Appears in Proc. of the 5th Usenix Workshop on Hot Topics in Cloud Computing (HotCloud), June 2013.

Cloud Computing for the Power Grid: From Service Composition to Assured Clouds

György Dán
KTH Royal Institute of Technology
Stockholm, Sweden

Rakesh B. Bobba George Gross Roy H. Campbell
University of Illinois Urbana-Champaign
Urbana-Champaign, USA

Abstract

The electric power industry is one of the few industries where cloud computing has not yet found much adoption, even though electric power utilities rely heavily on communications and computation to plan, operate and analyze power systems. In this paper we explore the reasons for this phenomenon. We identify a variety of power system applications that could benefit from cloud computing. We then discuss the security requirements of these applications, and explore the design space for providing the security properties through application layer composition and via assured cloud computing. We argue that a combination of these two approaches will be needed to meet diverse application requirements at a cost that can justify the use of cloud computing.

1 Introduction

The electric power industry was among the first adopters of mainframe computers in the 1960s, to build computerized automation systems that can efficiently monitor and control the power grid. Major power utilities today operate multiple redundant data centers with hundreds to thousands of utility grade computers to satisfy the computational needs of monitoring and control applications. Some utilities have recently migrated their customer facing applications to hosting service providers, but the use of virtualization and cloud computing has found little or no adoption in operations related applications.

In contrast, the banking industry, which has been traditionally risk averse, is embracing cloud technology in an effort to meet customer demands quickly. Although most cloud deployments in the banking industry have been private clouds, the first core banking system went into production on a public cloud recently on a proprietary platform[1].

[1] http://www.temenos.com/products/t24/

One thus wonders whether the electric power industry is avoiding the use of cloud technology because it does not offer any significant benefit compared to dedicated computing infrastructures, or because with today's cloud technology the risks by far outweigh the benefits. If it is the latter, what kind of developments would be needed in cloud technology to mitigate the risks. Could the risks be mitigated by choosing the right cloud deployment model and how would the deployment model affect the potential benefits. Alternatively, could some of the risks be mitigated without explicit support of the cloud infrastructures, and if so, what would be their cost.

In this paper we explore these questions by characterizing the computational and security requirements of various existing and future core power system applications. We contrast these with the state-of-the-art in cloud computing and identify a number of points where current cloud technology falls short. Finally, we discuss potential solutions that could address the shortcomings at the application layer, keeping in mind that these solutions might easily nullify the benefits of moving to the cloud. Our primary focus is on core yet less critical power system applications, whose temporary unavailability would not immediately impact the reliable operation of the power system.

2 Power Applications in the Cloud

We start with identifying what would make it beneficial to move power applications to the cloud. We consider applications used in the power markets, in planning and in operations.

Time-varying computational needs Many of the application used in the planning and operation of power systems have time varying computational needs, for one of two reasons. Some applications are used periodically or occasionally, but need significant computational capacity when used. Others are used continuously but the amount

of computations they need is a function of the state of the power system, which changes with time.

Forecasting and Planning (FAP) are performed over various timescales ranging from minutes up to ten years. Short term forecasts are used for operations, and are obtained based on historical loads and weather forecasts using different prediction methods, e.g., neural networks. Weather forecasts are also used in predicting the transmission capacity of the grid, and with the penetration of renewables, to forecast wind and solar generation by numerically solving vast numbers of systems of differential equations. Studying the feasibility of integrating renewable energy sources into a constrained transmission network requires extensive high fidelity simulations under various load, generation and weather conditions. These analyses are highly parallelizable and could significantly benefit from the elasticity provided by cloud computing. The computation of long term load forecasts is based on historical measured demands and on expected societal developments and is typically aimed at identifying the peak load over a time period. The forecasted peak load is used in generation and transmission studies to support the planning and coordination of infrastructure maintenance (planned outages) and that of infrastructural investments [9]. Long terms forecasts are performed rather infrequently.

Locational Marginal Pricing (LMP) is used to calculate the price of electricity as a function of the forecasted demand for the real-time (spot) market and for the day ahead market.

In integrated utilities the LMP is obtained by solving the optimal power flow (OPF) problem given the predicted loads, known generation costs and transmission capacity, and is often called economic dispatch. In its simplest form the OPF problem can be cast as a quadratic optimization problem subject to network capacity and generation constraints and known demands, but in reality the OPF needs to be solved for uncertain demands over several connected congestion areas and under security constraints over a receding horizon and often involves stochastic dynamic programming.

In power markets electricity prices are determined by market operators or by independent system operators. Offers and bids are submitted by market participants before the market closing time for the considered time interval (entire day or particular hour), which can be several days or hours ahead of the considered time interval. After that the bids, offers and capacity constraints are published by the market operator. The locational marginal prices are determined based on the load forecasts, the generators' submitted offers, and the load serving entities' submitted bids, considering power system capacity constraints and transmission losses.

In both cases, the frequency of the computations for the real-time prices can be as high as once every five minutes, but day ahead computations are typically performed once a day.

Topology/parameter error detection (TPD) is performed by power system operators to identify potential errors in the parameters of the system models they maintain. Topology error detection relies on measured data and can be performed at different timescales. It can be performed jointly with state estimation and bad-data detection, which involve solving a non-linear weighted least squares minimization problem and calculating residuals. This type of calculation is performed as often as once per minute based on the most recently obtained measurement data, but due to potential bad measurement data it is not very reliable. More reliable topology error detection can be performed based on historical measurement data and using various signal processing algorithms, such as independent component analysis [10]. Such processing is typically done off-line once sufficient measurement data are available.

Contingency analysis (CA) is used to identify whether a set of identified contingencies (a contingency is the failure of one or more system components) would result in an unstable system given the instantaneous system state. The impact of a contingency is calculated by solving a non-linear weighted least squares (WLS) estimation problem using an iterative algorithm. Alternatively, the impact can be approximated by using a linearized system model (called DC model) in which case it involves the solution of a linear WLS problem, which can be obtained through matrix multiplication and inversion. The number of contingencies that needs to be considered depends on the instantaneous load of the power system, the higher the load the more contingencies might have to be considered. The number of contingencies considered in practice is limited by the computational power available in the control center, and is often constrained to considering the loss of a single components known as $n-1$ security. CA is typically performed every time the system state is recalculated, which can happen as often as once a minute. Cloud-based contingency analysis could allow an operator to scale the number of considered contingencies freely as a function of the instantaneous system state.

Reliable data storage Power system operators maintain data historians, which are databases of past system states and events. Historians are are maintained for various reasons: to enable the reconstruction of events that led to system failure (forensics), to facilitate improving efficiency through e.g. data mining (planning) and for compliance with regulatory requirements. Although traditional SCADA historians need to store data for a few thousand data points a few times per minute, it is not uncommon for utilities to generate and store 100TB of data

annually. With the advent of Phasor Measurement Units (PMUs), which sample the system 30 times a second or more in contrast to once every few minutes as was done previously, the amount of data to be stored will soar. Current best practices implement reliable storage by replicating data at various locations, although data are only occasionally accessed from the replicas. Cloud-based storage could be a cost-efficient solution for such quasi write-only data with current pricing models of cloud storage, which typically involve charges based on the amount of data stored in the cloud and the amount of data downloaded from the cloud.

Ease of access Operational data are increasingly used for business purposes, such as billing and planning, and are also needed for coordination among operators and for reporting to regulators. The industry standard solution for allowing access to operational data to business applications and other entities is to place data in a demilitarized zone. Since the number of entities that need access to the data can potentially be large, configuring and enforcing access control without compromising control center security is a challenge. Cloud-based storage would allow an operator to move the data outside its security perimeter, and thus instead of managing access of multiple entities into its security perimeter, the operator only has to ensure that the transfer of the data to the cloud is secure and manage access to the data. Cloud-based access to operational data by trusted parties could be applicable to traditional SCADA measurement historians, but it is especially useful for sharing PMU data to realize wide-area monitoring, which is identified as a priority by the U.S. Federal Energy Regulatory Commission (FERC), and would be compatible with the NASPInet[2] architecture under development for sharing PMU data.

Open market of utility-oriented computing services Today, power system operators are virtually locked to a particular SCADA/energy management system (EMS) vendor. It is often cheaper for an operator to stay with a vendor and purchase additional EMS features at a higher cost than offered by competitors rather than to pay for integrating systems developed by different vendors. Adoption of cloud based computing may usher in more open and standard data formats and communication protocols, opening up a market of data analysis and forecasting as a service, and fostering innovation.

While the above examples suggest that cloud computing could provide significant cost and performance benefits for various power system applications, unprecedented data storage and processing requirements due to the advent of PMUs, and unprecedented need for fast simulations to integrate renewables into a constrained network may become the most important drivers for the adoption of cloud computing by the power industry.

[2]https://www.naspi.org/site/Module/Home/naspinet.aspx

3 Security Requirements

The power grid is a critical infrastructure on which many other critical services and consequently the safety and health of the society depend upon. In fact loss of power could impact the availability or accessibility of the cloud infrastructure itself. Therefore, the migration of power system applications to the cloud would require stringent security guarantees in order to meet the reliability needs of power system operation and be compliant with regulations. The requirements can be very diverse, depending on the application.

Confidentiality and Privacy A variety of information used by power system applications have to be kept confidential by regulation. Such data include power system characteristics, such as the active topology and the system parameters, and potentially the current system state. There are two distinct reasons for the need of confidentiality. First, an adversary could leverage the information to infer expected congestion in the power system and could use this knowledge for insider trading in the forward (e.g., day-ahead) power markets. Second, an adversary with knowledge of the current power system state could infer a critical contingency (i.e., a failure scenario that would cause instability) and could launch a targeted attack to implement it. The confidentiality requirements of the data might depend on the timescale at which the computation is performed. For example, load forecasts, offers and bids submitted for calculating the price in the day-ahead market can be used to infer future electricity prices, and would need to be kept confidential. The same data used for real-time pricing cannot be used for obtaining profit, and can therefore be public. At the same time, the current system state might need to be kept confidential as it can be used to infer potential contingencies.

An even stricter requirement could be that the execution of certain applications should not be visible to third parties, as one could potentially use the information to infer the state of the system. For example, by observing significant CA being performed an attacker could infer that a power system is critically loaded.

Integrity/Correctness The integrity of data and computations is paramount to power applications. Depending on the application, the results of computations can impact planning or operational decisions which in turn may have profound implications for the grid. For example, computational results may indicate feasibility of renewable integration or may list important contingencies under certain conditions. If the integrity of computations or data is violated operators may take undesirable planning

or control actions, which can affect grid stability.

Availability Availability is a major concern for power system applications that provide real-time situational awareness and control. Under normal conditions a power system can operate without operator intervention by design, and thus without relying on power system applications. Nevertheless, real-time situational awareness is essential when load conditions change and intervention is needed to maintain system stability. NERC CIP-002-5 defines a bulk electric system critical cyber asset as one whose disfunction over 15 minutes would affect the reliable operation of the power system. Placing such cyber assets into the cloud would thus impose a requirement on the availability of the network infrastructure connecting to the cloud and of the cloud infrastructure itself, even in the face of adversarial activity. Power system operators prefer 24×7 availability with immediate response times for situational awareness and control applications.

Legal compliance Operators of critical infrastructures, like the power system, have to comply to various policies mandated by regulatory agencies. For example, legislation in many countries prohibits the storage and processing of sensitive data on foreign territory. Thus, the use of a cloud-based storage whose physical location is not controllable by the operator could raise legal issues, potentially even if the data are encrypted, as the question of data ownership might be disputed.

4 The Solution Space

Having outlined potential power applications and their security requirements, we now explore the solution space. We consider application deployment in a cloud infrastructure managed by a non-trusted third party, such as a community, sovereign, public or hybrid cloud [13]. A private cloud would not raise many of the security issues, but the benefits of migration are not substantial unless the potential for multiplexing various applications of the same power operator is significant.

4.1 Composing Assured Services

At one end of the solution space security guarantees have to be implemented at the application layer with no support for assured computing from the cloud provider, as is the case in today's public clouds.

In this scenario the necessary confidentiality and privacy protection could potentially be achieved using encryption techniques. While single-user encrypted storage can be implemented easily, efficient multi-party access to encrypted data is not straightforward, and search and query on encrypted data are challenging. Computations could in principle be performed directly on ciphertext (encrypted data) using homomorphic encryption (*e.g.,* [7]). This would allow a grid operator to send encrypted instances of computational problems to the cloud, computations in the cloud are performed only on encrypted data, and the operator obtains the solution by decrypting the output sent back from the cloud. However, today's homomorphic encryption techniques are limited in the types of computation that they can support, and those that could compute arbitrary functions [7] impose significant computational overhead and are far from practical.

Alternative approaches to protecting the confidentiality include transforming the original computational problem [1], introducing invertible perturbations or noise into the problem, splitting the problem and solving each subproblem in a different cloud platform. Each of the alternative approaches leaks information to varying degrees and may or may not be suitable depending on the application under consideration. Further, these approaches have to be developed specifically for every power system application. Transforming a power flow optimization is quite a different problem compared to transforming state estimation, for example. These approaches not only impose computational overhead on the operator to transform the problem and to recover the results, but they may also induce unwanted latencies, and some of them may also impact the accuracy of the results, for example, depending on the invertibility of introduced perturbations.

To ensure the integrity of the computations and the results, several instances of the same problem can be executed on different platforms and their results accepted using, for example, a majority voting scheme. For certain problems, a low complexity verification of the result may be possible. For example, for certain decision problems, *e.g.,* NP-complete problems such as the existence of a Hamiltonian in a graph, it can easily be verified if a result actually solves the problem, but a negative result - there is no feasible solution - cannot be verified. However, most power system applications do not solve decision problems, but various optimization problems for which the verification of the result is often as computationally intensive as its computation. Another option could be to solve a low fidelity version of the problem inhouse and compare the results with those returned from the cloud. For example, lists of contingencies computed using a full AC model (non-linear) could be compared to a list of contingencies computed locally using a DC (linearized) model or linear sensitivities. Again each of these approaches imposes overheads or costs on the operator. For example, solving replicated instances will increase the costs linearly with the number of instances.

The availability of cloud-based computations can be improved by replicating network connections to cloud service providers and by replicating cloud-based computations and storage over several independent clouds.

The use of several clouds is costly for two reasons. First, different clouds might provide different application programming interfaces, services and management solutions, which makes cross-cloud application development complex and expensive in lack of a standardized abstraction. Second, simultaneous execution of the same computation on different clouds provides the highest level of availability but at the price of increased cost; similar to replicated storage. Stateless failover could be used to migrate computations from one cloud to the other, but this requires the decomposition of power system applications to support stateless failover. If this is not possible, stateful failover could be used but this could potentially require the live migration of virtual machines between different cloud providers. Alternatively, an abstraction for application checkpointing could be used, by preference one that is compatible between different cloud providers. The question is whether such checkpointing could be implemented at a reasonable cost in terms of extra computational load (and thus cost).

4.2 Assured Cloud Computing

At the other end of the solution spectrum is assured cloud computing. Cloud providers, especially community cloud providers, could aim to provide all the desired security properties through assured cloud platforms. Such settings should aim to address threats from undesirable inter-tenant interactions, and from malicious insiders. Confidentiality and integrity of communications to and from the cloud, and within the cloud can be ensured by standard cryptographic protocols (*e.g.,* TLS, IPSec). So here we focus on the security of the client VMs to ensure confidentiality of their data and integrity of their computations.

To ensure that the confidentiality of the data and the integrity of the computations are not violated by undesirable inter-tenant interactions, strong guarantees of isolation between tenant VMs are needed. To ensure the same protection against malicious insiders (*e.g.,* server administrators), protections for client VMs against the host Virtual Machine Managers (VMMs) are needed. The threat of undesirable interactions is realistic given i) commodity VMMs execute full-fledged operating systems and there are a number of vulnerabilities (*e.g.,* [4, 5]) in those VMMs that malicious tenant VMs may be able to exploit, and ii) VMMs have privileges to inspect the state of client VMs.

Both these issues received considerable attention from the research community. Techniques for strengthening VMMs against certain classes of attacks [17], and architectures for elimination of VMMs [15] are among the solutions proposed to mitigate the issue of vulnerabilities in commodity VMMs. To prevent privilege escalation through compromise of VMM and to protect client VMs from the VMM, many approaches that use privilege separation (*e.g.,* [12, 3, 2]) to disaggregate the privileges of the host VM (VMM) have been proposed. Of these the self servicing cloud computing (SSC) model proposed in [2] is notable as it reduces risks to the client VMs through the host while allowing tenants flexible control over their VMs. These solutions may address the issue of confidentiality and integrity violations that may result from the compromise of the host VM, but they do not address confidentiality violations through side-channels (*e.g.,* [14, 18]). Also, unless formal verification of the VMM layer becomes feasible, one cannot assure the absence of vulnerabilities at the VMM layer. Furthermore, some of these solutions come with their own challenges. For example, SSC depends on Trusted Platform Modules (TPMs) and raises the issue of managing and distributing TPM keys to end users.

In order to ensure availability and timeliness for power grid applications, cloud infrastructures should be enhanced to provide better reliability, automated redundant provisioning and failure handling. The GridCloud Project [11] is undertaking research in this direction. While redundant provisioning and failure handling may improve reliability, timeliness may still be impacted by malicious tenants (*e.g.,* [16]) and by the inherent performance heterogeneity in clouds [6]. In order to ensure an appropriate sharing of resources among tenants, to mitigate resource starvation attacks by malicious tenants and to reduce the impact of performance heterogeneity, better resource allocation and scheduling algorithms (*e.g.,* [8]) may be needed at the VMM layer. It is important to note that the appropriate resource sharing might be far from fair if, for example, power grid applications coexist with best effort applications in the cloud.

Besides computational resources, the network resources have to be provisioned to ensure security and availability. The network traffic of different applications could be isolated using Virtual LANs or using flow isolation through software defined networking. There are several known exploits for VLAN hopping, which allows compromising VLAN traffic isolation, although these exploits can be mitigated by proper system configuration. The security of software defined networking (*e.g.,* OpenFlow) based approaches are not well understood.

Finally, sovereign clouds, which are a specific form of community clouds whose physical location is guaranteed to be within a nation, could address the legal issues that arise from the physical location of cloud data centers. An alternative way of settling the issue of sovereignty is to include requirements in SLAs regarding the physical placement of the data. Restrictions on the physical placement of certain data might appear as a new constraint in cloud storage optimization and selective replication.

4.3 Putting it all together

While there are potential solutions that can address individual security concerns, either in the form of cloud assurance or implemented at the application layer, there are no ready-made solutions that address all security concerns and can provide all the necessary security properties. Further research is needed to understand whether the individual solutions can be composed to provide the necessary security properties at an adequate level. Given the heterogeneity of the security requirements of the power applications, we believe that it will not be economical to provide all security features in the form of cloud assurances, but a basic set of assured cloud computing services will be used to compose secure and available applications.

It is not clear either what the cost of providing the security features would be and whether providing these properties would increase the price of cloud computing to make it unattractive. Even if providing all the security properties is not cost effective, an understanding of what properties can be provided at a reasonable cost is needed. Today, data centers are built using commodity hardware meant for personal or enterprise use. But as the cloud computing paradigm takes off it is conceivable that hardware tailored for multi-tenancy with certain built-in protections will become available paving the way for assured cloud computing that can support power grid applications at a reasonable cost.

5 Conclusion

We believe that there is a potential for moving operational power grid applications to cloud computing. We expect that the main driver will not only be cost savings on computation, but the ability to occasionally scale computations as needed, which can improve power system reliability, and the ease of storing, managing and exchanging data between different entities. The question is whether the security requirements can be addressed at a complexity and cost that do not out weigh the benefits of moving to clouds. This question is particularly interesting, as the diverse requirements open up a potentially rich area of composing secure cloud services for power applications on top of partially trusted cloud infrastructures. Furthermore, the algorithmic characteristics and computational demands of power system applications make the solution space significantly different from that of other industries, *e.g.,* the banking industry.

Acknowledgements

This material is based upon work supported in part by the Department of Energy under Award Number DE-OE0000097. G. Dán was a Fulbright research scholar at UIUC while this work was performed.

References

[1] BORDEN, A. R., MOLZAHN, D. K., RAMANATHAN, P., AND LESIEUTRE, B. C. Confidentiality-preserving optimal power flow for cloud computing. In *Allerton Control Conference* (2012).

[2] BUTT, S., LAGAR-CAVILLA, H. A., SRIVASTAVA, A., AND GANAPATHY, V. Self-service cloud computing. In *Proc. of ACM CCS* (2012), pp. 253–264.

[3] COLP, P., NANAVATI, M., ZHU, J., AIELLO, W., COKER, G., DEEGAN, T., LOSCOCCO, P., AND WARFIELD, A. Breaking up is hard to do: security and functionality in a commodity hypervisor. In *Proc. of ACM SOSP* (2011), pp. 189–202.

[4] CVE-2007-4993. Xen guest root escapes to dom0 via pygrub.

[5] CVE-2008-2100. Vmware buffer overflows in vix api let local users execute arbitrary code in host os.

[6] FARLEY, B., JUELS, A., VARADARAJAN, V., RISTENPART, T., BOWERS, K. D., AND SWIFT, M. M. More for your money: exploiting performance heterogeneity in public clouds. In *Proc. of ACM Symposium on Cloud Computing* (2012), pp. 1–14.

[7] GENTRY, C. Computing arbitrary functions of encrypted data. *Commun. ACM 53*, 3 (Mar. 2010), 97–105.

[8] GHODSI, A., ZAHARIA, M., HINDMAN, B., KONWINSKI, A., SHENKER, S., AND STOICA, I. Dominant resource fairness: fair allocation of multiple resource types. In *Proc. of USENIX NSDI* (2011), pp. 24–24.

[9] HENDERSON, M., WONG, P., AND PLATTS, J. Power system planning process and issues. In *Proc. of IEEE PES General Meeting* (2009).

[10] LIAO, H., AND NIEBUR, D. Load profile estimation in electric transmission networks using independent component analysis. *IEEE Trans. on Power Systems 18*, 2 (2003), 707–715.

[11] MAHESHWARI, K., LIM, M., WANG, L., BIRMAN, K., AND VAN RENESSE, R. Toward a reliable, secure and fault tolerant smart grid state estimation in the cloud. *IEEE PES Innovative Smart Grid Technologies* (2013).

[12] MURRAY, D. G., MILOS, G., AND HAND, S. Improving Xen security through disaggregation. In *Proc. of ACM Intl. Conf. on Virtual Execution Environments* (2008), pp. 151–160.

[13] PETER MELL, T. G. The NIST Definition of Cloud Computing. NIST Special Publication 800-145, September 2011.

[14] RISTENPART, T., TROMER, E., SHACHAM, H., AND SAVAGE, S. Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. In *Proc. of ACM CCS* (2009), pp. 199–212.

[15] SZEFER, J., KELLER, E., LEE, R. B., AND REXFORD, J. Eliminating the hypervisor attack surface for a more secure cloud. In *Proc. of ACM CCS* (2011), pp. 401–412.

[16] VARADARAJAN, V., KOOBURAT, T., FARLEY, B., RISTENPART, T., AND SWIFT, M. M. Resource-freeing attacks: improve your cloud performance (at your neighbor's expense). In *Proc. of ACM CCS* (2012), pp. 281–292.

[17] WANG, Z., AND JIANG, X. Hypersafe: A lightweight approach to provide lifetime hypervisor control-flow integrity. In *Proc. of IEEE Symp. on Security and Privacy (SP)* (2010), pp. 380–395.

[18] ZHANG, Y., JUELS, A., REITER, M. K., AND RISTENPART, T. Cross-vm side channels and their use to extract private keys. In *Proc. of ACM CCS* (2012), pp. 305–316.