for $i = 1, \ldots, N$, the smart meter readings for the ith household at time t, X_t^i, follows the same distribution as X_t. The N households are separated into C clusters and each cluster has n_c households. We define S_t to be aggregated sum for a cluster of smart meter readings at time t, which defines the underlying distribution for S_t^c for $c = 1, \ldots, C$. This data set is used to compare the utility and protective efficacy of data-masking methods in the specific application of protecting individual households' electricity data. In particular, the **privacy requirements of Smart Meter data** are to simultaneously

1. *enable accurate estimation* of the sum $S_t^c = \sum_{i=1}^{n_c} X_t^i$ for a defined cluster of n_c dwellings at each time t; and
2. *protect individual value* $\{X_t^i\}$ from accurate estimation by the electricity suppliers.

Considering these two privacy requirements, we introduce three testing criteria for Smart Meter data. The first measure is an accuracy measure, namely, the *probability of tolerable relative error*

$$p_{\delta,S}(t) = P\left(\left|\frac{\hat{S}_t - S_t}{S_t}\right| < \delta\right), \qquad p_{\delta,X}(t) = P\left(\left|\frac{\hat{X}_t - X_t}{X_t}\right| < \delta\right),$$

where $\delta > 0$ is the tolerance threshold, e.g. $\delta = 0.1$ suggests a 10% tolerance in differences between a masked value and its true value. We

[4] See `ebgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/ESSnet_Big_Data` for more information.

denote $\hat{S}_t$ and $\hat{X}_t$ as the estimated values for cluster sum at time t and individual household at time t, respectively. The calculations of $\hat{S}_t$ and $\hat{X}_t$ vary according to the implemented data masking methods. Referred to as the *rate of error* or *accuracy rate*, p_δ in both cases are estimated by the sample proportions

$$\hat{p}_{\delta,S}(t) = \frac{1}{C}\sum_{c=1}^{C} I\left(\left|\frac{\hat{s}_t^c - s_t^c}{s_t^c}\right| < \delta\right), \quad \hat{p}_{\delta,X}(t) = \frac{1}{N}\sum_{i=1}^{N} I\left(\left|\frac{\hat{x}_t^i - x_t^i}{x_t^i}\right| < \delta\right),$$

where $I(\cdot)$ is the indicator function and the lower cases x and $\hat{x}$ are the actual observations of variables X and $\hat{X}$, respectively. The cluster sum $s_t^c = \sum_{i=1}^{n_c} x_t^i$ is the realizations of S_t^c. The estimated masked cluster sum $\hat{s}_t^c$ is similar. We use the accuracy measures $p_{\delta,S}(t)$ and $p_{\delta,X}(t)$ differently. For estimation of the sum S_t, we wish high accuracy to maximise utility, corresponding to high $p_{\delta,S}(t)$. For the individual household values X_t we wish to maximise privacy and minimise malicious estimation accuracy, corresponding to low $p_{\delta,X}(t)$. The latter is critical in measuring the vulnerability of data masking methods.

The second and third measures used in this paper are *mean relative error* (MRE) and *mean unsigned relative error* (MURE), namely

$$\text{MRE} = \frac{1}{C}\sum_{c=1}^{C}\frac{\hat{s}_t^c - s_t^c}{s_t^c}, \qquad \text{MURE} = \frac{1}{C}\sum_{c=1}^{C}\left|\frac{\hat{s}_t^c - s_t^c}{s_t^c}\right|.$$

Different from the rate of error $p_{\delta,S}$, MRE and MURE quantify the overall or average relative error across all clusters at a time t. Smaller value means more accurate estimation of the sum.

3 The Multiplicative Noise Method

Noise multiplication or addition has a long history of being used for data masking, mainly in official statistics and remote sensing imagery (see [4], [5], [6], [3]). These two methods are similar in nature but each has quite different privacy properties. [10] provided detailed review for both the additive and multiplicative noise methods, and they pointed out that it is difficult to generate the noise values under the additive noise method given that the appropriate selection of noise distribution depends on the original data values. By contrast the multiplicative noise method is able to provide uniform record level protection in terms of noise coefficient of variation to all values. The Smart Meter data has large variations among electricity consumption values, thus we consider a multiplicative masking technique for smart meter data, and then propose a novel distribution appropriate for such noise to result in the greatest protective benefit.

Consider the aggregate sum of consumption from a group of n_c households' smart meters at each time t. The smart meter at each house records its consumption levels as a regularly-spaced time-series, with the measurement from the ith node at time t denoted X_t^i. In general, the concept of employing Multiplicative Noise is to simply calculate the perturbed values $\tilde{X}_t$, such that

$$\tilde{X}_t = X_t \times M_t,$$

where M_t is a random variable sampled independently from a distribution with mean μ_M and standard deviation σ_M for time t. Our goal is to ensure that the electricity supplier is able to estimate the true cluster sum $S_t^c = \sum_{i=1}^{n_c} X_t^i$ without disclosure of the individual household values $\{X_t^i\}$, in order to protect the privacy of each household.

3.1 Shifting

Non-negativity in electricity consumption is valid, given current should never flow in the opposite direction to ordinary power usage in Smart Meter data. It is still possible for a zero value to occur from time to time in real life. This creates two primary issues. First, from the practical perspective of the simulation metrics, the relative error p_δ does not exist due to division by zero. This itself is not insurmountable, but it is desirable to avoid it. The second issue is more serious: multiplicative noise applied to a signal $X_t^i = 0$ causes no perturbation, thus the masking is nullified. This poses a high disclosure risk for households with zero signals. Indeed, given that zero signals are often a result of dwelling vacancy that persists across many hours, even days, zero signal households must be protected.

[8] offered a solution to the zero signal issues by formulating a minor modification to the scheme, referred to as *shifting*. The application of the shifting approach can be found in [11]. The algorithm for applying the multiplicative noise method to Smart Meter data is given as follows:

Algorithm: Multiplicative Noise with Shifting for masking Smart Meter data

1. Each node shifts their measurement by a value a, which is known to the electricity suppliers :
   $Y_t = X_t + a$ for $a > 0$.

2. Each node samples from the distribution M, multiplies: $\tilde{Y}_t = Y_t \times M_t = (X_t + a)M_t$.
3. The electricity suppliers receive noisy, shifted values $\{\tilde{Y}_t^i\}$, and aggregate them to obtain $\tilde{S}_t^c = \sum_{i=1}^{n_c} \tilde{Y}_t^i = \sum_{i=1}^{n_c} (X_t^i + a)M_t^i$.
4. The electricity suppliers use the noise mean μ_M to obtain the *unbiased* estimate for the cth cluster:

$$\hat{S}_{t,M}^c = \frac{\tilde{S}_t^c}{\mu_M} - n_c a = \frac{\sum_{i=1}^{n_c} \tilde{Y}_t^i}{\mu_M} - n_c a = \frac{\sum_{i=1}^{n_c} (X_t^i + a)M_t^i}{\mu_M} - n_c a\,.$$

Here n_c denotes the number of households in the cth cluster for $c = 1, \ldots, C$. Note that some information about the noise distribution is generally available to electricity suppliers, e.g., the distribution mean, in order to use μ_M for the estimate $\hat{S}_{t,M}^c$.

We discuss this algorithm with respect to the two privacy requirements listed in Sec. 2. Requirement 1 is about the estimation accuracy of the sum, for which an unbiased sum estimate, $\hat{S}_t^c$, is defined, which needs to de-scale and de-shift the perturbed sum $\tilde{S}_t^c$. It is unbiased in the following sense:

$$\mathrm{E}(\hat{S}_{t,M}^c | \{X_t^i\}) = \sum_{i=1}^{n_c} \left[(X_t^i + a) \frac{\mathrm{E}(M_t^i)}{\mu_M} \right] - n_c a = \sum_{i=1}^{N} X_t^i + (n_c a - n_c a) = S_t^c\ ,$$

where $\mathrm{E}(M_t^i) = \mu_M$. The sum estimate $\hat{S}_{t,M}^c$ satisfies two properties:

Property 1. Estimator $\hat{S}_t^c$ is an unbiased estimator for the true sum S_t^c given the true values $\{X_t^i\}$.

Property 2. The conditional standard error of the estimated sum $\hat{S}_{t,M}^c$ is
✓ $\{\sigma_M/\mu_M\}\sqrt{\sum_{i=1}^{n_c} \left(X_t^i + a\right)^2}$, given $\{X_t^i\}$ values.

These properties imply that the electricity suppliers are able to unbiasedly estimate the true cluster sum using $\hat{S}_t^c$ (Property 1) with a known estimation error propotional to the true electricity consumption (Property 2).

For Privacy Requirement 2, we need to consider the how well the individual value is protected. Given that noise mean μ_M and shifting parameter a are known to electricity suppliers, it is possible for suppliers to construct an unbiased estimator $\hat{Y}_{t,M} = \tilde{Y}_t/\mu_M = (Y_t M_t)/\mu_M$ of Y_t.

We are essentially comparing shifted estimate against shifted true value, and this avoids division by zero as described above. The individual value protection will be evaluated empirically in Sec. 4.

To the attentive reader, this may raise the question: *So what are we really protecting now, the true value X_t or the shifted true value Y_t?* It is true that since we apply an additive transformation to form Y_t, the relative error on the scale of Y_t is not equal to the relative error on the scale of X_t. Given the shifting value a is public knowledge, the argument we make as resolution is that if the values Y_t are sufficiently protected (i.e. $p_{\delta,Y}$ using $\hat{Y}_{t,M}$ values is sufficiently low), then X_t^i is sufficiently protected also. Conversely, if Y_t is poorly protected and vulnerable to disclosure, i.e. $p_{\delta,Y}$ is high, then X_t is also vulnerable, since it can be estimated accurately by a simple shift of an accurate estimate $\hat{Y}_{t,M}$.

3.2 Noise Distribution: Twin Uniform Distribution

Given that both perturbed values and the mean of the noise often are known to the electricity supplier, the distribution of the multiplicative noise plays an important role in protecting the values of the original data. An obvious choice would be a simple uniform distribution. The problem with the simple uniform distribution is that with a desired level of accuracy in estimation of the sum, the range of the parameters (maximum noise - minimum noise) need to be small, and as a result, the masked data is not sufficiently different from the original values. Broadly speaking, we need a distribution that is less distributed around the mean and also has small variance.

Here we introduce the *twin uniform distribution* for a multiplicative noise M, which has the following density

$$f_M(m) = \begin{cases} \frac{1}{2\mu(\alpha_{\max} - \alpha_{\min})}, & m \in A \\ 0, & \text{else}\ , \end{cases} \tag{1}$$

where $A = [\mu(1-\alpha_{\max}), \mu(1-\alpha_{\min})] \cup [\mu(1+\alpha_{\min}), \mu(1+\alpha_{\max})]$ and we denotes (1) as $M \sim \text{TwinUnif}(\mu, \alpha_{\min}, \alpha_{\max})$. The twin uniform distribution is essentially a mixture distribution of the two uniform distributions $\text{Unif}(\mu(1-\alpha_{\max}), \mu(1-\alpha_{\min}))$ and $\text{Unif}(\mu(1+\alpha_{\min}), \mu(1+\alpha_{\max}))$, mixed with equal weight of probability (Fig. 1). The mean and variance for a random noise M with $\text{TwinUnif}(\mu, \alpha_{\min}, \alpha_{\max})$ distribution are

$$\text{E}(M) = \mu \quad \text{and} \quad \text{Var}(M) = \underline{\frac{\mu^2(\alpha_{\max}^2 + \alpha_{\max}\alpha_{\min} + \alpha_{\min}^2)}{3}}\ .$$

The distribution is relatively simple, making it feasible for implementation in the inexpensive hardware of typical smart meters. When using a twin uniform noise distribution, the conditional variance of the sum estimator for a particular cluster of size n_c (Property 2) is

$$\mathrm{Var}(\hat{S}^c_{t,M}|\{X^i_t\}) = \sqrt{\frac{\alpha^2_{\mathrm{max}} + \alpha_{\mathrm{max}}\alpha_{\mathrm{min}} + \alpha^2_{\mathrm{min}}}{3} \sum_{i=1}^{n_c}} \left(X^i_t + a\right)^2 . \tag{1}$$

✓

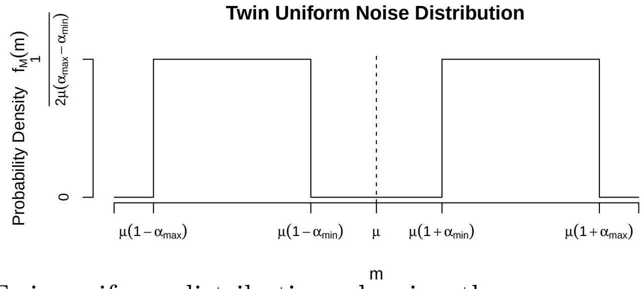


Fig. 1: Twin uniform distribution, showing the gap around the mean μ that prevents accurate estimation of X^i_t by the electricity supplier.

4 Results and Discussion

In this section, we evaluate the proposed multiplicative noise algorithm with twin uniform noise distribution through three criteria, using the ESSNET data described in Section 2. The cluster size is $n_c = 100$ for $N = 3371$ households with 96 hourly time points (4 days). We masked Smart Meter data using a twin uniform distribution (1) with $\mu = 27$ and setting the shift parameter $a = 0.7$. We discuss the performance of the proposed multiplicative noise masking method with various parameter values on three aspects relevant to the Smart Meter data, namely, *privacy*, *utility*, and *practicality*.

4.1 Privacy

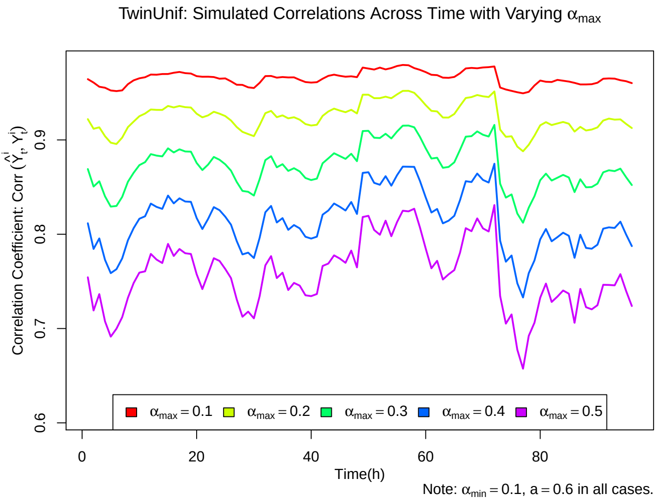
We firstly consider the *probability of tolerable relative error* for measuring disclosure risk for individual data (Privacy Requirement 2, Sec. 2) for

$\hat{Y}_{t,M}$, the estimator of Y_t. This is estimated by the sample proportion $\hat{p}_{\delta,Y}$, i.e.,

$$\hat{p}_{\delta,Y}(t) = \frac{1}{N}\sum_{i=1}^{N} I\left(\left|\frac{\hat{y}^i_{t,M} - y^i_t}{y^i_t}\right| < \delta\right).$$

To ensure the privacy of individual data, it needs to satisfy $\hat{p}_{\delta,Y}(t) = 0$ for all t at given δ level. Theoretically, $p_{\delta,Y}(\alpha_{min}, \alpha_{max}) = (\delta - \alpha_{min})/(\alpha_{max} - \alpha_{min})$. This means that when taking $\alpha_{min} = \delta$, it guarantees that none of the masked value will be close enough to the true value.

In addition to value disclosure, risks of the multiplicative noise scheme may arise from other forms of estimation by malicious agents, for example, *regression risk*, which is the risk of improved estimation by fitting a model to the received data $\tilde{Y}_t$. This is measured by the correlation of estimator $\hat{Y}_{t,M}$ with true values Y_t, corr$(\hat{Y}_{t,M}, Y_t)$. [9] concluded that a correlation below 0.8 shows large enough estimation uncertainty and is considered reasonable protection for individual value against regression risk. Choosing $\delta = \alpha_{min} = 0.1$ and ensuring $\hat{p}_{\delta,Y}(t) = 0$ for all t, Fig. 2 displays the sample correlations across time when varying α_{max}. The results show that the correlations decreases (better protection) when α_{max} increases (larger masking noise), and $\alpha_{max} \geq 0.4$ limits the correlations approximately to 0.8, suggesting good individual value protection.



Note: $\alpha_{min} = 0.1$, a = 0.6 in all cases.

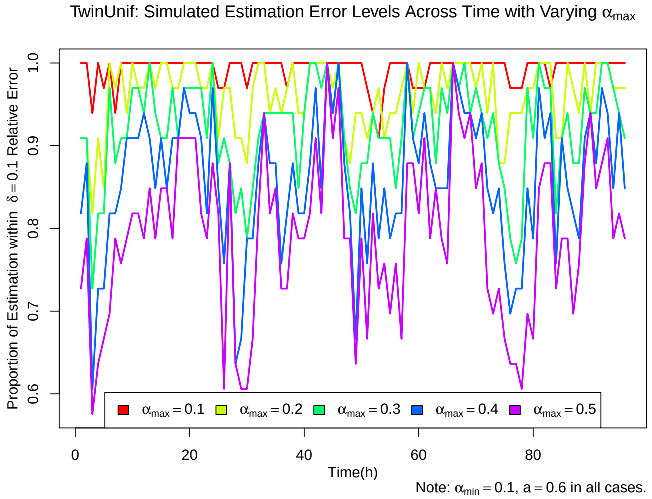
Fig. 2: Correlations for the Multiplicative scheme, corr$(\hat{Y}_{t,M}, Y_t)$.

4.2 Utility

The Smart Meter data also requires accurate estimation for the sum $S_t^c = \sum_{i=1}^{n_c} X_t^i$ over each cluster of dwellings (Privacy Requirement 1, Sec. 2). The first measure is the *probability of tolerable relative error* for sum estimation, $p_{\delta,S}$, which is estimated using the sample proportion, namely

$$\hat{p}_{\delta,S} = \frac{1}{C}\sum_{c=1}^{C} I\left(\left|\frac{\hat{s}_t^c - s_t^c}{s_t^c}\right| < \delta\right).$$

At a given level of δ, larger $\hat{p}_{\delta,S}$ indicates better utility, and perfect utility has $\hat{p}_{\delta,S} = 1$. Figure 3 suggests that multiplicative noise method achieves reasonable accuracy rates with accuracy threshold $\delta = 0.1$. As expected, larger noise (α_{max}) is associated with larger loss of utility. The cyclic behaviour of the accuracy rate corresponds to the pattern in electricity assumption, which again suggests a need for smaller α_{max} value for better utility loss.



Note: $\alpha_{min} = 0.1$, a = 0.6 in all cases.

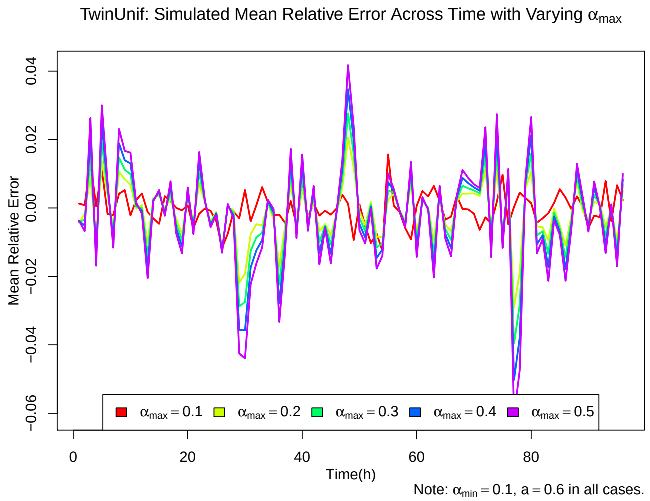
Fig. 3: Simulated sum accuracy rates $\hat{p}_\delta$ with accuracy threshold $\delta = 0.1$ with Multiplicative Noise (Twin Uniform).

The other measures of sum accuracy are the *mean relative error* (MRE) and *mean unsigned relative error* (MURE),

$$\text{MRE} = \frac{1}{C}\sum_{c=1}^{C}\frac{\hat{s}_t^c - s_t^c}{s_t^c}, \qquad \text{MURE} = \frac{1}{C}\sum_{c=1}^{C}\left|\frac{\hat{s}_t^c - s_t^c}{s_t^c}\right|,$$

where the sums are taken over all clusters $c = 1, \ldots, C$ at each time t. As explained in Section 2, distinct from the rate of error $p_{\delta,S}$, both MRE and MURE measure an average relative error. Here we focus on the MRE measure, which includes both overestimation and underestimation, the errors of opposite sign offset each other. MURE sums the absolute differences but shows similar result to MRE. The result of MURE is available upon request.

Figure 4 plots MRE for multiplicative noise scheme across time. Similar to the accuracy rates $\hat{p}_{\delta,S}$, MRE is larger for masking scheme with larger α_{max}. The cyclic pattern is less obvious comparing to the accuracy rates, although one can observe large relative errors at several time points. It again suggests that the utility loss is associated with true electricity assumption volume.



Note: $\alpha_{min} = 0.1$, a = 0.6 in all cases.

Fig. 4: Simulated mean relative error (MRE) across time with Multiplicative Noise (Twin Uniform).

4.3 Practicality

Beyond the numerical properties of privacy and utility, there are other characteristics to factor into evaluation. One of the drawbacks of the DREAM method is its vulnerability to meter failure due to the nature of paired encryption [1]. If one meter fails to collect data at a time point due to malfunction, DREAM algorithm is unable to generate shared keys, which leads to failure in encryption. The multiplicative noise method has superior robustness to technical failure. The absence of a small number of meter readings within a cluster masked under the multiplicative noise scheme will cause underestimation of S_t, but this is not catastrophic. The electricity suppliers can still form a reasonable makeshift estimate by simply scaling up the received noisy sum $\tilde{S}_t$ by a factor $n_c/(n_c - f)$, where n_c denotes the number of meters in the cluster, f the number of 'failed' or absent meters.

The multiplicative noise method is also flexible in its choice of clustering. Since each individual meter's measurements are sent independently of others, clusters may be formed and reformed using these different combinations of dwellings. Any sufficiently large cluster will have comparable sum estimation accuracy as any other large cluster.

5 Concluding Remarks

Considering the privacy and utility performance, the multiplicative noise scheme can achieve reasonable individual data privacy and adequately accurate sum estimation, but it needs careful calibration of masking parameters. The multiplicative noise method is also highly practicable for smart meters, showing simplicity in application and flexibility in defining clusters. The main challenge is that it is tricky to achieve an optimal trade-off of privacy for accurate sum estimation. The Multiplicative algorithm in full generality has great practical merit, but investigation into the optimal noise distribution M is much needed.

The valuable features of the Twin Uniform distribution are its simplicity and the capacity to eliminate the chance of correctly estimating the true values through the gap in support about its mean. Such a gap can be created using other mixture distributions beyond the uniform distribution. Figure 5 provides two possible alternatives, the Twin Triangular distribution and the Twin (Truncated) Exponential distribution. Both distributions concentrate likelihood more heavily towards the mean of the distribution, which could have potential benefit of reducing utility loss. It is of interest to investigate the impacts of these distributions on the value protection and estimation accuracy both empirically and theoretically.

Another interesting area for further research is choosing the values for masking parameters. Given that there are several masking parameters in the multiplicative noise method, an optimisation process can be proposed in order to maximize accuracy and privacy protection simultaneously. In addition, parameters are not required to be consistent across the different clusters, with the only requirement being a consistent mean μ_M across

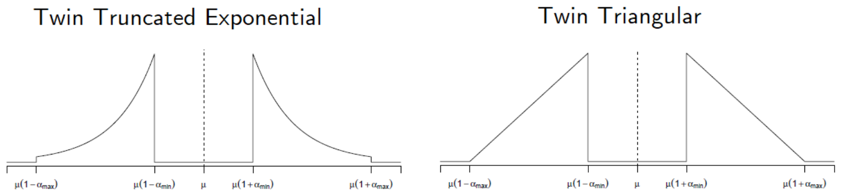


Fig. 5: Alternative options for noise distribution.

the family of distributions. By allowing variability of masking parameters between and within clusters, it is plausible to achieve greater accuracy.

References

1. Ács, G., Castelluccia, C.: I have a dream!(differentially private smart metering). Information hiding 6958, 118–132 (2011)
2. Bohli, J., Sorge, C., Ugus, O.: A privacy model for smart metering. Communications Workshops (ICC), 2010 IEEE International Conference 1, 1–5 (2010)
3. Corner, B.R., Narayanan, R.M., Reichenbach, S.E.: Noise estimation in remote sensing imagery using data masking. International Journal of Romote Sensing 24(4), 689–702 (2003)
4. Evans, B.T.: Effects on trend statistics of the use of multiplicative noise for disclosure limitation. Proceedings of the Section on Government Statistics and Section on Social Statistics, American Statistical Association (1997)
5. Evans, T., Zayatz, L., Slanta, J.: Using noise for disclosure limitation of establishment tabular data. Journal of Official Statistic 14(4), 537–551 (1998)
6. Kim, J.J., Winkler, W.E.: Multiplicative noise for masking continuous data. Proceedings of the Annual Meeting of the American Statistical Association 5(9), 1–17 (2001)
7. Laforet, F., Buchmann, E., Böhm, K.: Individual privacy constraints on time-series data. Information Systems 54, 74–91 (2015)
8. Lin, Y.X.: Protecting values close to zero under the multiplicative noise method. UNESCO Chair in Data Privacy. Lecture Notes in Computer Sciences series 11126 1, 247–262 (2018)
9. Ma, Y., Lin, Y., Sarathy, R.: The vlnerability of multiplicative noise protection to correlation- attacks on continuous microdata. Working Paper (2017)
10. Nayak, T.K., Sinha, B., Zayatz, L.: Statistical properties of multiplicative noise masking for confidentiality protection. Journal of Official Statistics 27(3), 527–544 (2011)
11. Wakefield, B., Lin, Y.X.: Efficiency and sample size determination of protected data. Privacy in Statistical Databases. UNESCO Chair in Data Privacy International Conference Proceedings 1, 263–278 (2018)