Model-based Performance Evaluation of Large-Scale Smart Metering Architectures

Johannes Kroß,
Andreas Brunnert,
Christian Prehofer
fortiss GmbH
Guerickestr. 25
80805 Munich, Germany
{kross,brunnert,prehofer}
@fortiss.org

Thomas A. Runkler
Siemens AG
Corporate Technology
Otto-Hahn-Ring 6
81739 Munich, Germany
thomas.runkler@siemens.com

Helmut Krcmar
Chair for Information Systems
Technische Universität
München
Boltzmannstr. 3
85748 Garching, Germany
krcmar@in.tum.de

ABSTRACT

Smart meter devices are used to monitor and control energy consumption and are interlinked with smart grids. Their growing use leads to an extensive amount of available data to be processed and causes smart grids to evolve to large-scale systems of systems. Guaranteeing appropriate scalability and performance characteristics is a tremendous challenge. In this paper, we focus on the provisioning of sufficient computing capacity to efficiently analyze the produced data in such a distributed system. For this purpose, we show the use of performance models to plan and simulate this distributed computation in smart grid systems. It demonstrates how different system architectures can be evaluated and required capacities can be estimated to cope with the occurring data volume. We analyze response times for time-critical tasks and assess the scalability of smart grid systems.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Modeling techniques

Keywords

Smart Meter, Smart Grid, Advanced Metering Infrastructure, Performance, Evaluation

1. INTRODUCTION

Smart meter devices are replacing conventional energy meters in several countries and form the basis to manage and monitor energy consumption [10]. These devices are interlinked as part of smart grids and allow two-way communication via interfaces so data can be automatically exchanged between smart meters and energy management operators. A smart grid connects smart energy devices such as smart meters to a distributed energy delivery network that allows for automatic communication and management of devices [10].

LT'15, February 1, 2015, Austin, Texas, USA.

http://dx.doi.org/10.1145/2693182.2693184.

In order to built up a smart grid, an advanced metering infrastructure (AMI) is necessary to manage smart meter devices. It is responsible for connecting distributed smart meters and storing their data [11].

In realistic scenarios, often more than hundred thousand devices are managed by a few central systems and data is continuously exchanged. Hence, the produced data volume is enormous and can easily cause performance issues. The responsibility of these systems (called smart grids) is to store the data produced by smart meter devices and manage them. An additional time-critical task is to calculate optimized energy plans based on the collected data. Therefore, they must combine analytic capabilities with real-time processing [3]. As we consider here large distributed systems, the distribution of this processing is important and has not been considered before. Since the introduction of smart meter devices is growing in many countries, these systems must also be able to scale-up to continuously reach and comply with their performance goals.

Performance models provide a common way to mirror systems and simulate their behavior to guarantee such non-functional requirements [2]. They allow for predicting and measuring performance metrics such as throughput, response time and resource utilization. Performance models can be used for capacity planning as well as to answer sizing questions. They also enable developers to examine design alternatives of architectures and find optimized system configurations. By being able to simulate different workloads on such models, they also support to evaluate a system's scalability.

This paper shows how performance models can be used to model large distributed smart grid systems and simulate hundreds of thousands connected smart meters. While prior work on smart grid performance has mainly focused on the networking aspects [7, 5, 9], we focus here on the computation required for the analysis of the data in large distributed systems. We develop two prototype models presenting two different use cases. For each model, we implement two infrastructure approaches and simulate them with varying amount of smart meter devices. Although both models are kept as simple as possible, they involve the specification of multiple parameters and allow us to already address common problems in the smart grid context. Therefore, we prioritize to analyze the performance metrics utilization, throughput and scalability in this paper.

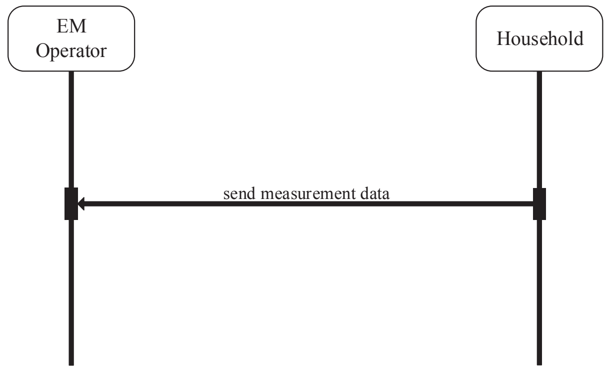

Figure 1: Read smart meters

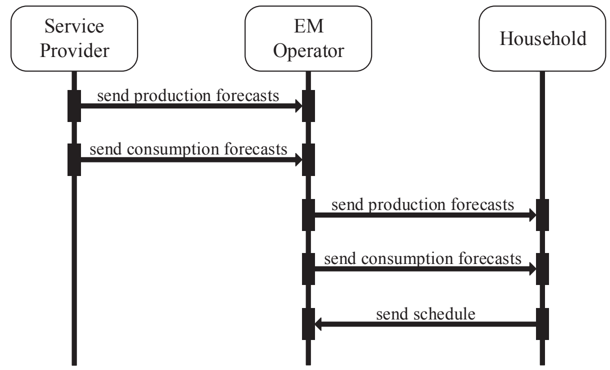

Figure 2: Local optimization

2. USE CASE AND DESIGN OPTIONS

In order to model and evaluate smart meter devices in smart grids we selected two common use cases for our performance model - 'read smart meters' and 'local optimization' - as depicted in Figures 1 and 2. These two use cases are adapted from the E-Energy report [4]. The first use case contains an energy manager (EM) operator and a household as main actors. Smart meters of a household regularly send their measurement data to the EM operator at an interval of five to twenty minutes depending on the type of device. Workload mainly appears at the EM operator who must be able to handle and store the occurring data amount as well as analyze information.

Figure 2 depicts the use case 'local optimization'. The same actors are involved as in the first use case, but it additionally includes a service provider. The latter sends its energy production and consumption forecast data to the EM operator who forwards the information to the household so its smart devices can be locally optimized. Afterwards, the household reports its schedule to the EM operator. In this use case, the EM operator is also connected to the household, but must be able to cope with the data of a household only once a day.

Usually, several thousand households are linked to the EM operator resulting in huge, frequent data volumes. One of the focus questions for the performance evaluation in this paper is therefore the optimal design for an AMI architecture with regard to performance. There are several aims to regard for the operation of an AMI and different solutions have to be weighed. A load reduction at the EM operator as well as minimal hardware costs are desired, but also a reliable solution that is able to cope with network failures and scales up in future. We evaluate two design alternatives for an AMI:

1. A centralized architecture in which an EM operator is directly linked with households as shown in Figure 3.

2. A hierarchical, decentralized architecture similar to [8] as presented in Figure 4. Here, an aggregation system is directly interlinked with households and pre-analyzes their measurement data before it sends them to the EM operator.

3. PERFORMANCE MODEL AND EXPERIMENT DESIGN

To depict and evaluate these design alternatives, we use the Palladio Component Model (PCM) [1]. The PCM metamodel is divided in several sub-models according to different developer roles. Component developers model system components and their behavior, software architects assemble different components into a system, system deployers specify the resource environment and assign resources to different components and, lastly, domain experts create a usage model which describes the workload and usage scenario [1].

We use components to represent each actor of the two use cases and specify their behavior and interfaces based on the descriptions in the E-Energy report [4]. Since we are not interested in investigating resources at households, only one component is modeled as an abstraction that represents all households and causes the same load as several thousand households. For the use case 'read smart meters', smart meter devices send 1 kilobyte (KByte) of data every five to twenty minutes to the EM operator in a centralized architecture and to the aggregator in a decentralized architecture. The aggregator reports the aggregation results in chunks of 100 megabytes (MBytes) to the EM operator four times a day.

For the second use case, smart meters are locally optimized once a day. Production and consumption forecast data of the service provider comprise 10 MBytes. Both forecast data from the EM provider in a centralized and from the aggregation system in a decentralized architecture to each household are sent as messages with a size of 1 KByte each. Similar to the first use case, the aggregator reports 100 MBytes of pre-analyzed data to the EM operator after it has received the schedule of all households.

For the CPU resource demands, the processing time is specified relative to the amount of data that needs to be processed. The EM operator requires 2 milliseconds (ms) CPU time for each KByte, whereas one aggregator demands 1 ms CPU time for each KByte. These demands are based on the assumption that aggregation systems only aggregate data, whereas the EM operator processes the data more intensively using analytical algorithms. For the hard disk drive (HDD) resource demands, both actors demand the full data size of a message. In order to allow for a better comparison between the EM operator and aggregation systems we spec-

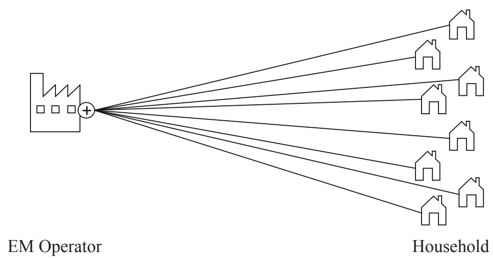

Figure 3: Centralized architecture

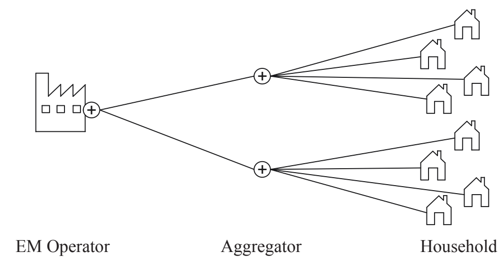

Figure 4: Decentralized architecture

Table 1: Variant table

Workload		System	Hardware
Use Case	Households	architecture	environment
Read smart meters	100,000	Centralized	Constant
Read smart meters	100,000	Decentralized	Constant
Read smart meters	150,000	Centralized	Constant
Read smart meters	150,000	Decentralized	Constant
Read smart meters	200,000	Centralized	Constant
Read smart meters	200,000	Decentralized	Constant
Local optimization	100,000	Centralized	Constant
Local optimization	100,000	Decentralized	Constant
Local optimization	150,000	Centralized	Constant
Local optimization	150,000	Decentralized	Constant
Local optimization	200,000	Centralized	Constant
Local optimization	200,000	Decentralized	Constant

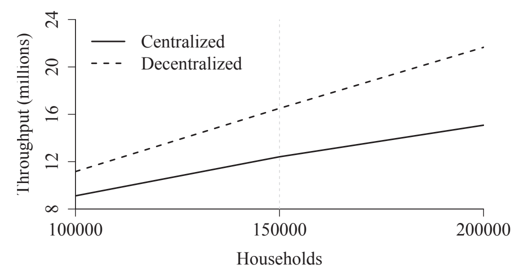

Figure 5: Throughput for use case 'read smart meters'

ified the same hardware resources for all components. We set the processing rate to 60 work units per simulation unit for the CPU, 8,760 MBytes per simulation unit for the HDD and 7,680 MBytes per simulation unit for the network. For the simulation, we determined the simulation unit to be one minute and simulated both use cases 24 hours.

In order to evaluate the modeled AMI and smart grid system, we specify and simulate several different variants of our model as shown in Table 1. Besides the two use cases, the amount of households is alternated from 100,000 over 150,000 to 200,000. Furthermore, the system architecture is also varied. For the decentralized architecture, we modeled an amount of four aggregation systems which are evenly distributed to households so, for instance, for an amount of 100,000 households each of the four aggregators is connected to 25,000 households.

4. SIMULATION RESULTS

Figure 5 shows the predicted throughput of the EM operator in a centralized architecture and the combined throughput of the four aggregation systems in a decentralized one for the use case 'read smart meters'. Having four aggregation systems in parallel, the throughput in a decentralized architecture is increased compared to a centralized one. Whereas the centralized approach is only able to process 15,080,025 smart meter messages of 200,000 households within one day, the decentralized approach can manage 21,655,736. The latter also scales better to higher load levels than a single EM operator as aggregations systems only have to handle a quarter of households.

Table 2 shows the predicted mean CPU utilization during simulating the use case 'read smart meters'. Here, the values at the EM operator are considerably higher than the mean CPU utilization per aggregator. This has several reasons. An aggregator has only to handle a quarter of the workload compared to the EM operator. Furthermore, the resource demand for the EM operator is increased compared to the aggregator as it fully analyzes the data as described in Section 3. Regarding scalability, both systems tend to scale even with increasing user amount.

Table 2: Mean CPU utilization for use case 'read smart meters'

User amount	Centralized (EM operator)	Decentralized (Mean per aggregator)
100,000	21.10%	3.82%
150,000	28.73%	5.37%
200,000	34.91%	6.86%

Considering the use case 'local optimization', we simulated a time-critical scenario where the EM operator tries to send all consumption forecasts as fast as possible to each household. Here, response time is the most interesting metric as listed as minute values in Table 3. Clearly, response times in a decentralized architecture are faster than in a centralized one due to four parallelized aggregation systems. The response time between the EM operator and the four aggregation systems is here not considered as it is negligibly small. For 200,000 households the centralized architecture needs 42.34 minutes to reach all households, which equals 12.70 milliseconds per household. The decentralized architecture only takes 7.62 minutes, which equals 2.29 milliseconds per household if aggregators are fully parallelized.

Answering our focus question, a decentralized approach seems to allow for higher throughput and faster response times. Having several smaller aggregations systems can also

Table 3: Response time sending consumption forecasts for use case 'local optimization'

Households	Centralized (EM operator)	Decentralized (Mean per aggregator)
100,000	21.62 minutes	5.79 minutes
150,000	31.98 minutes	6.81 minutes
200,000	42.34 minutes	7.62 minutes

distribute load in peak times and decrease utilizations of the main system which may have cost advantages regarding required hardware resources. Since we modeled two simple use cases separately, however, it is not possible to predict major impacts of both architectures on the overall system performance. Thus, our model will serve as a basis for future efforts to address several more complex scenarios.

5. RELATED WORK

Several approaches have been developed to model AMI that allow for analyzing and simulating smart grids. However, most of them mainly focus on evaluating the network. Mora et al. [7] modeled the network communication as needed for smart grids, but did not consider hardware resources such as CPU load for processing. Lin et al. [5] used a similar approach, but with a co-simulation setup where only network is simulated. This procedure is also not suitable to evaluate larger sets of scenarios as we intend to do in this paper. Besides different network technologies, the paper of Wang et al. [9] surveys architectures for various smart grid scenarios, including discussion of requirements. The paper does however not discuss how to distribute data collection and analysis in such networks. From the EM operator point of view required hardware, response times and system scalability of smart grids are vital bottlenecks that cannot be answered with the above mentioned models.

6. CONCLUSION AND FUTURE WORK

This paper shows how performance models can be applied to evaluate smart metering infrastructures including the computation needed for data analysis. Although there are several approaches available that model and analyze smart grids, they specialize on network traffic and not on predicting performance metrics such as resource utilization, throughput and response times. The application of performance models for this purpose allows for predicting such performance metrics and supports EM operators while answering sizing questions and evaluating different architecture designs.

We implemented two common use cases for smart metering infrastructures and evaluated them using a centralized and a decentralized architecture. Different amounts of households have been applied to generate workload. Therefore, response time, throughput and CPU utilization have been predicted and evaluated for different parameter configurations. Regarding system performance the results indicate advantages in response times and throughput of a decentralized architecture over a centralized one.

One of the key challenges in this work was to run simulations in scales of up to 200,000 households using the available PCM simulation engines. Initially, we employed the process-driven simulation engine SimuCom [1], which was not able to scale up to this workload. Therefore, we switched to the event-oriented PCM simulation engine EventSim [6]. Although it was able to scale up to these loads, simulating way beyond these load levels might require different approaches.

In future work, we plan to extend the models with resource demand measurements for specific data analytic platforms. Furthermore, we plan to include reliability evaluations in the simulations. Afterwards, the creation of performance models for additional smart grid scenarios is planned, including simulations of several smart grid use cases in parallel.

7. REFERENCES

[1] S. Becker, H. Koziolek, and R. Reussner. The palladio component model for model-driven performance prediction. *The Journal of Systems and Software*, 82(1):3–22, Jan. 2009.

[2] A. Brunnert, C. Vögele, A. Danciu, M. Pfaff, M. Mayer, and H. Krcmar. Performance management work. *Business & Information Systems Engineering*, 6(3):177–179, 2014.

[3] D. Ilic, S. Karnouskos, and M. Wilhelm. A comparative analysis of smart metering data aggregation performance. In *Proceedings of the 11th IEEE International Conference on Industrial Informatics*, pages 434–439. IEEE, July 2013.

[4] M. Irlbeck and V. Koutsoumpas. E-Energy Abschlussbericht. Report, Technische Universität München, 2013.

[5] H. Lin, S. Sambamoorthy, S. Shukla, J. Thorp, and L. Mili. Power system and communication network co-simulation for smart grid applications. In *Proceedings of the First IEEE PES Conference on Innovative Smart Grid Technologies*, pages 1–6. IEEE, 2011.

[6] P. Merkle and J. Henss. EventSim – an event-driven Palladio software architecture simulator. In *Proceedings of the 2nd Palladio Days (appeared as technical report)*, Karlsruhe Reports in Informatics 2011,32, pages 15–22, Karlsruhe, 2011. KIT, Fakultät für Informatik.

[7] R. Mora, A. López, D. Román, A. Sendín, and I. Berganza. Communications architecture of smart grids to manage the electrical demand. In *Proceedings of the 3rd Workshop on Power Line Communications*, 2009.

[8] T. Overman and R. Sackman. High assurance smart grid: Smart grid control systems communications architecture. In *Proceedings of the First IEEE International Conference on Smart Grid Communications*, pages 19–24. IEEE, Oct 2010.

[9] W. Wang, Y. Xu, and M. Khanna. A survey on the communication architectures in smart grid. *Computer Networks*, 55(15):3604–3629, 2011.

[10] J. Zheng, D. Gao, and L. Lin. Smart meters in smart grid: An overview. In *Proceedings of the 5th IEEE Conference on Green Technologies*, pages 57–64. IEEE, April 2013.

[11] J. Zheng, Z. Li, and A. Dagnino. Speeding up processing data from millions of smart meters. In *Proceedings of the 5th ACM/SPEC International Conference on Performance Engineering*, ICPE '14, pages 27–37, New York, NY, USA, 2014. ACM.