Scalable, Secure Energy Information Management for Demand-Response Analysis

Yogesh Simmhan[1,2], Saima Aman[1], Baohua Cao[1], Mike Giakkoupis[1], Alok Kumbhare[1], Qunzhi Zhou[1], Karthik Gomadam[1] and Viktor Prasanna[1]

[1]*University of Southern California, Los Angeles CA* [2]*Microsoft Research, Los Angeles CA*

The advent and growth of smart energy grids is increasing the ability to monitor and communicate power supply, pricing, and demand among utility providers and consumers. While the smart meter infrastructure is expanding at a rapid rate to enable communication using emerging standards, the software architecture to collect, manage, analyze, scale, and secure the information constantly streaming from the grid is still being designed. Effective integration, analysis, and feedback of energy information are essential for the benefits of smart grid to propagate to the various stakeholders: power utilities, residential, commercial, and institutional consumers, and service and application providers. This can lead to a lower peak demand on utilities to ensure regular supply of quality power, reduced consumption, and costs for consumers by making them aware of and giving them control over their power profile, and help organizations to plan and optimize energy usage to meet sustainability goals.

Managing the energy information lifecycle – from the events streaming from smart meters through the smart grid, to meaningful analysis and feedback to utilities and consumers – presents several opportunities for software systems research identified below.

1. Security and Privacy in Stream Processing Systems

The communication between stakeholders in the smart grid is periodic, taking the form of streams of control and data events such as power usage, production, pricing, feedback, and query/response interactions, which are transmitted across a widely distributed infrastructure. These entities have different roles, and the streams they produce and consume have diverse data privacy requirements that need to be strictly enforced to ensure secured data is not leaked to unauthorized entities. For e.g., a utility company wants its production data to be accessible only to ORG-X's internal processing system. A residential consumer wishes to authorize a specific analysis application she is subscribed to, and no other, to have access to home appliances she is using and their usage pattern. A medical emergency or disaster preparedness agency imposes relaxed privacy requirements during emergency situations.

A Federated stream processing system should have the capabilities to specify and enforce these different policies, and allow controlled access to objects of different granularities (e.g. streams, events types and attributes) by a subscribed application. The types of accesses themselves may be restricted to specific APIs based on application contracts. The system must be resistant to different forms of attack, such as impersonation, denial of service, and data injection, by a nefarious entity.

Our approach to address these issues involves examining existing protocols for communication networks, and demonstrating their applicability (or not) for the smart grid. In particular, we will consider two specific challenges imposed by the smart grid in the context of cyber-security – the real time nature of data and the large scale at which it operates.

2. Semantic Complex Event Processing for Information Integration in the Smart Grid

The benefits of Smart Grid include applications that predict peak-demand, provide dynamic pricing, and perform automatic outage analysis to improve the reliability and efficiency of power grids. The information for these decisions comes from diverse sources: smart meters that report near real-time power usage and quality, intelligent thermostats that measure and control buildings heat and humidity, and even weather forecast from online services. Taking advantage of the large quantity of data influx requires meaningful integration of heterogeneous data streams and combining them with prior knowledge.

We propose a solution to this crucial information integration problem using a combination of complex event processing (CEP) and Semantic Web technology. CEP abstracts incoming data as high-level events that provide insight into new information as it arrives. Semantic Web provides an ontology-based framework for data to be formalized, shared, and reused across application and community boundaries. Our Smart Grid semantic complex event processing (SCEP) approach models Smart Grid data as events and complex event patterns using ontologies developed for the domain. This helps to detect event patterns over multiple data streams based on semantic pattern matching. Addition analysis can establish causal, membership, temporal and spatial relationships between Smart Grid events.

The consequence of this is a more meaningful synthesis of structural smart grid events with better understanding of their relationships. This makes it easier to identify the most relevant events from the distributed and heterogeneous data sources, analyze their impact, and take subsequent action in real time. Since the event data conforms to a formal semantics, domain experts will be able to encode their knowledge of meaningful event patterns and define responses to events using declarative languages.

3. Cloud Computing for Scalable Smart Grid Information Processing and Management

Smart grids are large-scale systems, encompassing millions of power consumers, each of whom is an information provider and consumer. The information transferred in the form of event streams represents power usage, current pricing, system load, current power production ability, and weather data. These high frequency streams (10-100/min) from consumers/utilities accumulate to gigabytes of data per day. Besides storing and managing them for future queries and analysis, complex computation and prediction models need to be run on the streams in realtime to respond to consumer targets and smart grid load.

The software architecture for managing, processing, and responding to smart grid information requires a scalable infrastructure. Cloud computing is emerging as a viable platform for on-demand, flexible access to collocated compute and data resources with low administrative overhead. Clouds present several benefits for our smart grid software architecture. Commercial cloud datacenters are built to support online access by millions of web users. This allows Cloud IP endpoints to accept input streams from smart grid information providers at scale. Structured and unstructured Cloud data storage can scale to terabytes of data and billions of tuples, and help manage the SG information over time. Virtual machines (VM) in the Cloud can encapsulate different information processing requirements. Specific VM roles can be assigned to data ingest, transformation, analysis and mining, with the number of instances of each calibrated to match the current processing load. For e.g., a VM role can categorize a consumer based on consumption pattern, while another can run a prediction model to extrapolate future load, while yet another can provide feedback to consumers and utilities on curtailing consumption to meet energy targets and predictions of peak-load.

Several research issues present themselves in constructing a smart grid software architecture in the Cloud. Issues of security and privacy have to be addressed. Compute requirements need to be dynamically estimated to ensure sufficient VMs are available for low-latency response, and yet limit Cloud resource usage to curtail costs. Stream data computing and workflows abstractions are needed to map consumer and utility policies to Cloud computation that are monitored and reliably executed by agents. Data has to be efficiently modeled in databases and blob storage to support optimal computation, and access by authorized external smart grid applications through web service protocols. Provenance of demand and response data has to be tracked across consumers, smart grid applications, and utilities. The outcome of these research efforts will be a highly scalable software system deployed on a hybrid Cloud consisting of both public and private Clouds, and able to reliably process realtime streaming data.

4. Machine Learning for Predictive Modeling of Energy Usage

Two key issues in smart energy management are to predict the energy usage behavior of customers and to track the changing usage trends across consumer groups over time. This can make customers aware of their energy usage with respect to personal usage targets, assist in choosing appliances by predicting their impact on consumption, and help explain building-wise consumption patterns. Utility companies benefit from the ability to predict peak demand, target incentives to specific customer groups, and predict the impact of new customers or neighborhoods.

We propose a predictive energy usage model using novel machine learning methods to address these two issues, and provide accurate estimates of total energy usage and changing trends. The variable of interest here is the energy units consumed in a certain time period by entities at granularities such as customers, organizations, appliances, and buildings. Our model considers attributes like customer demographics, weather, time of day, and geographical information – identifying these attributes is in itself a challenge we tackle. In addition, the dynamic nature of these attributes and changes in customer behavior can pose additional difficulties to the model.

One way to handle consumer diversity and make the large-scale problem tractable is to cluster customers into relatively homogeneous groups and model each group separately. There are others approaches we will explore, motivated by prior work in domains like intelligent transportation management and marketing that can be related to the smart grid domain. While the former domain is concerned with the problem of predicting traffic patterns, the latter deals with predicting the purchase behavior of the customers.

The approaches will be tested against different datasets provided by the utility companies, and in the context of USC buildings and customers. The approaches will be compared on several aspects for different types of prediction problems. A good prediction model will help estimate the energy demands for different times of the year and facilitate the formulation of a demand-response policy.