REPORT

Smart Meter Aggregation Assessment Final Report

Private and confidential

Prepared for: Energy Networks Association (ENA)

Project No: 96240
Document Version: 1.5
Date: 8 June 2015

Version History

Date	Version	Author(s)	Notes
12 March 2015	1.0	Ana Duran	Final report for release
26 March 2015	1.1	Ana Duran	Updated report following ENA review
29 April 2015	1.2	Ana Duran	Updated report following ENA second review
27 May 2015	1.3	Ana Duran	Updated report following ENA third review
5 June 2015	1.4	Ana Duran	Updated report following ENA fourth review
8 June 2015	1.5	Ana Duran	Updated report following ENA fifth review

Final Approval

Approval Type	Date	Version	EA Technology Issue Authority
Business approval	12 March 2015	1.0	Mark Sprawson
Business approval	26 March 2015	1.1	Mark Sprawson
Business approval	29 April 2015	1.2	Mark Sprawson
Business approval	27 May 2015	1.3	Mark Sprawson
Business approval	5 June 2015	1.4	Mark Sprawson
Business approval	8 June 2015	1.5	Mark Sprawson

EA Technology Limited, Capenhurst Technology Park, Capenhurst, Chester, CH1 6ES;
Tel: 0151 339 4181 Fax: 0151 347 2404

http://www.eatechnology.com

Registered in England number 2566313

Executive summary

Background

This document summarises the content of the full Smart Meter Aggregation Assessment Final Report produced by EA Technology on behalf of the ENA to assess the relationship between data aggregation and privacy. A further report is being prepared to assess the relationship between data aggregation and the delivery of network benefits.

Consumer behaviour and electricity demand on the low voltage (LV) network are anticipated to change dramatically over the coming years due to electrification of heat and transport, decarbonisation of electricity production and widespread take-up of micro-generation, posing a significant challenge to electricity Distribution Network Operators (DNOs). Having increased visibility of demands on the LV network via the roll-out of smart meters to all customers could be of material benefit to DNOs in assisting them manage their networks and plan reinforcement. In turn, the use of smart meter data will benefit end customers as a consequence of DNOs being able to make more informed decisions and hence invest more efficiently in the network. Previous ENA work has shown the likely benefits to be £27.9m over ED1 and a further £41.3m over ED2, although this is dependent upon the realisation of the forecast increases in Low Carbon Technologies (LCT).

Under licence condition SLC10a DNOs are unable to access raw load profile data (time series consumption data) from individual smart meters due to concerns over personal privacy issues with customers. This project has been instigated to investigate the relationship between aggregation and anonymity. A second project has been instigated to evaluate the reduction in financial benefits as a consequence of DNOs not being able to make use of individual consumption profiles from individual smart meters to inform their network investment decision process and/or if anonymity concerns result in aggregation levels that prevent access to sufficiently granular consumption data.

This paper summarises the work to assess the level of aggregation required to achieve a high degree of anonymity. It is important to note that aggregation on its own cannot ensure anonymity; for example the aggregate of consumption profiles from customers A, B and C subtracted from the aggregate of consumption profiles from customers A, B, C and D will produce the consumption profile for customer D. This simple example illustrates that whilst aggregation can help provide anonymity, DNOs will need to build controls into their IT systems and/or business processes to preclude the possibility of aggregated data being abused. Hence, the objective of this work was to establish the relationship between the number of consumption profiles aggregated and the likelihood of being able to establish the consumption profile of an individual customer from this aggregated profile.

Establishing an individual profile from an aggregated profile

This section describes the methodology adopted and the results obtained from the studies performed to assess the possibility of being able to recreate an individual consumer consumption profile from a profile formed by aggregating different numbers of consumption profiles.

To model this it was necessary to establish some typical LV demand profiles representative of individual household load consumption. For this EA Technology used a well-established model which simulates over a 24 hour period the use of domestic appliances within UK dwellings. This model, called the CREST model[1], was used to create individual household half hourly load consumption data, representative of ten real feeders with the number of customers on each feeder ranging from 9 to 124.

[1] Centre for Renewable Energy Systems Technology at Loughborough University - http://www.lboro.ac.uk/research/crest/

These modelled profiles were aggregated, at feeder level, and validated against the feeder monitoring data from the Northern Powergrid Customer Led Network Revolution project (CLNR)[2].

Once the profiles were confirmed as being representative of real LV networks, they were subject to consecutive aggregated customer privacy studies using three different methods of analysis:

- **Method 1 – Visual inspection:** graphically compared the aggregated half hourly profiles. Aggregation was undertaken in increased order from 2 customers up to the total number of customers on the feeder and the optimum aggregation value to preserve anonymity was one that visually showed appreciable differences to the previous aggregated profile but aggregated profiles thereafter showed a lower degree of graphical variations.

- **Method 2 – Correlation analysis:** investigated how correlated an additional profile was to a group of aggregated profiles. The correlation coefficient between profiles was used to numerically quantify the similarities between profiles and to calculate the extra differentiation that the addition of a further profile could add into the group.

- **Method 3 – Clustering analysis:** used a K-means clustering approach that determined the average number of customers with household load profiles similar to that of the group, and how likely an individual customer load profile could be estimated from the aggregated group load profile.

These methods were assessed considering typical conventional network loads under balanced load conditions and results from each methodology were considered. The review concluded that the visual inspection method provided less objective results than the correlation analysis, while the clustering analysis offered less granular comparisons than the correlation analysis.

For these reasons the correlation analysis was the chosen methodology to be taken forward for the detailed analysis, examining 330 different future load cases and assessing the effects of LCT and uncertainties associated with customers' phase connectivity which affects load balance on the network.

These 330 analyses were obtained from 33 LCT penetration scenarios analysed over 10 different feeders. The LCTs considered for the analysis were: photovoltaics (PV), Heat Pumps (HP) and Electric Vehicles (EV) – both fast and slow charging points. The 330 analysed cases covered situations with domestic customers without any LCTs and included several combinations of increasing LCT penetration up to a maximum number of customers with LCT loads[3] and analysing winter/summer seasonal variations on both suburban and rural feeders.

In the following table the visibility risk is presented as a key metric in the evaluation of suitable aggregation levels. Visibility risk is defined as the likelihood of an individual customer consumption profile being derived from the aggregated group load profile. In other words, if someone had access to the aggregated profile, what would be the probability of deriving one individual profile from it. Hence the lower the visibility risk the greater the customer privacy. The visibility risk has been calculated as the median of the 330 results obtained from the analysed cases and results are presented in Table 1.

[2] Northern Powergrid feeders were deemed to be appropriate for the analysis having already been investigated for their national representation on the Low Carbon Network Fund (LCN Fund) project and having had monitoring systems fitted in place which allowed the simulated load consumption data to be validated against real feeders.
[3] Based on DECC's Credit Purchase scenario projections for 2020 - this scenario bears the closest resemblance to the uptake of LCTs expected by DNOs in the RIIO-ED1 period following the results of stakeholder engagement activities undertaken by all DNOs.

Table 1 Visibility risk for each aggregation level

Aggregation level	Visibility risk (%) (median correlation results)
1	100%
2	22%
3	20%
4	17%
5	15%

Figure 1 below shows that an aggregation level of two reduces the visibility risk by approximately 80% and very little improvements for increasing levels of aggregation happen thereafter.

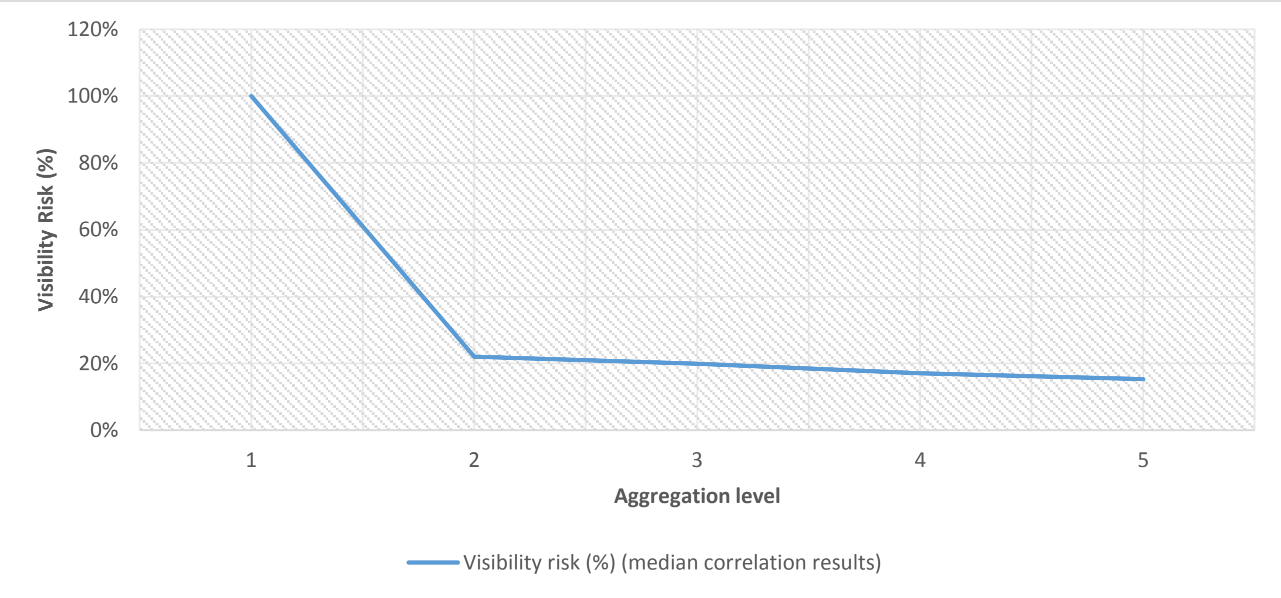

Figure 1 Visibility risk for each aggregation level

The detailed analysis shows that aggregating two consumption profiles reduces the possibility of being able to establish an individual customer's consumption profile from 100% to 22%, and that further aggregation would provide only a marginal reduction in visibility risk. Hence aggregation of two profiles, coupled with the development and implementation of DNO IT systems and/or business processes, to address the inherent flaws in using aggregation to ensure anonymity, is proposed as being the approach to meet the requirements of SLC10a.

Contents

Figures

Tables

Appendices

1. Background & Introduction

Consumer behaviour and electricity demand on the low voltage (LV) network are anticipated to change dramatically over the coming years due to electrification of heat and transport, decarbonisation of electricity production and widespread take-up of micro-generation, posing a significant challenge to electricity Distribution Network Operators (DNOs). Having increased visibility of demands on the LV network via the roll-out of smart meters to all customers could be of material benefit to DNOs in assisting them manage their networks and plan reinforcement. In turn, the use of smart meter data will benefit end customers as a consequence of DNOs being able to make more informed decisions and hence invest more efficiently in the network. Previous ENA work has shown the likely benefits to be £27.9m over ED1 and a further £41.3m over ED2, although this is dependent upon the realisation of the forecast increases in Low Carbon Technologies (LCT).

The roll-out of smart meters to all customers is one means by which this greater level of visibility may be achieved. However, it is to be noted that under licence condition SLC10a DNOs are unable to access raw load profile data (time series consumption data) from individual smart meters due to concerns over personal privacy issues with customers.

Although DNOs will receive non-aggregated (defined as attributable to an identifiable customer) load profile data, DNOs are obliged to aggregate the data as soon as possible and can only store and analyse the data in an aggregated manner.

As such, this project has been instigated to investigate the relationship between aggregation and anonymity. A second project has been instigated to evaluate the reduction in financial benefits as a consequence of DNOs not being able to make use of individual consumption profiles from individual smart meters to inform their network investment decision process and/or if anonymity concerns result in aggregation levels that prevent access to sufficiently granular consumption data.

2. Scope and Objectives

This project establishes the relationship between the number of consumption profiles aggregated and the likelihood of being able to establish the consumption profile of an individual customer from this aggregated profile. As such, the scope and objectives of this project are summarized in sections 2.1 and 2.2 below.

2.1 Scope of project

The scope of this project was to analyse suburban and rural feeders that have a high proportion of domestic customers.

The study assessed the impact of today's load and future load at different points on the LV network where different levels of load imbalances for the ED1 period were also examined.

2.2 Objective of project

The objective of this project was to develop an assessment of the minimum number of consumers' data sets that should be aggregated to provide a high degree of anonymity, while providing DNOs with the highest possible network visibility for planning their investment decisions if smart meter data can only be assessed in an aggregated manner.

3. Analysis Overview

To determine the minimum required number of domestic smart meters that should be aggregated to ensure customers' privacy, it was necessary to establish some typical LV demand profiles representative of individual household load consumption. For this EA Technology used a well-established model which simulates over a 24 hour period the use of domestic appliances within UK dwellings. This model, called the CREST model[4], was used to create individual household half hourly load consumption data, representative of ten real feeders with the number of customers on each feeder ranging from 9 to 124.

These modelled profiles were aggregated, at feeder level, and validated against the feeder monitoring data from the Northern Powergrid Customer Led Network Revolution project (CLNR).

Once the profiles were confirmed as being representative of real LV networks, they were subject to consecutive aggregated customer privacy studies using three different methods of analysis:

- Method 1 – Visual inspection
- Method 2 – Correlation analysis
- Method 3 – Clustering analysis

These methods were initially assessed considering today's network loads with balanced conditions and then further explored with a wide range of low carbon technology (LCT) penetration scenarios and the uncertainties associated with customers' phase connectivity.

3.1 Data selection, compilation and validation

To avoid anonymity issues related to licence condition SLC10a, individual household half hourly load profiles (typical conventional network loads), were generated using the CREST model (Centre for Renewable Energy Systems Technology).

The CREST model is a high-resolution model of domestic whole house electricity demand which simulates over a 24 hour period, at a one minute time resolution, the use of domestic appliances within a single UK dwelling. The model uses the appliances as the basic building blocks, where "appliances" refers to any individual domestic electricity load, such as washing machine, vacuum cleaner, kettle, television etc. The appliances in the model are configured using statistics describing their mean total annual energy demand and associated power use characteristics, including steady-state consumption or typical use cycles as appropriate. The simulation also incorporates and utilises previously developed models of active occupancy and domestic lighting and requires the input of other parameters such as the month of the year, the household consumption type (domestic economy 7 or domestic unrestricted), the average number of residents that live at the dwelling and whether a weekday or a weekend day simulation is required.

EA Technology modified the CREST model to allow it to simulate multiple household profiles on each run and estimate the electricity demands on half hour periods. This enabled the generation of 30 minute time resolution profiles for each of the households in a feeder.

From all the customizable parameters in CREST, several needed to be selected up front. It was determined that the average household occupancy level was three and hence this was selected for the simulations. The time of year at which to perform the analysis was selected to be weekdays in January, due to this being the period of the year, together with December, when maximum demands occur. Information regarding the number of dwellings per feeder and the individual household consumption type were taken from the selected suburban and rural feeders pertaining to the Northern Powergrid network. These feeders, containing terraced streets, 3/4 bed semi-detached or

[4] Centre for Renewable Energy Systems Technology at Loughborough University - http://www.lboro.ac.uk/research/crest/

detached houses, rural villages and farms were chosen because they have a high proportion of domestic customers and represent around 60% of the LV circuits in Great Britain (GB)[5].

Northern Powergrid feeders were deemed to be appropriate for the analysis having already been investigated for their national representation in the Customer Led Network Revolution project and having had monitoring systems fitted in place which allowed the simulated load consumption data to be validated against real feeders. Table 2 shows the amount of feeders analysed in this smart meter aggregation project under each feeder type. Further details can be found in Appendix I

Table 2 Feeders selected for the smart meter aggregation analysis

LV Feeder type	Household type	Proportion of LV circuits in GB (Transform Model)	CREST model analysed feeders
Suburban	Terraced Street	38.8%	2
	3/4 bed semi-detached or detached houses	15.2%	6
Rural	Rural village / farms	5.5%	2

Feeder validation was performed via comparison of the feeder profiles obtained from the Customer Led Network Revolution iHost system and the aggregation of the individual household half hourly consumption profiles from the CREST model. This was done for a 24 hour period. All modelled data was comparable to the ten CLNR (iHost) feeder profiles with two exceptions:

- Case 1 – The aggregated CREST profiles showed a lower consumption than the iHost profile due to lack of data regarding the customer type. These customers were not taken into consideration in the CREST model.
- Case 2 – The aggregated CREST profiles showed a higher consumption than the iHost profile because of incomplete iHost monitoring data. The CLNR monitoring system did not take into consideration all the customers on the feeder.

Figure 2 shows a graphical representation of the feeder load validation in kW for one of the selected suburban feeders (Wooler Ramsey - Feeder B). The blue and green trends correspond to the feeder loads obtained from the aggregated individual household modelled data from the CREST model and the Northern Powergrid CLNR iHost monitoring data respectively.

[5] The proportion figures related to LV suburban and rural feeder circuits in Great Britain (GB) have been taken from the Transform Model.

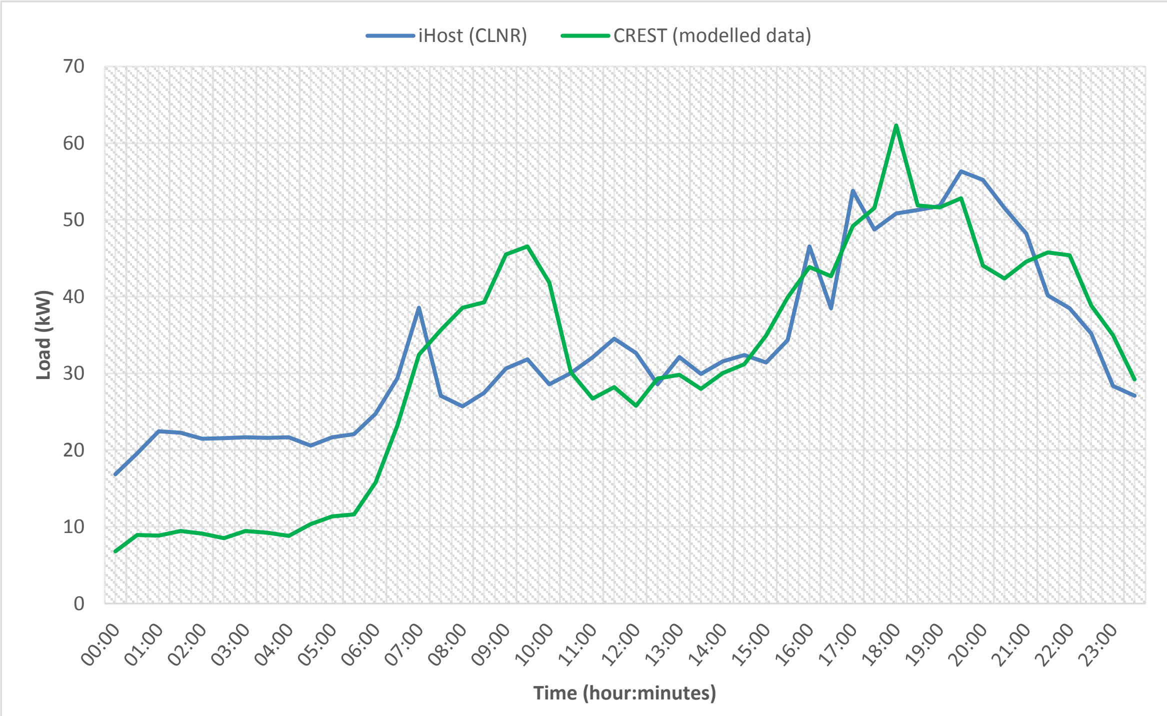

Figure 2 **Modelled data validation**

4. Data analysis

Once the data obtained from the CREST model was validated as being representative of real consumers on LV network feeders, EA Technology proceeded to analyse the data with the effect that today's and the forecast future load, based on the projected uptake of LCTs, have on the required aggregation level of smart meter data. This also took into consideration the uncertainties associated with unbalanced load connections. In determining the appropriate level of LCT uptake, the governmental scenario that most closely represents the views of the DNOs following their stakeholder engagement consultations for their RIIO-ED1 business planning purposes was selected. This is sometimes referred to as 'DECC Scenario 4' or 'Credit Purchase'.

In this way, the analysis was split up in two parts described in section 4.1 and 4.2. The first part of the analysis (section 4.1) investigated the appropriateness of the methodologies proposed for the privacy study with today's load under balanced conditions, and the second part of the analysis (section 4.2) progressed with those methodologies that provided more robust and objective outputs, and expanded the initial results to accommodate the effect of LCT penetration scenarios and the material effect that load imbalance may have on the aggregation level.

4.1 Today's network loads with balanced conditions

This section describes the investigations undertaken to examine the data from the CREST model (individual household half hourly consumption profiles), representative of each individual household on the feeders, under three different analysis methodologies. The explored methods were:

- Method 1 – Visual inspection
- Method 2 – Correlation analysis
- Method 3 – Clustering analysis

All three methodologies assessed how similar or dissimilar a profile is to the increased aggregation of individual household profiles for the analysed feeders[6].

4.1.1 Method 1 – Visual inspection

This methodology was based on a graphical comparison of the average normalised aggregated half hourly profiles. The profiles were aggregated in increased order from 2 customers up to the total number of customers on the feeder and the optimum aggregation value to preserve anonymity was that one that visually showed appreciable differences to the previous aggregated profile but aggregated profiles thereafter showed a lower degree of graphical variations.

This methodology tried to find an aggregation level where peaks in consumption were smoother to a level such that the fewer the differences with further aggregated profiles, the higher the difficulty to derive an individual household profile from the aggregated one and vice versa. At the same time this optimum aggregation level was selected such that the aggregation of further profiles thereafter did not provide meaningful differences that justified a higher aggregation level.

Figure 3 shows the visual comparison of average aggregated profiles for 2, 6, 7, 8 and 74 customers for one of the feeders under investigation (Wooler Ramsey - Feeder B). 74 is the total number of customers in the feeder, 2 the minimum number of aggregated profiles, 7 is the optimum aggregation profile for this particular feeder and the aggregation of 6 and 8 customers is only represented for comparison purposes with the aggregation profile of 7 customers.

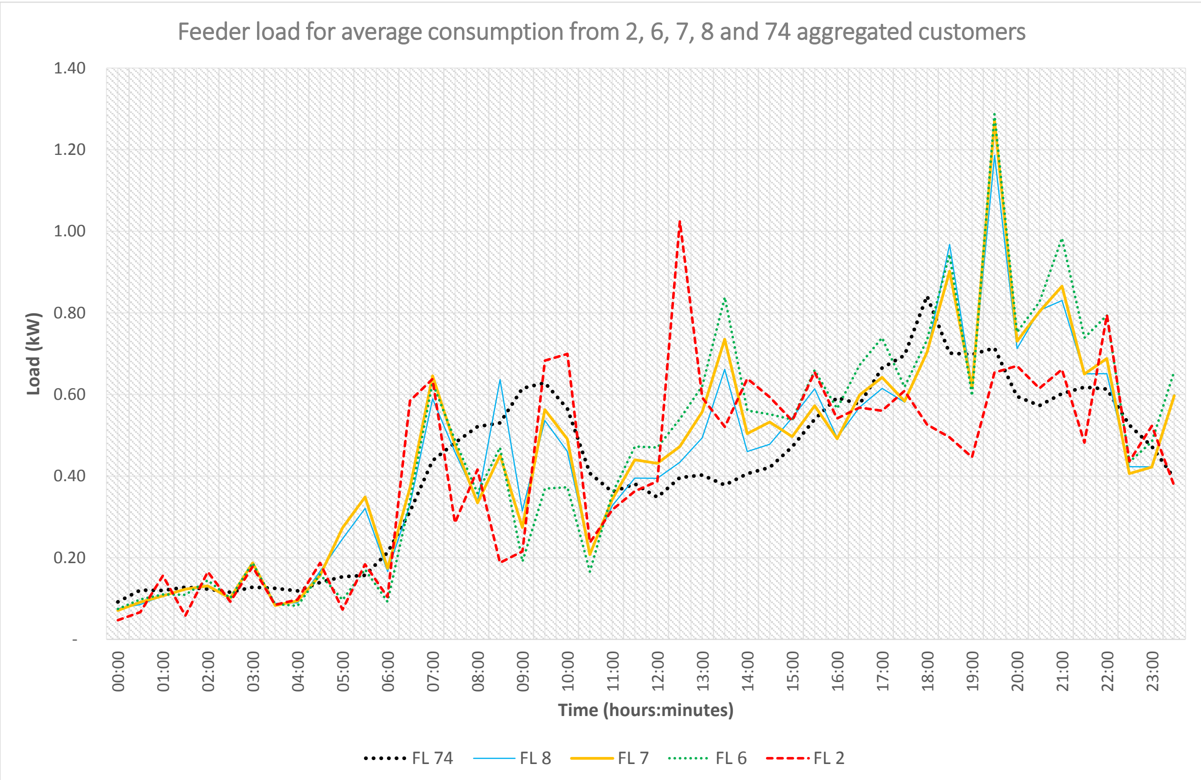

Figure 3 Method 1 - Visual inspection for smart meter aggregation to ensure customer's privacy

It can be seen in Figure 3 how the aggregation of 6 customers (FL6) follows very closely the aggregation of 2 customers' profile (FL2) up to 07:00, between 14:30 and 16:00 and from 22:00 onwards. The aggregation of 7 customers (FL7) starts following a more generic profile, smoother

[6] Method 1 and method 2 performed the analysis on each of the 10 selected feeders whereas method 3 was performed on a more reduced data set due to the complexity of the analysis methodology.

and more similar to the total feeder aggregation profile (FL74) and the aggregation of 8 customers' profiles (FL8) follows a similar pattern when compared to FL7.

4.1.2 Method 2 – Correlation analysis

This methodology investigated how correlated an additional profile was to a group of aggregated profiles. The correlation coefficient between profiles was used to numerically quantify the similarities between profiles. This correlation coefficient varied between -1 and 1, with -1 representing a perfect decreasing (indirect) linear relationship, 1 a perfect increasing (direct) linear relationship and 0 uncorrelated data.

The closer the correlation coefficient to 0 the better for the privacy analysis as it meant that the profiles under analysis did not share points in common that could be attributable to a single profile.

In order to determine the level at which the correlation coefficient was low enough as to decide on the minimum required aggregation level, a ratio between correlation coefficients was analysed so that, in percentage terms, it showed the extra differentiation that the addition of a further profile could add into the group. The optimum aggregation level was one such that by adding thereafter further profiles into the group they would only reduce the similarities in between the profiles by a maximum of 10%.

The methodology is further explained in Figure 4 and related comments below:

Cn	1	2	3	4	5	6	7	8	...	15
CORREL 1	1.00									
CORREL 2	0.59	0.93								
CORREL 3	0.49	0.80	0.65							
CORREL 4	0.47	0.76	0.60	0.47						
CORREL 5	0.41	0.72	0.55	0.40	0.70					
CORREL 6	0.43	0.65	0.55	0.40	0.71	0.69				
CORREL 7	0.37	0.63	0.55	0.32	0.69	0.69	0.67			
CORREL 8	0.36	0.65	0.55	0.33	0.71	0.67	0.62	0.68		
CORREL 9	0.30	0.67	0.55	0.33	0.70	0.60	0.64	0.65		
CORREL 10	0.31	0.68	0.51	0.36	0.66	0.59	0.62	0.64		
CORREL 11	0.30	0.67	0.50	0.38	0.67	0.62	0.58	0.66		
CORREL 12	0.34	0.64	0.50	0.40	0.65	0.61	0.53	0.69		
CORREL 13	0.31	0.62	0.48	0.41	0.65	0.61	0.52	0.68		
CORREL 14	0.33	0.60	0.47	0.46	0.63	0.62	0.49	0.66		
CORREL 15	0.31	0.58	0.49	0.44	0.64	0.61	0.53	0.65		0.47

	% differences										Average
CORREL 1	69%										69%
CORREL 2	47%	37%									42%
CORREL 3	36%	27%	24%								29%
CORREL 4	34%	24%	19%	7%							21%
CORREL 5	25%	19%	11%	-9%	9%						11%
CORREL 6	**28%**	**11%**	**10%**	**-10%**	**9%**	**11%**					**10%**
CORREL 7	16%	7%	12%	-36%	7%	12%	20%				5%
CORREL 8	14%	10%	10%	-33%	9%	8%	14%	5%			5%
CORREL 9	-4%	13%	10%	-33%	8%	-1%	17%	0%			5%
CORREL 10	-1%	14%	4%	-21%	2%	-3%	13%	0%			3%
CORREL 11	-4%	13%	2%	-15%	4%	1%	8%	2%			2%
CORREL 12	9%	8%	1%	-9%	2%	0%	-1%	7%			2%
CORREL 13	1%	6%	-2%	-7%	1%	0%	-3%	5%			0%
CORREL 14	7%	3%	-4%	4%	-2%	1%	-9%	3%			1%
CORREL 15	0%	0%	0%	0%	0%	0%	0%	0%	...	0%	0%

Figure 4 Correlation analysis - Calculation of % differentiation with increased aggregation

Figure 4 (left) shows the correlation coefficients obtained for each aggregation group. This has been calculated in such a way that each cell represents the correlation coefficient between each customer's HH load profile (the customer number is defined in the columns) and the aggregated HH load profile of the customers in that group (the group of customers is specified in the rows). Existing customers' HH load profiles in the group were maintained (not replaced by new customers' HH load profiles) every time a new customer's HH load profile was added into the group. This was done to check the effect that only one change at a time had into the customers' privacy evaluation.

To further clarify this, following are presented a few examples:

- "row: correlation 1, column: 1 = 1.00" represents the correlation coefficient between the HH load profile of customer 1 and the aggregated profile of the group of this customer. 1 means that similarity between customer 1 and the aggregated profile of this customer is 100%
- "row: correlation 2, column: 1 = 0.59" represents the correlation coefficient between the HH load profile of customer 1 and the aggregated profile of the group of these 2 customers, where customer 1's HH load profile is the profile selected for the correlation coefficient above described. 0.59 means that similarity between customer 1 and the aggregated profile of these 2 customers is 59%
- "row: correlation 2, column: 2 = 0.93" represents the correlation coefficient between the HH load profile of customer 2 and the aggregated profile of the group of these 2 customers, where customer 1's HH load profile is the profile selected for the previously mentioned

correlation coefficients. 0.93 means that similarity between customer 2 and the aggregated profile of these 2 customers is 93%

- "row: correlation 3, column: 1 = 0.49" represents the correlation coefficient between the HH load profile of customer 1 and the aggregated profile of the group of these 3 customers, where customers 1's and 2's HH load profiles are the profiles selected for the previously mentioned correlation coefficients. 0.49 means that similarity between customer 3 and the aggregated profile of these 3 customers is 49%

On the other side Figure 4 (right) shows:

a) the ratio between each customer's correlation coefficient, for each group of customers, and that same customer's correlation coefficient for the aggregated group of 15[7] customers, e.g.:

$$69\% = (1 - \frac{0.31}{1}) \times 100 \quad ; \quad 47\% = (1 - \frac{0.31}{0.59}) \times 100 \quad ; \quad 37\% = (1 - \frac{0.58}{0.93}) \times 100$$

b) the average of the ratios for each of the aggregation groups. These average figures represent the further differentiation, in percentage terms, that each increased profile aggregation adds into the group. This is shown in Figure 5 below for one of the feeders under investigation (Wooler Ramsey - Feeder B).

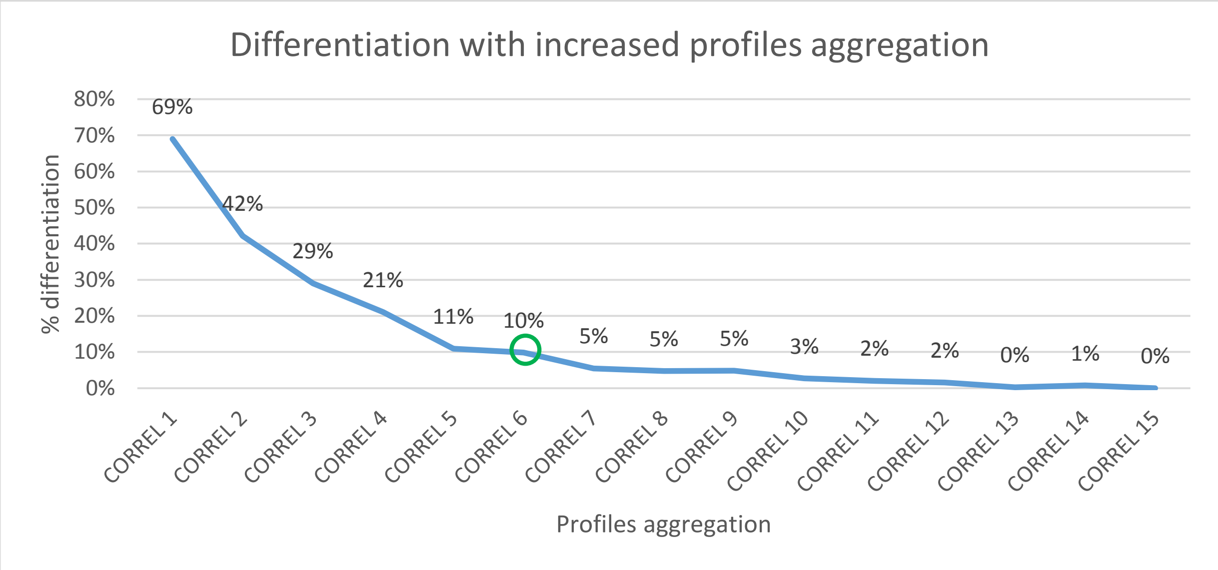

Figure 5 Correlation analysis - Added % differentiation with increased aggregation

The graph above indicates that the aggregation of 2 profiles (correlation 2), for example, in comparison with the aggregation of 15 profiles (correlation 15) has a 42% improvement scope for differentiating the profiles. It also indicates that by adding 4 more profiles (correlation 6), the differentiation level in between the 6 profiles and the group would drop down to 10% meaning that by selecting 6 profiles in this feeder, they would only be 10% similar to the aggregated profile.

4.1.3 Method 3 – Clustering analysis

This methodology assessed the similarities of the individual household profile of a group of *n* consumers, with *n* being a number between 2 and 10 consumers, to the aggregated load profile of the group.

[7] The highest smart meter data aggregation level obtained in Method 1 – Visual inspection, that suggested a high customers' privacy level for all the 10 analysed feeders, was 10. It was assumed for Method 2 – Correlation analysis that 1.5 times the maximum aggregation level obtained in Method 1, 15 customers, was a sufficiently wide group of profiles to which the addition of further profiles into the group would not improve the consumer's anonymity significantly further.

The analysis was performed 10 times for each of the 10 groups of profiles, each group containing:

- Group of 2 customers: 2 individual household load profiles and their aggregated profile
- Group of 3 customers: 3 individual household profiles and their aggregated profile
- …
- Group of 10 customers: 10 individual household profiles and their aggregated profile

The profiles selected from the CREST model pertained to only one of the feeders under investigation, Wooler Ramsey - Feeder B, as opposed to the other two methods that analysed all the 10 selected feeders. The profiles were selected randomly for each of the 10 times that the groups of *n* customers were analysed and all profiles had the same chances of being picked up for each of the group analysis. This meant that each group of 2 customers, for example, always had two different individual profiles picked up plus their aggregated profile, but each of the 10 times this group of 2 customers analysis was performed these same profiles could have been picked up again.

The total number of profile combinations therefore analysed in the cluster analysis was 630 and the total analysed cases were 100.

The method used a K-means clustering approach which determined the average number of customers with household load profiles similar to that of the group. This average number indicated how likely an individual customer load profile could be estimated from the aggregated group load profile.

Results from the clustering analysis showed that, for the feeder under investigation, the average percentage of consumer load profiles that clustered with the group of 4 was 3%. Beyond the threshold of 4 consumers, the percentage of consumer profiles clustering with the aggregated profile of the group showed very little and slow decrease. This means that from the ten tests performed for each of the consumer groups, the average anonymity level of each load profile compared to each group load profile was 97% or higher; or in other words, the average likelihood of a profile being estimated from the aggregated group load profile for groups of 4 customers or above, was 3% or lower. This is represented in Table 3 below.

Table 3 Clustering analysis - Consumer aggregation threshold

Number of consumer per group	Percentage of consumers similar to the group
2	35%
3	7%
4	3%
5	3%
6	3%
7	1%
8	0%
9	0%
10	0%

The clustering methodology has been previously used in smart meter, energy consumption and electricity demand profile studies, amongst others, and links showing examples where this analysis has been used within the industry are provided in Appendix II.

4.1.4 Preferred analysis method

The three methods were assessed considering typical conventional network loads with balanced conditions and results from each methodology were considered.

The visual inspection method graphically showed the differences in the aggregated consumption profile as individual profiles were added into the group. Due to the fact that results were only visually assessed, they were considerably sensitive to the order in which the aggregated profile was constructed. Given that their interpretation was heavily dependent on the analyser's view it was considered that the conclusions from this methodology were too subjective.

The correlation analysis method showed in percentage terms (i.e. quantitatively) the difficulty for an individual/system to derive a specific customer profile from an aggregated group of profiles. Not only did this methodology provide numerically quantified results, but it also showed the progressive codifying effect that extra profiles added into the group provided to the aggregated profile (with a lesser impact ascribed to the order of aggregation).

The clustering methodology showed also in quantifiable terms the similarities between each individual profile pertaining to a group and the aggregated profile of that group. The higher the similarities, the easier for an individual/system to derive an individual profile from the group. The main difference between the correlation analysis and this methodology was that this technique was randomly selecting profiles from a pool of profiles for each of the 10 tests performed for each cluster (i.e. groups of 2, 3, 4, 5 etc. profiles) but there was no correlation between the profiles selected for each of the ten tests performed for each cluster, nor for the consecutive tests for the following aggregated clusters.

Taking into consideration the above points the review concluded that the correlation methodology provided more objective results than the visual inspection and it also considered the similarities between each profile and the consecutive aggregation of profiles, in contrast to the clustering analysis, which only contemplated the similarities between each profile and the aggregation of the group profiles. For these reasons the correlation analysis was the chosen methodology to be taken forward for the next set of studies.

4.2 Forecast network loads by 2020 and effects associated with unbalanced load conditions

The sections below describe the different LCT scenarios examined under the correlation analysis and the effect that they have considering the different feeder types, in addition to load imbalance conditions, on the smart meter aggregation level.

4.2.1 Low carbon technologies and scenarios under investigation

The low carbon technologies considered for this analysis were:

- Photovoltaics (PV): Half hourly load profiles have been taken from the Transform Model. These figures have been updated to reflect a maximum demand of 2.5kW in summer.
- Heat pumps (HP): It has been assumed for this analysis that HP loads have a flat daily consumption of 10kW during winter and 1kW during summer.
- Electric Vehicles (EV): Two different EV charging points have been considered:
    - Slow EV charging points (EVs): It was supposed that these charging points consume 3.5kW between 18:30 and 23:30
    - Fast EV charging points (EVf): It was supposed that these charging points consume 7kW between 18:00 and 21:00

These LCTs have been assessed for winter and summer periods and their penetration rates to 2020 followed those predicted by the previously discussed 'Credit Purchase' scenario for LCT uptake as outlined by the Department of Energy and Climate Change (DECC).

The above described loads in conjunction with today's loads were therefore studied in the correlation analysis for summer and winter periods, 33 different LCT scenarios and ten different feeder tests. These add up to a total of 330 analysis which are all following described.

Focusing on the 33 LCT scenarios first, Figure 6 shows the distribution of profiles, with and without LCTs, analysed for each scenario (highlighted in green).

Scenario	Today's custoner load (No LCTs)	Added customers with PV loads	Added customers with HP loads	Added customers with EV slow loads	Added customers with EV fast loads	Description
1	15					No LCT customers
2	15	15				Max number of customers having 1 LCT at a time
3	15		15			
4	15			3		
5	15				6	
6	15	15	15			Max number of customers having PV and another max LCT at a time
7	15	15		3		
8	15	15			6	
9	15		15	3		Max number of customers having HP and another max LCT at at time
10	15		15		6	
11	15	15	15	3		Max number of customers having 3 LCTs at a time
12	15	15	15		6	
13	15	7	10	1	1	Scenario 4 and some customers having more than one LCT
14	15	6	8		1	Scenario 4 but each customer only having one LCT
15	15	6	8		1	As per 14) but some customers having more than one LCT
16	15	6	2		1	As per 15) but with a reduced number of HP
17	15	3	3	3	3	Alternative LCTs (none, PV, HP, EVs, EVf, none, PV, HP, EVs, EVf etc.)
18	4	1				Average aggregation level customers with one LCT at a time
19	4		1			
20	4			1		
21	4				1	
22	15	1				15 customers with one LCT at a time
23	15		1			
24	15			1		
25	15				1	
26	15	2				15 customers with the same two LCTs at a time
27	15		2			
28	15			2		
29	15				2	
30	15	3				15 customers with the same three LCTs at a time
31	15		3			
32	15			3		
33	15				3	

Figure 6 33 LCT scenarios

All customers' HH profiles had today's load as a base (without LCTs) and LCTs were allocated as per Figure 6. To give some examples:

- Scenario 1 analysed 15 customers' HH profiles with today's loads (No LCTs),
- Scenario 2 analysed 15 customers' HH profiles with today's load and PV HH profiles added into all of them,
- Scenario 4 analysed 15 customers' HH profiles with today's load and three customers had the EV slow load added into them.
- Scenario 13 analysed 15 customers' HH profiles with today's load where seven customers had PV, ten had HP, one had EV slow and one had EV fast loads added into them.

When customers had EV slow and fast loads allocated within the same scenario, it would not have been realistic that a same customer had two different charging points, and therefore it was assumed that EV slow loads were allocated to the first customers' HH profiles and following customers' profiles had the EV fast loads added into them.

The 33 scenarios covered situations with domestic customers without any LCTs and included several combinations of increasing LCT penetration up to a maximum number of customers with LCT loads[8]. However, these were not the only analysed cases in this project.

Going a step forward, Table 4 shows the 10 different feeders that were also analysed for each of the 33 scenarios. These were categorised per household consumption and feeder type and seasonal load variations.

Table 4 Feeder test types

Customer consumption type	All unrestricted customers		14% economy 7 customers		25% economy 7 customers	All economy 7 customers
Feeder type	Suburban semidetached or detached feeder		Suburban terraced feeder		Suburban semidetached or detached feeder	Rural feeder
Season	Winter	Summer	Winter	Summer	Winter	Winter
Tests	Test A, B and C	Test D	Test A	Test D	Test A, B and C	Test A

- Test A, B and C used each one 15 different customers' HH profile data sets, without LCTs, and had the winter LCT HH load profiles added into them as per scenarios 1 to 33.
- Test D used the same Test A customers' HH profile data sets, without LCTs, but had the summer LCT HH load profiles added into them as per scenarios 1 to 33.

In summary, 330 different correlation analysis were performed, taking into account all the 33 scenarios and 10 feeder tests per scenario. The complete analysis matrix is presented below for clarification purposes.

Scenario	Today's custoner load (No LCTs)	Added customers with PV loads	Added customers with HP loads	Added customers with EV slow loads	Added customers with EV fast loads	Description	10 Feeder tests A - D per scenario
1	15					No LCT customers	
2	15	15				Max number of customers having 1 LCT at a time	
3	15		15				
4	15			3			
5	15				6		
6	15	15	15			Max number of customers having PV and another max LCT at a time	
7	15	15		3			
8	15	15			6		
9	15		15	3		Max number of customers having HP and another max LCT at at time	
10	15		15		6		
11	15	15	15	3		Max number of customers having 3 LCTs at a time	
12	15	15	15		6		
13	15	7	10	1	1	Scenario 4 and some customers having more than one LCT	
14	15	6	8		1	Scenario 4 but each customer only having one LCT	
15	15	6	8		1	As per 14) but some customers having more than one LCT	
16	15	6	2		1	As per 15) but with a reduced number of HP	
17	15	3	3	3	3	Alternative LCTs (none, PV, HP, EVs, EVf, none, PV, HP, EVs, EVf etc.)	
18	4	1				Average aggregation level customers with one LCT at a time	
19	4		1				
20	4			1			
21	4				1		
22	15	1				15 customers with one LCT at a time	
23	15		1				
24	15			1			
25	15				1		
26	15	2				15 customers with the same two LCTs at a time	
27	15		2				
28	15			2			
29	15				2		
30	15	3				15 customers with the same three LCTs at a time	
31	15		3				
32	15			3			
33	15				3		

Figure 7 330 correlation analysis containing 33 scenarios and 10 feeder tests per scenario

[8] Based on DECC's Credit Purchase scenario projections for 2020 - this scenario bears the closest resemblance to the uptake of LCTs expected by DNOs in the RIIO-ED1 period following the results of stakeholder engagement activities undertaken by all DNOs

An optimum smart meter aggregation figure was obtained from each of the 330 correlation analysis. This figure was calculated as the ratio between correlation coefficients so that, in percentage terms, it represented the extra differentiation that the addition of a further profile added into the group. This analysis followed the same methodology explained in section 4.1.2 with the only difference being that the selected aggregation level, this time, was one such that by adding thereafter further profiles into the group they would only reduce the similarities in between the profiles and the group profile by a maximum of 15% of the total aggregated profile.

The selection of 15% for this analysis is an arbitrary figure that was deemed by EA Technology at this stage to be a 'reasonable' level of risk for this portion of the analysis. It is not, however, necessarily a reflection of the 'optimum' level of risk.

4.2.2 Correlation analysis conclusions considering LCTs

Analysing the average results for all the 330 tests, applicable to customers with an 85% or higher differentiation on their profiles, it was concluded that:

- 3.6 customers can be aggregated on average for unrestricted consumers, whereas Economy 7 customers increase the average aggregation level to 5.54
- Economy 7 customers with high PV and/or HP penetration rates, that also have electric vehicles installed, increase the privacy aggregation number from 5.54 to 7
- Customers without LCTs, for more than 90% of the analysed cases, can be aggregated by a level of 4
- The aggregation level can be 3 for customers with a mix of LCTs, but this increases to 4 when there is only one type of LCTs involved in the analysis
- Summer and winter aggregation levels were also analysed with the following results:
  Winter: 3.5, summer: 4.2

In summary, for 90% of the analysed cases, 4 was the average aggregation level obtained from the analysis and this had a 95% confidence level for unrestricted customers. In other words, in 90% of the cases, the risk of being able to derive an individual profile was some 15% given an aggregation level of 4.

This result is predicated on the accepted level of risk being 15%. Had an initial target value of 20% been selected for the risk level, for example, the aggregation level would have been lower.

4.2.3 Visibility Risk for each aggregation level

The visibility risk is presented as a key metric in the evaluation of suitable aggregation levels. Visibility risk is defined as the likelihood of an individual customer consumption profile being derived from the aggregated group load profile. In other words, if someone had access to the aggregated profile, what would be the probability of deriving one individual profile from it. Hence the lower the visibility risk the greater the customer privacy.

The analysis summarised in section 4.2.2 was based on assessing the results from all the 330 tests, to establish the aggregation level for a pre-selected visibility risk 'target' value of 15%. It was found that, in 90% of the cases (i.e. 297 tests), this provided an aggregation level of 4. However, given that the analysis was carried out having pre-selected 15% as the 'target' value, it was felt that there was significant merit in exploring the visibility risk for different levels of aggregation (i.e. from 1 to 5). The results of this analysis is shown in Table 5.

It is noted that by considering all (rather than 90%) of the sample data an aggregation level of 4 results in a 17% visibility risk as opposed to a 15% visibility risk when 90% of the 330 tests are contemplated. It is also noted that for an aggregation level of 2 the visibility risk only increases by 5 percentage points to 22%. DNO distribution networks have considerable numbers of network circuits with very few customers connected. Increasing aggregation levels reduces the level of network visibility afforded to DNOs to undertake their regulatory duties. This report therefore suggests that selecting an aggregation level of 2 offers network companies greater visibility of distribution networks while still providing customers with a comparatively similar level of visibility risk to an aggregation level of 4.

Table 5 Visibility risk for each aggregation level

Aggregation level	Visibility risk (%) (median correlation results)
1	100%
2	22%
3	20%
4	17%
5	15%

Figure 8 below shows that an aggregation level of two reduces the visibility risk by approximately 80% and very little improvements for increasing levels of aggregation happen thereafter.

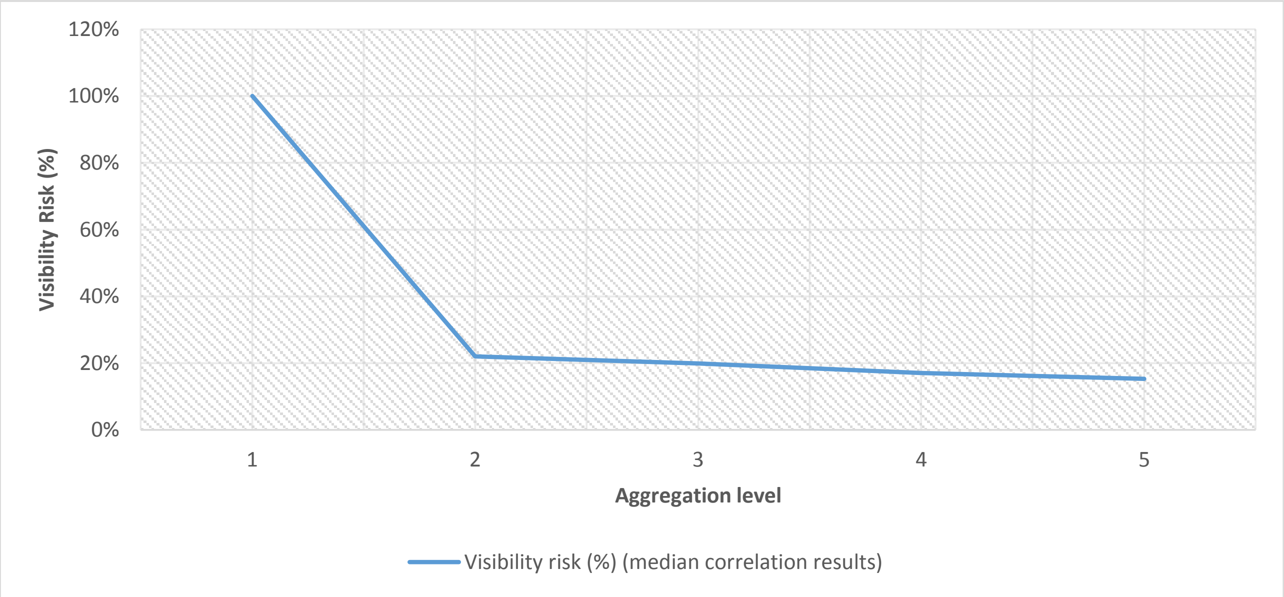

Figure 8 Visibility risk for each aggregation level

4.2.4 Load imbalance analysis

Customers' phase connectivity was also investigated as part of this project and the effect of LCTs penetration at the LV branch level was assessed.

It was supposed that the average cable size deployed at the points under investigation is a 95mm^2 LV cable (245A) for which it was studied, from HH load profiles, the maximum permissible imbalance demand (Amps) for each of the three phases. The method considered:

- Samples of 30 customers that were split between the three phases,
- Connectivity imbalances between 0% and 50%, with 10% imbalances steps,
- 50% and 60% LCT penetration rates, and
- Flat 6kW LCT loads, e.g. Heat Pumps.

The results concluded that:

- The $95mm^2$ would not be overloaded for a balanced network with a 6kW load penetration rate of 50%.
- The $95mm^2$ cable would not be overloaded for up to a 30% connectivity imbalance network with a 6kW load penetration rate up to 50%.
- The $95mm^2$ cable would not be overloaded for up to a 20% connectivity imbalance network with a 6kW load penetration rate of up to 60%

This means that when considering the availability of smart meter data, phase imbalance for LCT connections does not impose a material concern for Distributor Network Operators.

Further detailed results can be found in Appendix III.

5. Conclusions

From the various forms of analyses carried out, the metric of 'visibility risk' is deemed to be the most appropriate in evaluating the level of aggregation required to preserve customer anonymity.

The analysis indicates that moving from having individual customer profile data to data aggregated from two customers reduces the visibility risk by almost 80% while increased levels of aggregation do not greatly improve the level of anonymity.

Therefore this work concludes that aggregating two consumption profiles reduces the possibility of being able to establish an individual customer's consumption profile from 100% to 22%, and that further aggregation would provide only a marginal reduction in visibility risk. Hence aggregation of two profiles, coupled with the development and implementation of DNO IT systems and/or business processes, to address the inherent flaws in using aggregation to ensure anonymity, is proposed as being the approach to meet the requirements of SLC10a.

Appendix I Feeders selected for the smart meter Aggregation Analysis

LV feeder type	Household types	Substation feeder	Consumption customer type	Customer numbers
Suburban	3/4 bed semi-detached or detached houses	Wooler Ramsey, Feeder B	Domestic Unrestricted x 74	74
Suburban	3/4 bed semi-detached or detached houses	Rise Carr - Darlington Melrose, Feeder B	Domestic Economy 7 x 6 Domestic Unrestricted x 62	68
Suburban	3/4 bed semi-detached or detached houses	Rise Carr - Darlington Melrose, Feeder D	Domestic Economy 7 x 6 Domestic Unrestricted x 74	80
Suburban	Terraced Street	Wooler Ramsey, Feeder A	Domestic Economy 7 x 16 Domestic Unrestricted x 108	124
Suburban	3/4 bed semi-detached or detached houses	Rise Carr - Harrowgate Hill, Feeder C	Domestic Economy 7 x 1 Domestic Unrestricted x 11	12
Suburban	3/4 bed semi-detached or detached houses	Wooler St Mary, Feeder A	Domestic Economy 7 x 29 Domestic Unrestricted x 86	115
Suburban	3/4 bed semi-detached or detached houses	Wooler St Mary, Feeder C	Domestic Economy 7 x 5 Domestic Unrestricted x 11	16
Rural	Rural village / farms	Sidgate Lane, Feeder B	Consumption customer type	20
Rural	Rural village / farms	Sidgate Lane, Feeder A	Domestic Unrestricted x 74	14
Suburban	Terraced Street	Wooler Bridge, Feeder A	Domestic Economy 7 x 20	9

Appendix II Clustering methodology examples

- Smart meter driven segmentation: What your consumption says about you
  Albert, A.; Rajagopal, R.
  Power Systems, IEEE Transactions on (Volume: 28, Issue: 4)
  http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6545387&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D6545387

- A clustering approach to domestic electricity load profile characterisation using smart metering data
  Fintan McLoughlin, Aidan Duffy, Michael Conlon
  Applied Energy Volume 141, 1 March 2015, Pages 190–199
  http://www.sciencedirect.com/science/article/pii/S0306261914012963

- Household energy consumption segmentation using hourly data
  Jungsuk Kwac; Flora, J.; Rajagopal, R.
  Smart Grid, IEEE Transactions on (Volume: 5, Issue: 1)
  http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6693793&url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel7%2F5165411%2F6693741%2F06693793.pdf%3Farnumber%3D6693793

- Clustering analysis of residential electricity demand profiles
  Joshua D. Rhodes, Wesley J. Cole, Charles R. Upshaw, Thomas F. Edgar, Michael E. Webber
  Applied Energy Volume 135, 15 December 2014, Pages 461–471
  http://www.sciencedirect.com/science/article/pii/S0306261914009349

Appendix III Load imbalance analysis supportive materials

The load imbalance analysis investigated if phase connectivity could be an issue with LCT penetration at LV branch level.

For this study it was supposed that the most common cable installed at the points under investigation is a 95mm^2 LV cable (245A) for which it was analysed the maximum permissible imbalance demand (Amps) for each of the three phases from HH load profiles. The method considered:

- Samples of 30 customers that were split between the three phases,
- Connectivity imbalances between 0% and 50%, with 10% imbalances steps,
- 50% and 60% LCT penetration rates, and
- Flat 6kW LCT loads, e.g. Heat Pumps.

Table 6 and Table 7 below show the connectivity imbalance scenarios investigated for 6kW loads (HP) for 50% and 60% HP penetration rates respectively.

Table 6 Connectivity imbalance scenario – 50% HP penetration rates (15 HP out of the 30 customers)

Case	Cust.	connectivity imbalance (*phase 2 and 3 vs phase 1)	Phase 1 N° cust.	Phase 2 N° cust.	Phase 3 N° cust.	Phase 1		Phase 2		Phase 3		MD (kW) Ph 1	MD (kW) Ph 2	MD (kW) Ph 3
						N° cust. without HP	N° cust. with HP	N° cust. without HP	N° cust. with HP	N° cust. without HP	N° cust. with HP			
1	30	0%	10	10	10	5	5	5	5	5	5		160	165
2	30	10%	10	11	9	5	5	5	6	5	4		192	138
3	30	20%	10	12	8	5	5	5	7	5	3	171	220	111
4	30	30%	10	13	7	5	5	5	8	5	2		244	85
5	30	40%	10	14	6	5	5	5	9	5	1		276	58
6	30	50%	10	15	5	5	5	5	10	5	0		305	29

Table 7 Connectivity imbalance scenario – 60% HP penetration rates (18 HP out of the 30 customers)

Case	Cust.	connectivity imbalance (*phase 2 and 3 vs phase 1)	Phase 1 N˚ cust.	Phase 2 N˚ cust.	Phase 3 N˚ cust.	Phase 1		Phase 2		Phase 3		MD (kW) Ph 1	MD (kW) Ph 2	MD (kW) Ph 3
						N˚ cust. without HP	N˚ cust. with HP	N˚ cust. without HP	N˚ cust. with HP	N˚ cust. without HP	N˚ cust. with HP			
1	30	0%	10	10	10	4	6	4	6	4	6	197	186	191
2	30	10%	10	11	9	4	6	4	7	4	5		218	164
3	30	20%	10	12	8	4	6	4	8	4	4		245	138
4	30	30%	10	13	7	4	6	4	9	4	3		275	111
5	30	40%	10	14	6	4	6	4	10	4	2		302	84
6	30	50%	10	15	5	4	6	4	11	4	1		331	56

The results concluded that:

- The 95mm^2 would not be overloaded for a balanced network with a 6kW load penetration rate of 50%.
- The 95mm^2 cable would not be overloaded for up to a 30% connectivity imbalance network with a 6kW load penetration rate up to 50%.
- The 95mm^2 cable would not be overloaded for up to a 20% connectivity imbalance network with a 6kW load penetration rate of up to 60%

This means that phase imbalance for LCT connections does not impose a material concern for Distributor Network Operators.

www.eatechnology.com
Australia | China | Singapore | UAE | Europe | USA

Main reception: +44(0) 151 339 4181
EA Technology, Capenhurst Technology Park,
Capenhurst, Chester, CH1 6ES, United Kingdom

Global Footprint

We provide products, services and support for customers in 90 countries, through our offices in Australia, China, Europe, Singapore, UAE and USA, together with more than 40 distribution partners.

Our Expertise

We provide world-leading asset management solutions for power plant and networks.

Our customers include electricity generation, transmission and distribution companies, together with major power plant operators in the private and public sectors.

- Our products, services, management systems and knowledge enable customers to:
- Prevent outages
- Assess the condition of assets
- Understand why assets fail
- Optimise network operations
- Make smarter investment decisions
- Build smarter grids
- Achieve the latest standards
- Develop their power skills

Main reception: +44(0) 151 339 4181
EA Technology, Capenhurst Technology Park,
Capenhurst, Chester, CH1 6ES, United Kingdom