

Informatics Solutions for Smart Metering Systems Integration

Simona-Vasilica OPREA, Ion LUNGU

The Bucharest University of Economic Studies, Romania

simona.oprea@csie.ase.ro, ion.lungu@ie.ase.ro

In this paper different aspects regarding smart metering systems integration have been depicted. Smart metering systems, renewable energy sources integration and advanced tariff systems implementation require informatics solution that could automatically collect and process data, forecast the behavior of electricity consumers, analyze trends regarding electricity prices, optimize the consumption of consumers, provide friendly interfaces, etc. They are advanced technologies that represent solutions for insufficient conventional primary energy sources, gas emissions, dependency on energy sources located outside European Union and issues related to energy efficiency. This paper mainly describes several informatics solutions correlated with operational requirements for smart metering system and our proposal for simplified architecture of smart metering systems, with three distinct levels (base level, middle level and top level) and load profile calculation methods.

Keywords: Operational Requirements, Internet of Things, Architecture of Smart Metering Systems, Renewables Integration, Load Profile, Data Warehouse

1 Introduction

Smart grid concept is not a new one; it has been defined since October 1997 by three pioneers Khoi Vu, Miroslav M. Begovic, Damir Novosel in *Grids get smart protection and control*, published in *IEEE Journal Computer Applications in Power* [1]. Since 1997, thousands of papers, books and thesis have been written.

Smart grids include modern systems and equipment governed by informatics technology that collect data and take decisions based on electricity consumers and prosumers in order to improve efficiency, reliability, economic aspects and sustainability of generation, transmission, distribution and energy supply of electricity consumers.

The most known smart grid components are protections through relays and automatics, Energy Management System/Supervisory Control And Data Acquisition EMS/SCADA, smart meters, integrated communication systems, decision support systems and advanced interfaces.

Smart metering systems allow electricity consumers to have an active role, by managing programmable consumption, by using storage equipment and supplying as prosumers the national grid with electricity produced by small size wind generators or solar panels according

to the electricity price fluctuation.

Smart metering systems allows grid operators and suppliers to better plan their resources, reduce energy losses, estimate necessary energy that should be acquired in order to cover losses, eliminate costs regarding meters reading and integrate a higher volume of renewable energy sources.

In Romania, there are 9,000,000 electricity consumers, out of which 8,380,000 household electricity consumers, 600,000 small and medium size non-household electricity consumers and 20,000 big non-household electricity consumers. Only 75,000 consumers have smart metering with automatic reading (*AMR – automatic meter reading*) and approximately 15,300 consumers have complex smart meters (*AMI – advanced metering infrastructure and AMM – advanced metering management*) [2]. Meters communicate with concentrators and central points of network operators by optic fibre installed on overhead lines PLC (*Power Line Communication*), GPRS (*General Packet Radio Service*), etc. Only 1% of Romanian electricity consumers have installed smart meters.

European strategies regarding integration of renewable energy sources and implementation of smart metering systems have been taken

over into Romanian legislation as national targets. For implementation of smart metering systems, Romanian National Regulatory Authority approved gradually implementation of these systems by grid operators so that until 2020, about 80% of energy consumers should benefit from smart metering systems.

2 Operational Requirements of Smart Meters and Informatics Solutions

One important integration issue of smart metering systems consists in interoperability of these systems and their architecture. We should consider that smart metering systems usually belong to different grid operators; therefore a proper approach is to implement an open standard communication protocol.

Interoperability of smart metering systems could be performed by setting minimal operational requirements. These requirements were established by several regulations at European level [3]. They have been taken over into national level [4].

Operational requirements have been divided into five categories: requirements for electric-

ity consumers, grid operators, suppliers, security and data protection, distributed production.

The first category includes supplying updates of records or other information such as electricity price. Based on updated information, consumers can change behaviour: reduction of the peak consumption, increasing of off-peak consumption as a consequence of financial incentives. Electricity consumers' behaviour influences grid investment due to flattening of load curve, renewable energy sources integration, greenhouse emissions reduction, etc.

Information updates should be performed at certain intervals so that to allow consumer to react. It is recommended to update readings at 15 minutes.

For grid operators, smart metering systems should allow remote readings that will reduce operators' expenses regarding meters reading and billing based on real consumption. Also, smart metering systems should allow bidirectional communication and provide sufficient readings so that operators could better plan their resources (i.e. grid capacity).

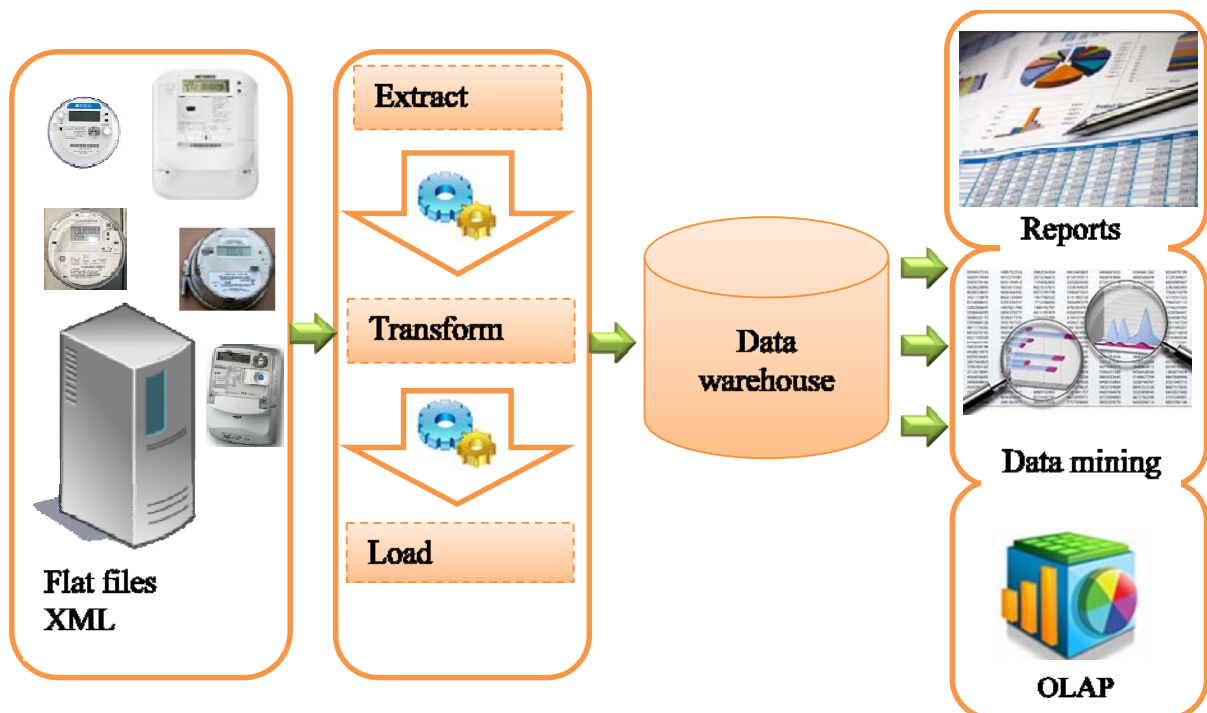


Fig. 1. Informatics solutions for smart metering data

After ETL process has been performed, these data could be stored into a data warehouse.

For each consumer could be set queries and reports. Results of these analyses are provided

to electricity consumers. Data processed as strategic reports is provided to top level managers of grid operators. For this task, business intelligence tools are recommended (data mining, OLAP) as in Figure 1.

From the electricity supplying point of view, smart metering systems should be correlated with advanced tariff systems, so that to inform

consumers regarding tariff options. Without advanced tariff systems, requirements at the electricity consumers' level could not reach the target. Smart metering systems should allow remote connection/disconnection of receivers by sensors and Internet of Things means, Figure 2, which will provide cost reduction.

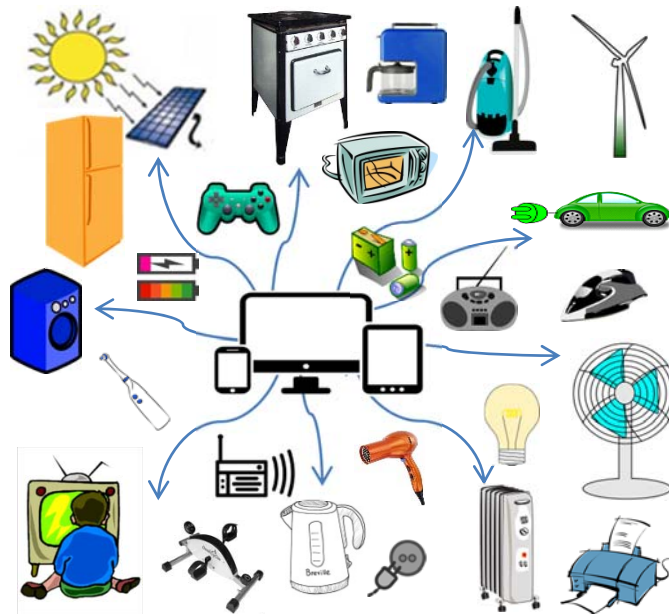


Fig. 2. Internet of Things concept

The advanced tariff systems are also involved into electricity consumption optimization tools. The most known advanced tariff systems are: time-of-use tariff, critical peak pricing tariff, real time tariff, etc.

As for security and data protection requirements, these are essential for both grid operators due to the decisions they take based on metering data and consumers that are not willing to disclose private information regarding their electricity consumption. Smart metering systems should be able to detect frauds. This will also reduce commercial losses.

In terms of distributed generation requirements, smart metering systems should also allow, apart from consumption of electricity from the grid, generation and supplying of electricity of local sources (small wind turbines, solar panels) to the grid. Smart metering systems should also allow metering of reactive energy. This way, local distributed generation is encouraged. Local generation has several advantages such as lower loading of

national grids, losses reduction due to long distances transmission and distribution, replacement of conventional sources with renewable energy ones, reaching national targets regarding RES integration, environment protection, etc.

The most difficult requirement is related to frequency and availability of updates to the electricity consumers. This operational requirement will enhance reactions from consumers' side.

From pilot projects experiences, incentives for consumers should be efficient, otherwise, they will be not interested to manage consumption and most of the operational requirements of smart metering systems will become useless.

3 Smart Metering System Architecture

Smart metering systems are electronic devices that measure electricity consumption, allow secured bidirectional data transmission from/to electricity consumers and supply more

information than a regular meter, by using telecommunications. Smart metering systems contain subsystems for metering, subsystems for data transmission and subsystems for data management that are provided by meters.

Smart metering systems hourly (or even at 15 minutes) measure electricity consumption using metering electronic devices, transmit recorded data and send it to unified and complex data management system. Heterogeneity of metering devices imposes data integration components by using universal standards that allow loading and validating data in a centralized manner.

According to [5] and [6], these components are part of advanced metering infrastructure – *AMI* that measures, collects, analyses energy

consumption and facilitate communications with metering devices. *AMI* includes hardware components, telecommunication networks, interfaces, controllers for electricity consumers, software components for meter data management – *MDM*. According to [7], the most important advanced architecture and infrastructure for smart metering systems are: Roadmap E-Energy/Smart Grid, NIST IOP Roadmap, IEC SMB SG 3, BMWi E-Energy Program, BDI initiative – Internet der Energy, Microsoft SERA, CIGRE D2.24, European Union Mandate CEN/CENELEC M/441, etc. Starting from common characteristics of these architectures, in Figure 3 we present simplified architecture of smart metering systems.

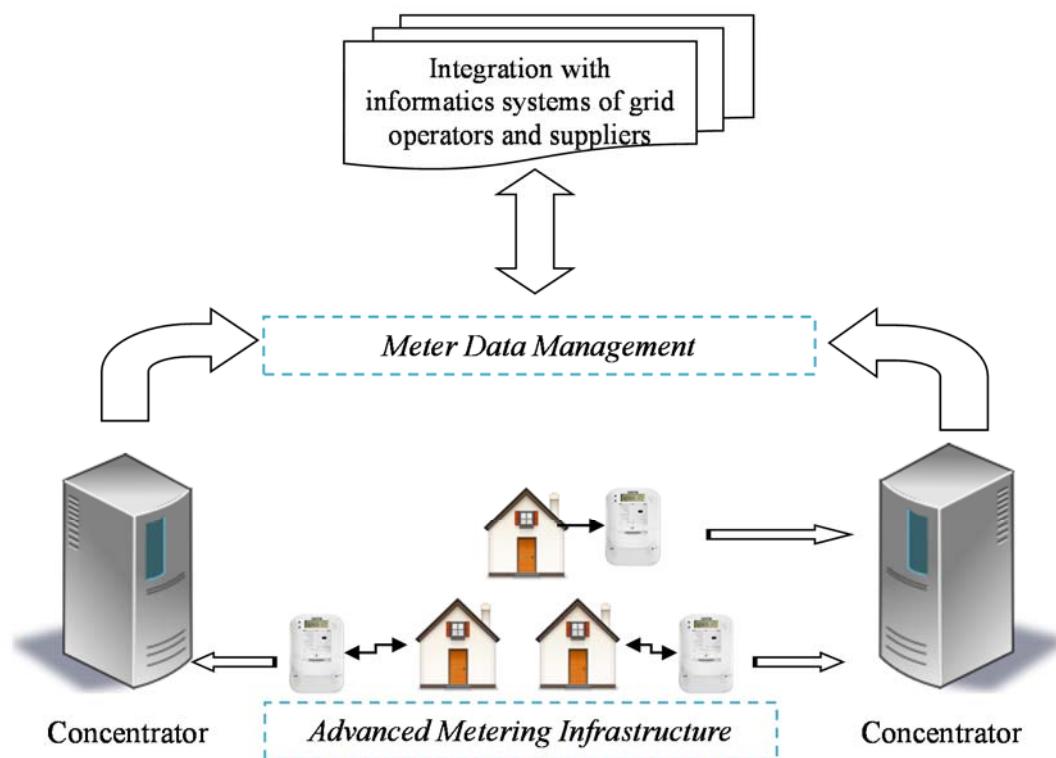


Fig. 3. Simplified architecture for SMS

By integration with grid operators and electricity suppliers' informatics systems and by analyses facilities, smart metering systems could be included into management information systems – *MIS*. According to [8], *MIS* are necessary for current tactic decision assistance for short term control and planning. *MIS* architecture is structured on four distinct levels, according to [8] and [9]:

- **Data level** – represent data sources level, database management system level and metadata dictionaries level. In order to organize data for data processing, it is necessary to follow a couple of stages: collection and validation of data from several sources, data cleansing and transforming, data loading into databases or data warehouses;

- **Models level** – represent the level of computing models and data analyses that are necessary for management requirements. As for obtaining models mathematic, statistics methods, data mining algorithms, predictive analyses could be used, etc.;
- **Interfaces level** – contains applications and tools for data dynamic analyses presented into an interactive form with flexible analyses facilities;
- **Telecommunications level** – contains the set of interconnection elements for previous levels and includes computer networks, communication devices and support for distributed software.

Starting from this architecture, we structure the architecture for *SMS* on three levels: operational level, middle level, top level.

New measuring devices are smart meters that are located at the interface between electricity consumers and national grid. Operational level contains data that is evaluated, processed and integrated into database. In the initial stage data is validated in order to ensure quality of results. From the database, data is utilized for invoicing application, sending messages to consumers via electricity meter or specific portals. We present this process with its stages in Figure 4.

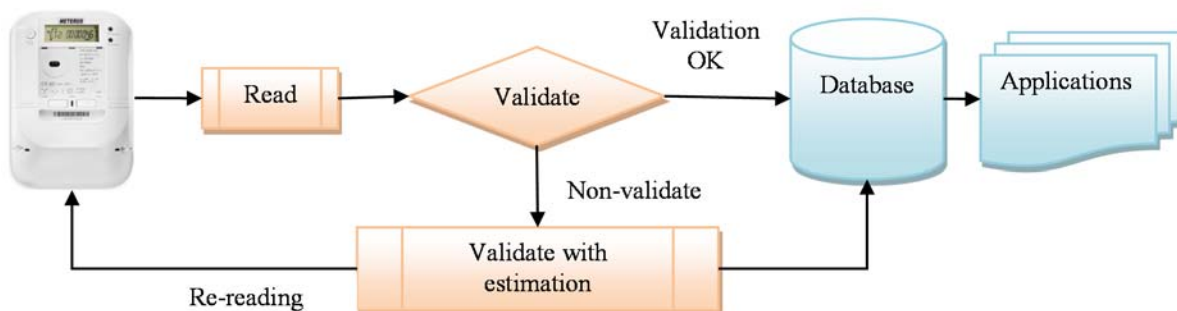


Fig. 4. The process of data collection from SMS and integration into database

Actual problems at this level refer to suppliers/grid operators' current informatics systems incapacity to process data provided by *SMS*. In the coming years, the volume of this data will be huge and will need powerful platforms and technologies that are able to perform respective processes. This data is characterized by velocity, variety and veracity that impose the necessity to study and develop *Big Data* techniques in order to obtain real time right and complex information. From monthly readings of meters to 15-minute readings, about 2900 monthly records for each single meter will be generated. Taking into account that in Romania there are about 9 million electricity consumption places, the monthly data volume will significantly increase that leads to changes of data storage and processing technologies.

Smart meters have to fulfil minimal functional requirement and transmit at least the following data: active and reactive energy consumption, consumed/generated active, reactive

power, active and reactive energy supplied to the national grid, etc. This data is processed at suppliers/grid operators' level. Out of it, consumption profiles are dynamically generated and electricity consumption optimisation applications based on electricity price are used by consumers. The electricity consumer will play a more and more active role being able to optimise consumption. The next step will be installation at large scale of smart sensors for detection of appliances consumptions and connection of some appliances to interface control (smart refrigerator or ovens, heating systems and air conditioning devices that can be programmed via Internet, etc.). The concept of smart house will generate new challenges regarding data processing and organisation that comes up from new appliances. Variety and velocity of the data that is provided by these appliances impose new technologies for parallel processing and data organising into *NoSQL databases*. *Big Data*

technologies for smart metering industry are presented in [10], [11].

All data collected from sensors are processed based on a standard flow [12]. Data collected from meters, sensors and smart appliances are transmitted via telecommunication system to the main server where cleansing and processing by *Apache Hadoop clusters* or *database in memory* take place. After data cleansing and validation, this is transmitted to the analytic server where models and algorithms are applied in order to obtain consumption profiles, optimisation and other computation regarding energy requirements, electricity prices, etc. Information is analysed via advanced analytic tools.

The middle level contains database and telecommunication system that ensures data transmission from consumers to suppliers/grid operators' database and from databases to electricity consumers. Data transmission can be done by different technologies: general packet radio service – *GPRS*, optic fibre, wireless. From the adopted communication system point of view, the most accepted solution in countries with high level of theft consists in common infrastructure for several utilities (electricity, natural gas and thermal energy) with an intermediary level known as *middleware*, concentrators and balancing meters that rapidly locate theft Figure 5.

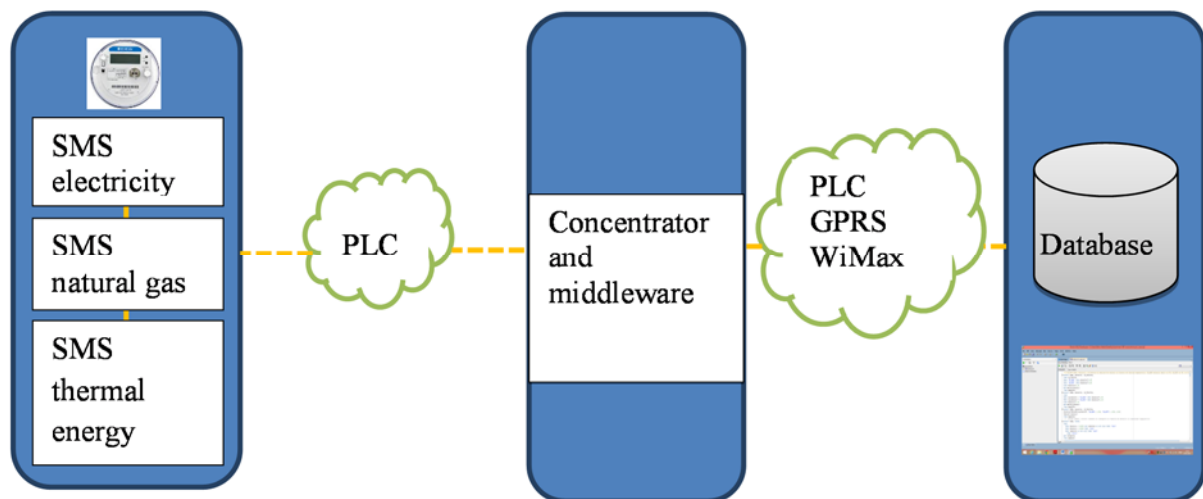


Fig. 5. The architecture of a common telecommunication system with intermediary level

In comparison with *GPRS* technology that has proportional costs with transmitted data volume, communication technology via power line carrier – *PLC* has more advantages because of the lower costs. *PLC* technology is used currently in order to monitor the state operation of National Power System. Taking into account equipment that belong to different grid operators, communication errors should be avoided by open standard protocols that are able to facilitate communication among meters, concentrators and database.

At top level the most important elements are represented by computation models, estimation of consumption profile, electricity price model, analyses and reports for strategic management.

Prosumers' behaviour estimation is a continually changing activity due to the evolution of appliances, generation opportunities from own sources (micro-generation, solar panels), social and demographic changes. At suppliers/grid operators' level, consumption estimation is very significant for planning, commercial and market activities. Thus, a good estimation of consumption will lead to efficient decisions regarding buying or selling of electricity at advantageous prices and setting attractive electricity prices for consumers.

Applications for advanced analyses and reports for suppliers include interactive interfaces by using dashboards, prediction analyses, *what if* scenarios, planning and reporting tools, etc. These are obtained with *business intelligence* technologies. By implementation of

smart metering systems and sensors, reinventing these applications represents a new challenge due to *business intelligence* technologies that have to be applied with *NoSQL databases*.

The applications that are designed for electricity consumers include information such as 15-minute consumption, electricity price, alerts

and consumption thresholds, comparison between their consumption and similar consumers located in the same area, estimations and predictions of electricity consumption, etc. In Figure 6 we propose an architectural model for smart metering system taking into account characteristics of previous presented levels.

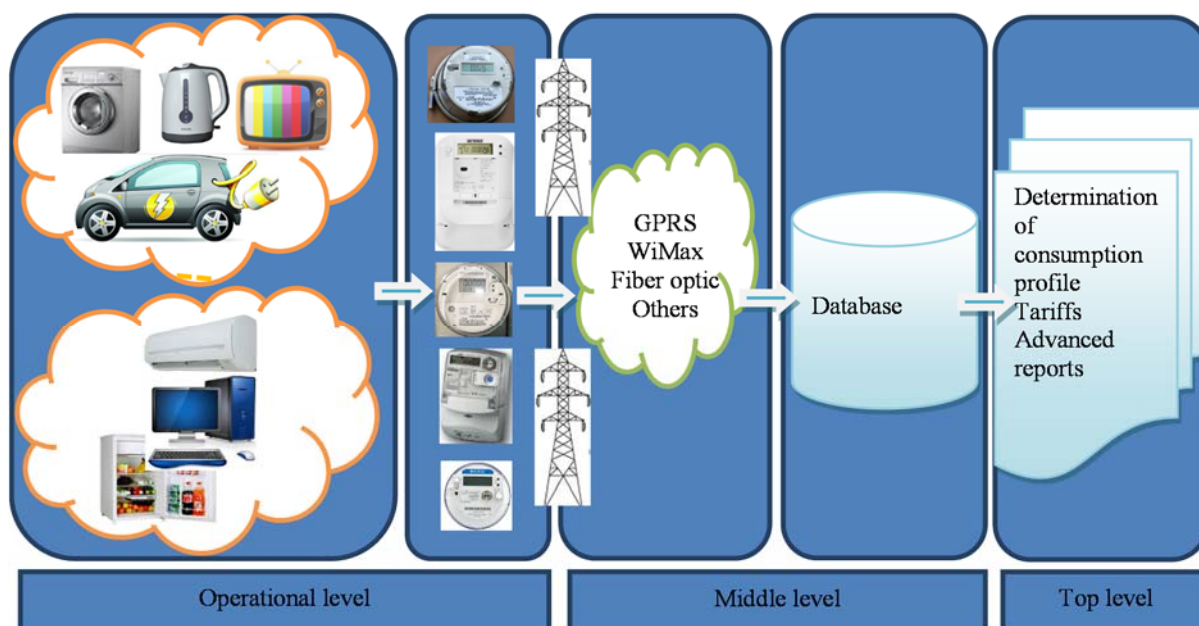


Fig. 6. Architectural model for SMS

The proposed architecture is flexible and can be adapted to new technologic trends regarding smart appliances, metering devices and sensors.

4 Informatics Solutions for Load Profile Calculation

Load profiles are graphical representation of electricity consumption variation in time. They are useful for certain activities such as: electricity market settlement, load forecasting and load optimization. They are related to the consumer type, temperature, season, working day, weekend day, etc. In paper [13] several methods for calculation of load profile were depicted. Multi-linear regression is used by English grid operator in order to calculate eight load profiles. Auto-regression with exogenous variables is applied for a sample of 1000 electricity consumers from Canada (Ontario) in order to calculate load profiles. Fuzzy C-means clustering method has been applied

by authors of papers [14] and [15]. They performed a study case that analyses 300 load curves of electricity consumption from Malaysia.

In order to design load profile, we propose two types of analyses related to data classification and clustering techniques.

Input data regarding electricity hourly consumption [16] were recorded in several cities in the United States of America from January to December 2014. Each record has consumption data related to heating, cooling, ventilation, interior and outside lightening, water heating, regular equipment (washing machine, refrigerator) and other smaller devices (i.e. TV, audio system, computer). In order to determine profiles, we imported data in Oracle Database 11g.

First, we analysed consumption distribution for different values intervals, based on consumption type and time intervals Figure 7.

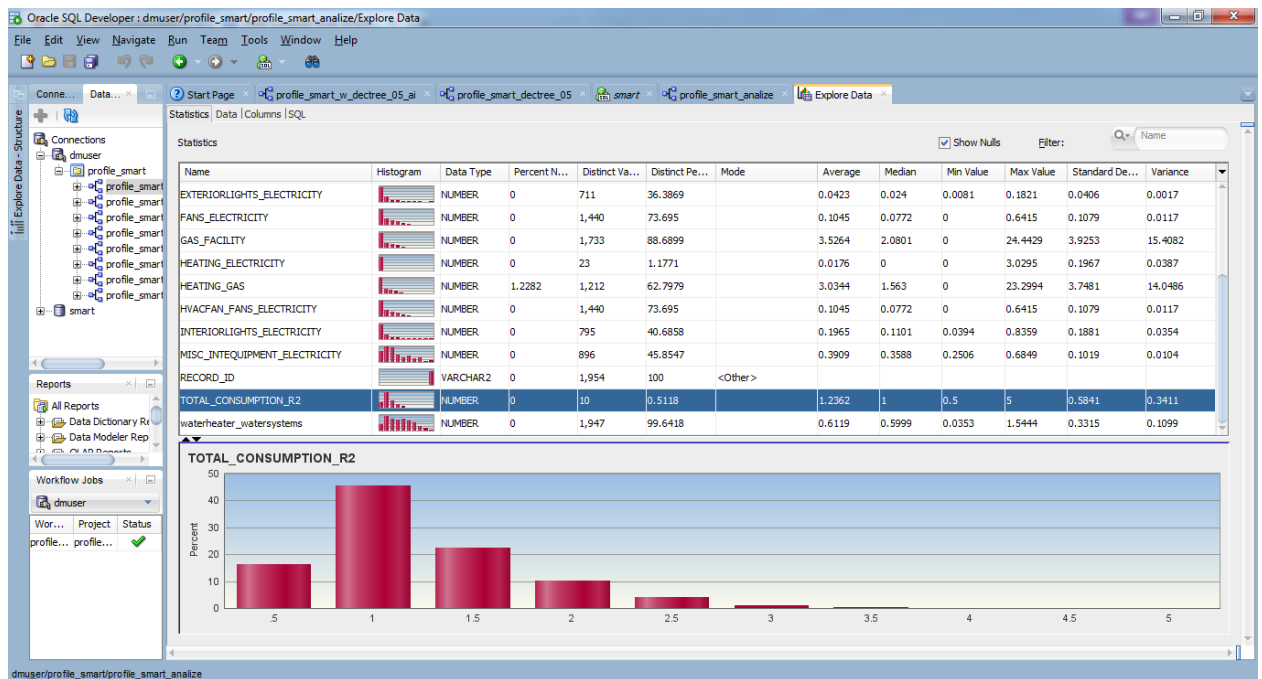


Fig. 7. Consumption distribution for different values intervals

We noticed that about 40% of the consumption is approximately 1 kWh, more than 20% of the consumption is 1.5 kWh and more than 15% of the consumption is about 5 kWh.

We also noticed that heating and interior equipment have similar trends as total consumption, therefore these attributes are significant in the total consumption.

Applying classification techniques, we built the model through decision trees and Support Vector Machines (SVM) algorithm taking into account all attributes without any data processing. In order to build the decision tree, we initially set maximum seven levels and for SVM algorithm, we considered linear function of similarity (kernel). The target variable is the total recorded consumption. The initial

results were not significant, both models showed a low accuracy (under 20%). Therefore, we processed the input data, especially those related to the total consumption. We rounded total consumption values, these values being from 1 to 8 and for values under 0.5, we considered them equal to 0.5. After applying the two methods, the accuracy improved, especially for decision tree – 87% Figure 8.

We obtained a tree with 6 levels and 36 leaf nodes, for each node accuracy is given. By analysing tree leaves, we noticed that nodes with the most cases (over 30 thousands) have the highest trust degree (cca. 90%), that is a good result for right estimation of load consumption.

Name:	"DM4J\$T299875542451_M"			
Average Accuracy:	0.6341188248			
Overall Accuracy:	0.8704922887			
Total Cost:	3345300.2757			

Model Performance	<input checked="" type="checkbox"/> Show Cost	
-------------------	---	--

Target	Total Actuals	Correctly Predicted ...	Cost	Cost %
.5	37,457	98.81	10,735.11	0.32
1	658,039	90.36	87,607.61	2.62
2	187,473	73.54	240,139.27	7.18
3	22,120	86.27	124,781.36	3.73
4	2,603	73.3	239,418.79	7.16
5	599	48.41	475,126.19	14.2
6	143	0	869,428.75	25.99
7	30	100	0	0
8	7	0	1,298,063.2	38.8

Confusion Matrix: Rows = Actual; Columns = Predicted	<input type="checkbox"/> Show Total and Cost	
--	--	--

	.5	1	2	3	4	5	6	7	8
.5	37,012	445	0	0	0	0	0	0	0
1	14,761	594,618	48,625	35	0	0	0	0	0
2	0	28,081	137,875	21,506	11	0	0	0	0
3	0	0	1,184	19,084	1,844	8	0	0	0
4	0	0	0	108	1,908	538	0	49	0
5	0	0	0	0	73	290	0	236	0
6	0	0	0	0	0	8	0	135	0
7	0	0	0	0	0	0	0	30	0
8	0	0	0	0	0	0	0	7	0

Fig. 8. Results obtained with decision tree algorithm

Although, the results are significant, built classes are sensitive to consumers' behaviour changes as we noticed within classes with lower number of elements where errors are high. In order to eliminate these drawbacks, we applied unsupervised learning technique: clustering.

In order to build clusters, we specified 10 clusters so that elements in each cluster to be similar, but different from elements from other

clusters. For similarity measure in each cluster, we considered variance (sum of square difference between main element and each individual element), choosing clusters with smallest variance.

We applied K-means method for building clusters; Euclidian distance has been calculated. Clusters have been described in Table 1.

Table 1. Description of K-means clusters

Cluster: 7 If 2.02310361 <= ELECTRICITY_HVAC <= 3.37183935 And 1.96318693 <= HEATING_ELECTRICITY <= 3.27197822 And 3.5 <= TOTAL_CONSUMPTION_R2 <= 5 And 0 <= HEATING_GAS <= 2.32994483 And 0 <= GAS_FACILITY <= 2.48639116 And 0.06908582 <= FANS_ELECTRICITY <= 0.13817163 And 0.06908582 <= HVACFAN_FANS_ELECTRICITY <= 0.13817163 And 0.03525717 <= WATERHEATER_WATERSYSTEMS <= 0.63983215 And 0.33502575 <= MISC_INTEQUIPMENT_ELECTRICITY <= 0.50995878 And 0.00807 <= EXTERIORLIGHTS_ELECTRICITY <= 0.06028729 Then Cluster is: 7	Cluster: 9 If 3.37183935 <= ELECTRICITY_HVAC <= 4.04620722 And 3.27197822 <= HEATING_ELECTRICITY <= 3.92637386 And 3.5 <= TOTAL_CONSUMPTION_R2 <= 5.75 And 0 <= HEATING_GAS <= 2.32994483 And 0 <= GAS_FACILITY <= 2.48639116 And 0.06908582 <= FANS_ELECTRICITY <= 0.13817163 And 0.06908582 <= HVACFAN_FANS_ELECTRICITY <= 0.13817163 And 0.03525717 <= WATERHEATER_WATERSYSTEMS <= 0.63983215 And 0.33502575 <= MISC_INTEQUIPMENT_ELECTRICITY <= 0.50995878 And 0.0459 <= APPL_INTEQUIPMENT_ELECTRICITY <= 0.35255899 Then Cluster is: 9
Cluster: 11 If 4.04620722 <= ELECTRICITY_HVAC <= 5.39494296 And 3.92637386 <= HEATING_ELECTRICITY <= 4.5807695	Cluster: 12 If 1.34873574 <= ELECTRICITY_HVAC <= 2.69747148 And 1.30879129 <= HEATING_ELECTRICITY <= 2.61758257

And 5 <= TOTAL_CONSUMPTION_R2 <= 6.5 And 0 <= HEATING_GAS <= 2.32994483 And 0 <= GAS_FACILITY <= 2.48639116 And 0.06908582 <= FANS_ELECTRICITY <= 0.20725745 And 0.06908582 <= HVACFAN_FANS_ELECTRICITY <= 0.20725745 And 0.03525717 <= WATERHEATER_WATERSYSTEMS <= 0.63983215 And 0.00807 <= EXTERIORLIGHTS_ELECTRICITY <= 0.06028729 And 0.33502575 <= MISC_INTEQUIPMENT_ELECTRICITY <= 0.50995878 Then Cluster is: 11	And 2 <= TOTAL_CONSUMPTION_R2 <= 4.25 And 0 <= HEATING_GAS <= 2.32994483 And 0 <= GAS_FACILITY <= 2.48639116 And 0 <= FANS_ELECTRICITY <= 0.06908582 And 0 <= HVACFAN_FANS_ELECTRICITY <= 0.06908582 And 0.03525717 <= WATERHEATER_WATERSYSTEMS <= 0.63983215 And 0.33502575 <= MISC_INTEQUIPMENT_ELECTRICITY <= 0.5974253 And 0.00807 <= EXTERIORLIGHTS_ELECTRICITY <= 0.06028729 Then Cluster is: 12
Cluster: 13 If 0.67436787 <= ELECTRICITY_HVAC <= 2.02310361 And 0.65439564 <= HEATING_ELECTRICITY <= 1.96318693 And 0 <= HEATING_GAS <= 2.32994483 And 0 <= GAS_FACILITY <= 2.48639116 And 0 <= FANS_ELECTRICITY <= 0.06908582 And 0 <= HVACFAN_FANS_ELECTRICITY <= 0.06908582 And 1.25 <= TOTAL_CONSUMPTION_R2 <= 3.5 And 0.03525717 <= WATERHEATER_WATERSYSTEMS <= 0.63983215 And 0.29129249 <= MISC_INTEQUIPMENT_ELECTRICITY <= 0.68489181 And 0.00807 <= EXTERIORLIGHTS_ELECTRICITY <= 0.06028729 Then Cluster is: 13	Cluster: 15 If 0.5 <= TOTAL_CONSUMPTION_R2 <= 2 And 0 <= HEATING_GAS <= 9.31977932 And 0 <= GAS_FACILITY <= 9.94556464 And 0 <= ELECTRICITY_HVAC <= 0.67436787 And 0 <= FANS_ELECTRICITY <= 0.27634326 And 0 <= COOLING_ELECTRICITY <= 0.33052529 And 0 <= HEATING_ELECTRICITY <= 0.65439564 And 0 <= HVACFAN_FANS_ELECTRICITY <= 0.27634326 And 0.03525717 <= WATERHEATER_WATERSYSTEMS <= 1.24440713 And 0.00807 <= EXTERIORLIGHTS_ELECTRICITY <= 0.07769305 Then Cluster is: 15
Cluster: 16 If 0.67436787 <= ELECTRICITY_HVAC <= 2.69747148 And 0.66105059 <= COOLING_ELECTRICITY <= 1.98315177 And 0.20725745 <= FANS_ELECTRICITY <= 0.48360071 And 0.20725745 <= HVACFAN_FANS_ELECTRICITY <= 0.48360071 And 2 <= TOTAL_CONSUMPTION_R2 <= 4.25 And 0 <= HEATING_GAS <= 2.32994483 And 0 <= GAS_FACILITY <= 2.48639116 And 0.18640092 <= WATERHEATER_WATERSYSTEMS <= 0.63983215 And 0.26494213 <= APPL_INTEQUIPMENT_ELECTRICITY <= 0.44017584 And 0.24755923 <= MISC_INTEQUIPMENT_ELECTRICITY <= 0.50995878 Then Cluster is: 16	Cluster: 17 If 0 <= HEATING_GAS <= 2.32994483 And 0 <= GAS_FACILITY <= 2.48639116 And 0 <= ELECTRICITY_HVAC <= 1.34873574 And 0 <= COOLING_ELECTRICITY <= 0.66105059 And 0.03525717 <= WATERHEATER_WATERSYSTEMS <= 0.7909759 And 0.24755923 <= MISC_INTEQUIPMENT_ELECTRICITY <= 0.55369204 And 1.25 <= TOTAL_CONSUMPTION_R2 <= 2.75 And 0.00807 <= EXTERIORLIGHTS_ELECTRICITY <= 0.11250458 And 0.03936766 <= INTERIORLIGHTS_ELECTRICITY <= 0.51730445 And 0.06908582 <= FANS_ELECTRICITY <= 0.27634326 Then Cluster is: 17
Cluster: 18 If 5.39494296 <= ELECTRICITY_HVAC <= 6.7436787 And 5.23516515 <= HEATING_ELECTRICITY <= 6.54395643 And 6.5 <= TOTAL_CONSUMPTION_R2 <= 8 And 0 <= GAS_FACILITY <= 2.48639116 And 0.13817163 <= FANS_ELECTRICITY <= 0.20725745 And 0.13817163 <= HVACFAN_FANS_ELECTRICITY <= 0.20725745 And 0.03525717 <= WATERHEATER_WATERSYSTEMS <= 0.63983215 And 0.00807 <= EXTERIORLIGHTS_ELECTRICITY <= 0.02547576 And 0.33502575 <= MISC_INTEQUIPMENT_ELECTRICITY <= 0.46622552 And 0.0459 <= APPL_INTEQUIPMENT_ELECTRICITY <= 0.26494213 Then Cluster is: 18	Cluster: 19 If 4.72057509 <= ELECTRICITY_HVAC <= 6.06931083 And 4.5807695 <= HEATING_ELECTRICITY <= 5.88956079 And 5.75 <= TOTAL_CONSUMPTION_R2 <= 7.25 And 0 <= HEATING_GAS <= 2.32994483 And 0 <= GAS_FACILITY <= 2.48639116 And 0.13817163 <= FANS_ELECTRICITY <= 0.20725745 And 0.13817163 <= HVACFAN_FANS_ELECTRICITY <= 0.20725745 And 0.03525717 <= WATERHEATER_WATERSYSTEMS <= 0.63983215 And 0.00807 <= EXTERIORLIGHTS_ELECTRICITY <= 0.06028729 And 0.33502575 <= MISC_INTEQUIPMENT_ELECTRICITY <= 0.50995878 Then Cluster is: 19

The trust degree and the number of elements of each cluster are shown in Figure 9. The

trust degree is relatively high for the most of the cases, over 85%.

Clusters Rules Results Build Settings Task		
<input type="checkbox"/> Show Attributes With Minimum Relevance Rank: 10		
Rules <input checked="" type="checkbox"/> Only Show Rules for Leaf Clusters		
Cluster ID	Confidence (%)	Support Count
3	89.82	248541
6	94.67	139219
9	85.32	280908
10	83.81	22878
12	69.37	5885
13	81.82	32871
15	92.68	127254
16	85.39	10151
18	79.93	1350
19	86.50	17438

Fig. 9. Trust degree for each cluster

For each cluster, rules have been generated based on attributes of each consumption category (heating, cooling, fanning, interior lightening, outside lightening, divers electric equipment).

Applying this method, we obtained 10 load profiles shown in Figure 10. It is obvious that P3, P4, P9 profiles differ in terms of shape and

level of consumption but P1, P2, P7, respectively P5, P6, P8, P10 profiles are similar as shape, but the level of consumption is different. Due to this aspect, we considered a lower number of clusters, but the results were not significant, the five load profile proved to be very similar so that the representation of consumers' consumption was too generalized.

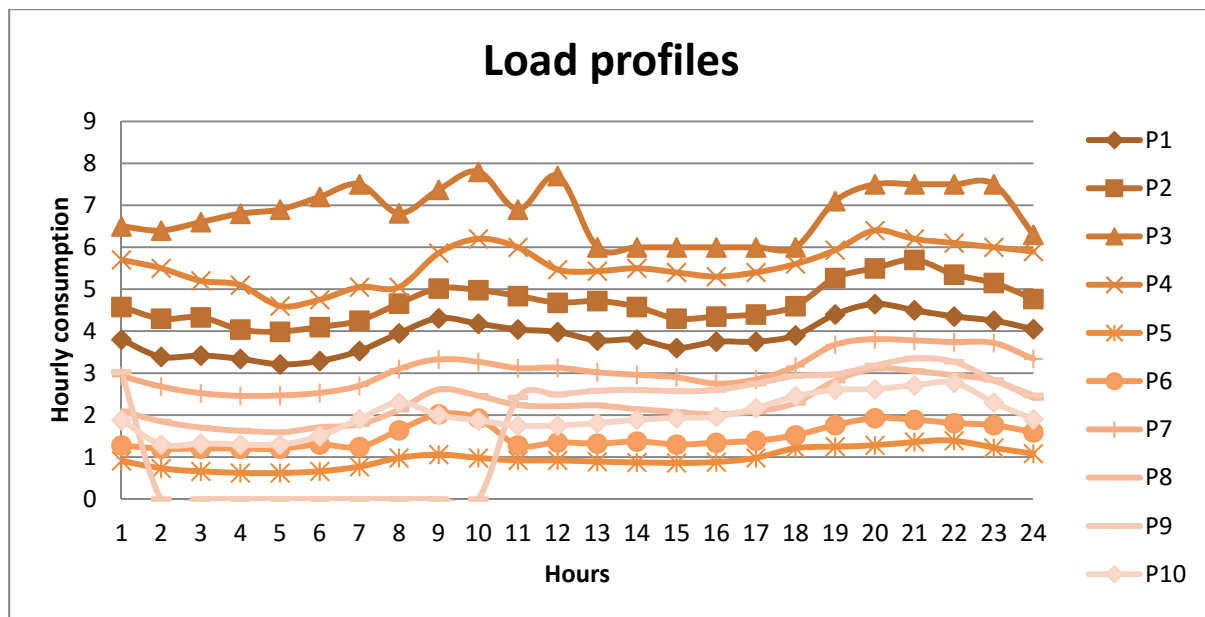


Fig. 10. Load profiles by K-means

We also applied O-cluster method (Orthogonal partitioning *clustering*) from Oracle. This method is using a recursive algorithm of data clustering by data orthogonal partitioning. We obtained hourly consumption patterns.

By analysing forming rules and each weight of consumption category within clusters, we

detailed each profile obtained by K-means method with hourly consumption patterns obtained by O-cluster method. For instance, we considered profile P5 and hourly consumption patterns related to this profile (Table 2).

Table 2. Repartition of hourly consumption patterns with P5 profile

Profile K-means	Pattern O-Cluster	Hour	Hourly consumption (kWh)	Hourly average consumption profile (kWh)
P5	T10	0:00	0.91	0.91
P5	T6	1:00	0.9	0.73
P5	T5	2:00	0.56	0.66
P5	T6	3:00	0.8	0.62
P5	T5	4:00	0.55	0.62
P5	T6	5:00	0.78	0.66
P5	T5	6:00	0.63	0.77
P5	T8	7:00	0.97	0.98
P5	T2	8:00	1.48	1.06
P5	T10	9:00	0.83	0.98
P5	T10	10:00	0.88	0.92

P5	T2	11:00	1.3	0.92
P5	T10	12:00	0.9	0.89
P5	T3	13:00	1.25	0.87
P5	T10	14:00	0.89	0.86
P5	T3	15:00	1.2	0.88
P5	T3	16:00	1.24	0.98
P5	T10	17:00	1.07	1.21
P5	T2	18:00	1.6	1.25
P5	T11	19:00	1.12	1.28
P5	T3	20:00	1.36	1.36
P5	T3	21:00	1.3	1.39
P5	T4	22:00	1.17	1.22
P5	T4	23:00	1.13	1.08

Hourly average consumption for P5 profile and effective hourly consumption of consumption patterns are shown in Figure 11.

With blue bar is shown hourly consumption (kWh) and with red line is represented hourly consumption profile.

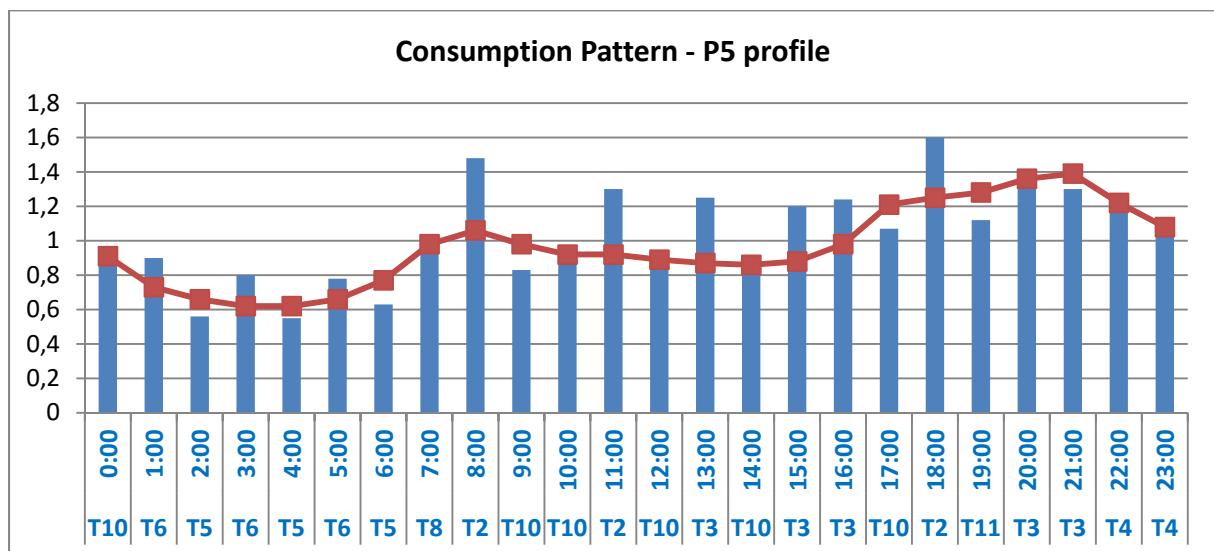


Fig. 11. Repartition of load patterns within P5 profile

We noticed that consumption patterns models dynamic behaviour of electricity consumers; load profiles are an approximation of variation of hourly consumption. Differences between patterns and load profile are small; therefore it validates our clustering model. Taking into account the results that we obtained, the most efficient method to determine electricity consumers' behaviours consists in setting load profiles and then building dynamic load patterns that will significantly contribute to the consumption optimization, market settlements, load forecasts, etc.

5 Conclusions

In paper authors described several informatics solutions correlated with operational requirements for smart metering system and our proposal for simplified architecture of smart metering systems, with three distinct levels (base level, middle level and top level) and load profile calculation methods.

Smart metering systems have been implemented due to numerous advantages. They allow a better integration of renewable energy sources by constantly informing consumers about their electricity consumption options. They also encourage better usage of national

grid by reducing load peaks and lifting load off-peaks. In these circumstances, electricity price is a very important key because it stimulates customers to change their behaviour.

Operational requirements regarding smart metering systems have been divided into five categories: requirements for electricity consumers, grid operators, suppliers, security and data protection, distributed production. Data provided by smart meters are extracted, transformed and loaded (ETL process) and then these data could be stored into a data warehouse. For each consumer could be set queries and reports. Results of these analyses are provided to electricity consumers. Data processed as strategic reports is provided to top level managers of grid operators. For this task, business intelligence tools are recommended (data mining, OLAP).

Such as management information system, smart metering systems have a similar architecture. In this paper, we proposed three-level architecture for smart metering system. It contains operational level, middle level and top level. Each level has particular characteristics that were described in this paper.

At *operational level*, the volume of data is tremendous and should be treated by using *Big Data* techniques. Variety and velocity of the data that is provided by consumers' meters and appliances impose new technologies for parallel processing and data organising into *NoSQL databases*. Data collected from meters, sensors and smart appliances are transmitted via telecommunication system to the main server where cleansing and processing by *Apache Hadoop clusters* or *database in memory* take place.

At *middle level*, data transmission can be done by different technologies: general packet radio service – *GPRS*, optic fibre, wireless. The most accepted solution in countries with high level of theft consists in common infrastructure for several utilities (electricity, natural gas and thermal energy) with an intermediary level known as *middleware*, concentrators and balancing meters that rapidly locate theft.

The *top level* includes interactive interfaces by using *dashboards*, prediction analyses, *what if* scenarios, planning and reporting tools. These

elements are obtained with *business intelligence* technologies. The applications that are designed for electricity consumers include information such as 15-minute consumption, electricity price, alerts and consumption thresholds, comparison between their consumption and similar consumers located in the same area, estimations and predictions of electricity consumption. The architecture proposed is flexible and can be adapted to new uncertain challenges that come along with smart metering implementation.

In order to design load profile, we propose two types of analyses related to data classification and clustering techniques. Applying classification techniques, we built the model through decision trees and Support Vector Machines (SVM) algorithm. After several data pre-processing adjustments, the accuracy improved, especially for decision tree – 87%. We obtained a tree with 6 levels and 36 leave nodes, for each node accuracy is given. By analysing tree leaves, we noticed that nodes with the most cases (over 30 thousands) have the highest trust degree (cca. 90%), that is a good result for right estimation of load consumption. Afterwards, we also applied unsupervised learning technique: clustering. In order to build clusters, we specified 10 clusters so that elements in each cluster to be similar, but different from elements from other clusters. For similarity measure in each cluster, we considered variance, choosing clusters with smallest variance.

We applied K-means method for building clusters; Euclidian distance has been calculated. The trust degree is relatively high for the most of the cases, over 85%. Applying this method, we obtained 10 load profiles. P3, P4, P9 profiles differ in terms of shape and level of consumption but P1, P2, P7, respectively P5, P6, P8, P10 profiles are similar as shape, but the level of consumption is different. Due to this aspect, we considered a lower number of clusters, but the results were not significant, the five load profile proved to be very similar so that the representation of consumers' consumption was too generalized.

We also applied O-cluster method (Orthogonal partitioning *clustering*) from Oracle. This

method is using a recursive algorithm of data clustering by data orthogonal partitioning. We obtained hourly consumption patterns. By analysing forming rules and each weight of consumption category within clusters, we detailed each profile obtained by K-means method with hourly consumption patterns obtained by O-cluster method. For instance, we considered profile P5 and hourly consumption patterns related to this profile.

We noticed that consumption patterns models dynamic behaviour of electricity consumers; load profiles are an approximation of variation of hourly consumption. Differences between patterns and load profile are small; therefore it validates our clustering model.

Acknowledgment

This paper presents some results of the research project: *Sistem inteligent pentru predicția, analiza și monitorizarea indicatorilor de performanță a proceselor tehnologice și de afaceri în domeniul energiei regenerabile (SIPAMER)*, research project, PNII – Parteneriate în domeniile prioritare, PCCA 2013, code 0996, no. 49/2014 funded by NASR.

References

- [1] K. Vu, M.M. Begovic, D. Novosel, “Grids get smart protection and control”, *IEEE Computer Applications in Power*, ISSN 0895-0156, 1997
<http://www.cs.ucsb.edu/~chong/290N-W10/smart-grid-alg>
- [2] A.T. Kearney (BERD), Studiul „Contorizarea inteligentă în România”, 2012, <http://www.anre.ro/ro/cauta?s=smart+metering>
- [3] Commission Recommendations of 9 March 2012 on preparations for the roll-out of smart metering systems (2012/148/UE)
<http://eur-lex.europa.eu/legal-content/RO/TXT/PDF/?uri=CELEX:32012H0148&from=RO>
- [4] Ordinul ANRE nr. 145/2014 privind implementarea sistemelor de măsurare inteligentă a energiei electrice
- [5] D.L. Petromanjanc, O. Momcilovic and I. Sceanovic, „Suggested architecture of smart metering system”, *Proceedings of the Romanian Academy, Series A, Volume 13, Number 3/2012, The Publishing House of the Romanian Academy*, 2012, pp. 278–285;
- [6] Electric Power Research Institute, „Advanced Metering Infrastructure (AMI)”, Available:<http://www.ferc.gov/event-calendar/Files/20070423091846EPRI%20%20Advanced%20Metering.pdf>, February, 2007;
- [7] Joint Working Group on Standards for Smart Grids, “Final Report of the CEN/CENELEC/ETSI Standards for Smart Grids”, *CEN/CENELEC/ETSI Joint Presidents Group (JPG)*, Available: <ftp://ftp.cen.eu/PUB/Publications/Brochures/SmartGrids.pdf>, 2011;
- [8] I. Lungu and A. Bâra, “Sisteme informatice executive”, *Editura ASE, București* 2007, ISBN 978 – 973 – 594 – 975 – 4;
- [9] M. Velicanu, I. Lungu, I. Botha, A. Bâra, A. Velicanu and E. Rednic, “Sisteme de baze de date evaluate”, *Editura ASE*, 2009, ISBN 978-606-505-217-8;
- [10] J. Bughin, M. Chui and J. Manyika, „Clouds, big data, and smart assets: Ten tech-enabled business trends to watch”, *McKinsey Quarterly*, Available: http://www.itglobal-services.de/files/100810_McK_Clouds_big_data_and%20smart%20assets.pdf, 2010;
- [11] S. Rusitschka, K. Eger and C. Gerdes, „Smart grid data cloud: A model for utilizing cloud computing in the smart grid domain”, *Smart Grid Communications*, 2010, ieeexplore.ieee.org;
- [12] M. Courtney, „How utilities are profiting from Big Data analytics”, *Engineering and Technology Magazine*, vol 9, issue 1, <http://eandt.theiet.org/magazine/2014/01/data-on-demand.cfm>, 2014.
- [13] S.V. Oprea, A. Bâra, I. Lungu, “Methods for electricity load profile calculation within deregulated markets”, accepted paper, *19th International Conference on System Theory, Control and Computing Joint*

Conference *SINTES 19, SACCS 15, SIMSIS 19*, October 14 - 16, 2015, Cheile Gradistei - Fundata Resort, Romania

- [14] K.L. Lo, Z. Zakaria, M.H. Sohod, "Determination of Consumers' Load Profiles based on Two-stage Fuzzy C-Means", *Proceedings of the 5th WSEAS Int. Conf. on Power Systems and Electromagnetic Compatibility*, Corfu, Greece, (pp. 212-217), 2005
- [15] Md. Jahangir Hossain, A.N.M. Enamul Kabir, Md. Mostafizur Rahman, Borhan Kabir, Md Rafiqul Islam, "Determination of Typical Load Profile of Consumers Using Fuzzy C-Means Clustering Algorithm", *International Journal of Soft Computing and Engineering (IJSCE)* ISSN: 2231-2307, Volume-1, Issue-5, 2011
- [16] Office of Energy Efficiency & Renewable Energy (EERE), OpenEI, <http://en.openei.org/datasets/dataset/commercial-and-residential-hourly-load-profiles-for-all-tmy3-locations-in-the-united-states>, consultat în septembrie 2015



Simona-Vasilica OPREA (b. July 14, 1978) is an Assistant at the Economic Informatics Department at the Faculty of Cybernetics, Statistics and Economic Informatics from the Bucharest University of Economic Studies of Bucharest. She has graduated the Polytechnic University in 2001, holds a Master Diploma in Infrastructure Management Program, Yokohama National University, Japan in 2007 and a PhD diploma from 2009. She is the author of over 25 articles. Domains of competence: wind power plant operation, investment opportunity analysis, studies of prognosis, etc.



Ion LUNGU (b. October 9, 1950) is a Professor at the Economic Informatics Department at the Faculty of Cybernetics, Statistics and Economic Informatics from the Academy of Economic Studies of Bucharest. He has graduated the Faculty of Economic Cybernetics in 1974, holds a PhD diploma in Economics from 1983. He is the author of 22 books in the domain of economic informatics, 57 published articles. He participated in more than 20 research projects. He is a CNCSIS expert evaluator and member of the scientific board for the

ISI indexed journal *Economic Computation and Economic Cybernetics Studies and Research*. He is also a member of INFOREC professional association and honorific member of Economic Independence academic association.