Chapter 19
Semantic Information Integration for Smart Grid Applications*

Yogesh Simmhan, Qunzhi Zhou, and Viktor Prasanna

Abstract. The Los Angeles Smart Grid Project aims to use informatics techniques to bring about a quantum leap in the way demand response load optimization is performed in utilities. Semantic information integration, from sources as diverse as Internet-connected smart meters and social networks, is a linchpin to support the advanced analytics and mining algorithms required for this. In association with it, semantic complex event processing system will allow consumer and utility managers to easily specify and enact energy policies continuously. We present the information systems architecture for the project that is under development, and discuss research issues that emerge from having to design a system that supports 1.4 million customers and a rich ecosystem of Smart Grid applications from users, third party vendors, the utility and regulators.

19.1 Introduction

The Los Angeles Smart Grid Demonstration Project is a Department of Energy Sponsored project[1] to investigate Smart Grid technology and innovative use of information technology to improve power usage within the largest municipal utility in the United States as it deploys smart meters across its service area [1]. The project, a collaboration between the Los Angeles Department of Water and Power (DWP), University of Southern California (USC), University of California–Los Angeles (UCLA) and NASA's Jet Propulsion Lab (JPL), is conducting research on five themes: software architecture for demand response optimization (USC and

---

Yogesh Simmhan · Viktor Prasanna
Ming Hsieh Department of Electrical Engineering

Yogesh Simmhan · Qunzhi Zhou · Viktor Prasanna
Center for Energy Informatics

Qunzhi Zhou · Viktor Prasanna
Computer Science Department, Viterbi School of Engineering,
University of Southern California, Los Angeles CA

* This material is based upon work supported by the Department of Energy under Award Number DE-OE0000192.

UCLA), consumer behavior (USC), electric vehicles (USC and UCLA), and cyber-security (USC and JPL), with the eventual target of demonstrating these within the Los Angeles smart power grid in 2014 and helping identify best practices for other utilities.

Demand response optimization (DR) is one of the key goals of the project [2]. DR is the ability for utilities to manage the demand for power to within the current available power generation capacity – particularly during the middle of the day when electricity consumption is at its peak and approaches power generation limits. The difference between peak and base electricity load is often as high as 50%, causing excess capacity to be provisioned by the utility.

Traditionally, DR is performed at a coarse time granularity and statically (Figure 19.1, left). Monthly usage data that is physically collected by utility personnel from individual consumer electricity meters is combined with the total electricity usage available centrally to the utility to forecast monthly trends. These are used to statically set annual incentives like time of use (TOU) pricing for different time periods of the day and shared with the consumers in the hope that they change their daily activities in the subsequent year.

Smart meters or Advanced Metering Infrastructure (AMI) form the enabling technology for a Smart Grids, and allow realtime, bi-directional communication between the utility and the consumer's home or building area network (HAN and BAN) using Internet and other protocols. Smart meters allow the utility to measure realtime electricity usage for each customer, and communicate pricing or load reduction signals to the consumer's appliances or HAN/BAN, which can be acted upon automatically (Figure 19.1, right). This dramatically reduces the time taken for the DR feedback to be set in motion and can be activated for finer granularities of consumers.

Our ongoing research into *software architecture for demand response optimization* aims to bring about a quantum leap in the way DR is performed in smart power utilities [2]. Rather than considering the fine grained information and control available through smart meters in isolation, as was done previously, we approach DR from an informatics perspective: where information from diverse sources – AMIs being just one of them – are integrated together and analyzed to

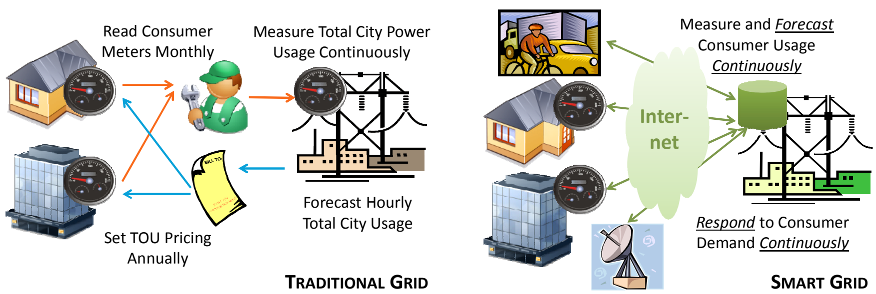

Fig. 19.1 Demand Response Optimization in Traditional and Smart Power Grids.

detect both direct predictors and indirect influencers of electricity usage. Load curtailment responses are similarly targeted both at electrical equipment and consumer behavior, and their effectiveness measured.

An effective *demand forecasting* and *load curtailment* software solution needs to meet three key characteristics for a large utility such as LA DWP, which serves 1.4 million consumers and consumes almost 1% of the total US power [1, 3, 4].

- **Intelligent, Adaptive DR:** Forecasting and response algorithms need to adapt to a dynamic city, where features that influence power usage constantly change, people move, traffic patterns change, and new electricity sources and sinks, like electric vehicles and solar panels, drastically change the energy landscape.
- **Translation of Policy to Practice:** A human in the loop of such a dynamic system with complex and multi-disciplinary attributes that impact energy use cannot keep up. Both customers and utility manager should be able to specify high level goals (e.g. "Keep load to within 90% of capacity", "Monthly electricity budget is $100") and delegate their enactment to software agents.
- **Securely Scale on Emerging Platforms:** The information analyzed by utilities is set for a dramatic increase, and will continue to grow on the order of tera- and even peta- bytes. The information architecture should deploy on scalable platforms such as Clouds. Ensuring data security and privacy will be of utmost importance as consumers get accustomed to an information-driven utility that can closely monitor their energy use.

Semantic information integration and *complex event processing* are two major components of our information architecture that address these requirements for effective DR.

Rapidly incorporate diverse information sources

As information attributes that directly and indirectly influence load forecasting and curtailment change over time, the information architecture must be able to incorporate new sources and deprecate old ones quickly. In addition, it needs to keep up with changing formats, qualities, and access policies of the sources. Semantics are needed to correlate diverse information that may refer to the same concept, and relate their impact on each other. Continuous arrival of data and online analysis by compute-intensive algorithms that mine for energy use patterns for power load forecasting means that information is constantly processed within the architecture and the warehouse contents is constantly changing.

Intuitive platform for goal-based policy specification

Information systems operate on queries. Utility managers and consumers operate on goals and policies. Means to specify higher level policies and automatically translate them into meaningful (semantic) queries needs to be present. The goals can be based on several constraints: energy conservation, social pressure, monetary benefits, ensuring quality of service, or scheduling maintenance. While the policies remain the same, their equivalent queries can change as information sources or forecast models change.

Besides the requirements of the utility for DR, other rich applications are enabled by an information-driven software architecture and will access the utility's information repository. These include consumer home automation software, third party vendors providing data analytics to consumers, mobile apps that notify users of peak demand signals and pricing incentives, data aggregators and regulators monitoring utility activities, and developers who wish to invent value added tools for customers.

In the ensuing sections, we present the design of the information systems architecture for the LA DWP smart grid project that is ongoing. In particular, we discuss the need for semantic information integration, and semantic complex event processing to support Smart Grid applications, and highlight the research issues they entail. This project is under active research, with initial prototyping underway to validate our architectural choices. In later sections, we present related work on these topics and present our future directions.

19.2 Semantic Information Integration

Smart Grid applications such as DR require a complete change in the energy information management paradigm. Traditional energy information systems tend to be vertically integrated, closed architectures that limit interoperability, extensibility and reuse. Customers are viewed as passive users of electricity (and information related to it), while utilities are charged with providing electricity as a commodity with fixed tariffs. As power utilities migrate to Smart Grids, information from heterogeneous sources including smart meters that report near real-time power usage and quality, intelligent thermostats that measure and control buildings' heat and humidity, and weather forecasts and traffic reports from online services are needed for accurate power *demand forecasting* and identification of effective *curtailment strategies* using data mining and analysis algorithms [2]. The information used by these DR algorithms is diverse in terms of the structures and semantics of data, software and hardware platforms used, types of interactions that they can support, and the size of the data items.

Figure 19.2 shows our information architecture that is being prototyped. Information sources at the bottom can communicate with the information integration framework through a service bus. Semi-structured data from micro-blogs and text comments are mined to extract structured data. These and other structured information arriving are semantically annotated using a Smart Grid ontology model using wrappers. The semantically meaningful instances of concept and relationship can then be recorded in a Cloud-hosted information repository for use by DR applications. Separately, the complex event processing (CEP) system extracts events from the semantically annotated information, constructs new abstract events based on rules, and reasons over the event cloud to match patterns based on policy and trigger actions for DR. A mining process additionally looks for new energy forecast and power curtailment pattern structures for future interest. Utility and external applications, shown on the left, access this semantic repository using the semantic model through the service bus. The features of the smart grid ontology model are described below, and the information repository schema conforms to it.

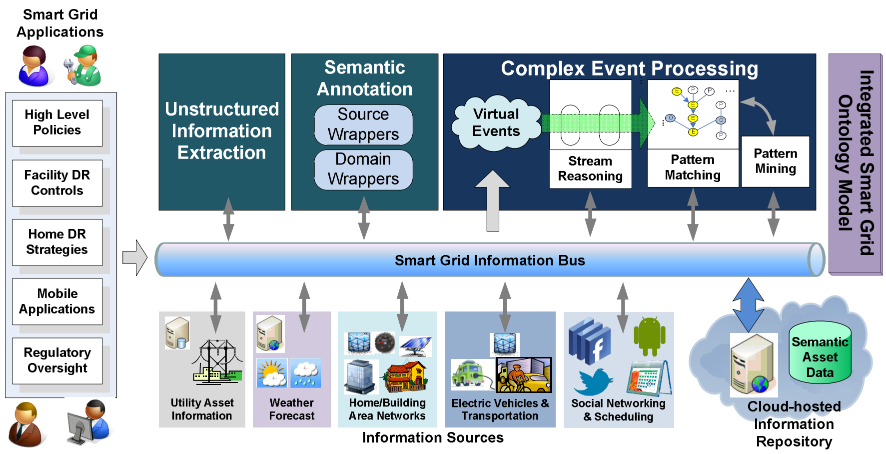

Fig. 19.2 Information Integration and CEP Achitecture for the Smart Grid.

The information integration architecture should support many-to-many points of interaction and information exchange [31]. The space of information sources and types is broad and will change as algorithms adapt. Utilities, software and hardware vendors are developing novel applications and products to meet consumer, utility and regulatory needs. While they are directly responsible for the functioning of their own software systems, they also depend on exchanging information with other systems. They need to negotiate agreements with the entities they interact with on how this information is to be exchanged. Conducting these negotiations between pairwise information sources and sinks is burdensome.

Semantic Web provides an ontology-based extensible framework for information to be shared and reused across application and domain boundaries [5]. It has been successful for information integration in domains such as eHealthCare [6], biology [7], and transportation [8]. Smart Grid data can be effectively used and interpreted when grounded in semantically meaningful terms. It is not just the utility managers who will use this information to determine policies, but also data mining and analysis applications, third party tools with which the data is shared, transmitted to other utilities and regulators, and even the consumer's application.

Applying Semantic Web for Smart Grid information integration goes beyond just a normalized form for exchanging meter, sensor or appliance data, and suggests integrating ontologies from different relevant domains – weather, traffic, and social networks, among others – and using sematic definitions for policies and software APIs using well understood semantic standards.

19.2.1 Integrated Smart Grid Ontology Model

We use the Semantic Web Ontology Language (OWL) to build the Integrated Smart Grid Ontology Model which stores the information about entities and their relationships in a structured format (Figure 19.2, right). We distinguish two types of ontologies: *domain ontologies* that are commonly used by utilities and power systems, and *external ontologies*, which describe concepts outside the electric grid such as weather and traffic, yet are required for information-driven energy use forecasting and curtailment by novel DR algorithms. These ontologies form the entities and relationships that for the schema for our information repository.

19.2.1.1 Domain and External Ontologies

Existing utility warehouses contribute concepts and relationships to the domain ontologies from three existing information sources: *Meter Data Management (MDM)*, *Customer Information System (CIS)* and *Outage Management System (OMS)*. CIS is used to describe the customers in the utility's service area and their address, billing records, and service requests. MDM records electricity usage information from electric meters, including their location, correlated with the customer who is being metered and for what time period. Earlier, this usage was populated by utility personnel who physically visit the customer location and record this data once a month. In a smart utility, this information is transmitted up to once a minute by the AMI head-end over a communication network, potentially using internet protocols, to a gateway at the utility that maps the AMI vendor specific data to a standard form. OMS is used to monitor the operation of the electricity distribution and transmission network. It helps analyze transmission reliability and narrow down sources of service outages to resolve them. Consequently, it captures the description of the entire utility's electric network itself [10].

In addition to the utility's information sources, other domain ontologies are contributed by the home and building area networks. These help capture the existence and activities of home appliance and electric vehicles (EVs), building plans such as surface area and year of construction, Heating, Ventilation and Air Conditioning (HVAC) equipment and their duty cycles, and so on [11].

Among external ontologies, we have identified four areas to initially support that affect or help forecast energy usage indirectly: *weather*, *traffic*, *scheduling* and *social networks*. Weather includes historical, current and forecast information about temperature, humidity, precipitation, cloud cover and wind speed. The first two affect the use of air-conditioning, precipitation additionally impacts traffic patterns, and cloud cover and wind speed affect power generation by solar and wind renewables. Traffic data shows the traffic flow and the rate of traffic exiting from arterial roads onto specific neighborhoods and predict impending energy use as customers arrive home. Additionally, movement of individual EVs is also modeled as they are a major energy sink when plugged in. Scheduling applies to both people (*when*, *where*, *with whom*) and facilities like classrooms, office rooms, and convention centers (*when*, *where*, *attendees* or *their count*). Social network data models people, their peers and friends, and their micro-blogs that is used for mining.

19.2.1.2 Ontology Normalization

Modeling the complete Smart Grid ontology is in itself an incremental process and a work in progress. It consists of identifying agreed upon standards for information modeling in that area, recognizing equivalent relationships between concepts in different domains, and defining additional relationships between concepts. Often, the ontologies do not exist in a directly usable form such as OWL. Sometimes, they do not exist as an information model as such, and need to be drawn out of communication protocols for message exchanges. This requires a keen understanding of the domain concepts by the modeler to establish their relationship.

However, this understanding does not have to extend to actually determining the exact influence that each concept has on the other with respect to energy use. That is something that will emerge from pattern specification and mining over observational data that will populate instances of the model concepts and relationships.

For example, the ontology model captures the concepts of and relationships between an office room in a building, its lighting system, its energy usage, and the infrared "people sensor" in the room, and that people in the office have a calendar schedule. These concepts and relationships are relatively easy to describe with limited knowledge of the different domains. This model captures the link between the person entering the room, the "person sensor" detecting it, triggering the lights to turn them on, and hence increasing the energy usage. This sequence is triggered at the time the person enters the room. What may be absent from the ontology model and discovered over time based on observational data for this model is: that the person's calendar has the room location and time of the meeting, that this person usually arrives early for most meetings, and that the fact that a meeting exists on her calendar means that the energy use for that room is going to increase at that point in future. This sequence is triggered at the time the calendar event is created – potentially hours or days in advance.

Information model standards for some of the domain ontologies exist and will be reused. The IEC Common Information Model (CIM) [12] provides an initial starting point to model power systems, including circuits and utility operations. It covers several aspects of MDM and OMS, but has to be extended with concepts for CIS. Also, the UML model of CIM needs to be mapped to OWL, which introduces issues that are documented [13]. ISO standards for building area network communications provide message protocols and APIs that can be used to derive concepts and relationships for an information model [14]. Weather concepts can be captured by JPL's SWEET ontology [15] and those for NextGen Network Enabled Weather (NNEW) [16]. Our prior work on ontologies for transportation networks is leveraged for the traffic models [17]. There is active work on defining standards for social network models, including a W3C incubator group [18][19]. Internet calendar standards provide the basic building blocks for scheduling [20].

19.2.2 Ingesting Information into Repository

Many information sources that contribute to our information repository do not provide semantic information; some do not even provide structured data. Two pre-processing steps of *information extraction* and *semantic annotation* need to take place before data from external sources is inserted into the repository in a semantically queryable form. These steps are exacerbated by information constantly arriving from different sources into our repository, thus distinguishing it from a traditional data-warehouse where the extract-transform-load (ETL) process is batched.

19.2.2.1 Information Extraction from Semi-structured Social Network Data

One of the primary sources of unstructured and semi-structured information useful for DR is from social networks. Information about friends on Facebook can be used for studying peer-pressure as a response mechanism for curtailing load (e.g. *"Your friends have signed up to reduce electricity usage by 10%. Would you like to?"*). Similarly, micro-blogging posts on Twitter and Facebook can be analyzed to detect user activity and indirectly forecast energy usage (e.g. *"Sick with #flu. Working from home tomorrow. #wfh"*). These are on the presumption that the utility has access and rights to this information, which is feasible through providing incentives for "friending" the utility and registering their twitter feed with it.

Two approaches for extracting structured information from unstructured social network data are being investigated. The first approach is to just use the structured content that exists in social network data and using statistics to draw *classifications* that can be fed into the information repository rather than the raw content. User IDs, peer network, hashtags and text tags, timestamps, twitter followers, and occasionally, location information are common structured information. It is possible to analyze the tags and twitter followers to classify users according to, say, their environmental consciousness or technology uptake based on specific issues they follow. This classification can help tune energy conservation messaging and technology tools for target users. Similarly, frequency, time of day and location of tweets/posts can be used to map energy use patterns of individuals.

Another approach for analyzing micro-blobs is to use Natural Language Processing (NLP) to evaluate user intent on energy use and curtailment. NLP techniques have been used for data extraction and curation from literature [21] and their use for examining Twitter feeds for emergency response is being actively researched [22][23]. Challenges to applying NLP to micro-blogging are introduced by the compact space available for content and the consequent use of unique abbreviations and sentence formulation. Too, posts may be related to conversation threads that span multiple mediums (IM, text, voice). Despite this, it may be possible to determine a small class of topic terms that affect energy use, such as a user's schedule or purchase of a smart appliance, to shape DR analysis.

19.2.2.2 Online Semantic Annotation

Structured information from external sources can have static mappings defined with related semantic ontology(ies). These mappings may be specific to an information source (e.g. *NOAA* for weather data) or an information type (e.g. *electricity usage* from Smart Meters). In our architecture, wrapper libraries that can access these individual mappings are responsible for processing information as it arrives from external sources, and enhancing them with a semantic concepts and relationships. It is these semantic instances that are actually recorded in the repository.

Individual agents may be responsible for specific data sources, which makes it easier to identify the right mapping to use. These agents can either pull information from the sources, or have it pushed to them. Part of the annotation process should include performing data transformations (such as unit conversion) to provide consistent information in the repository. Software vendors such as Oracle have specialized software solutions for utilities that help perform such conversions for smart meter data from different types of meters as part of a gateway service[2].

19.2.3 Cloud-Hosted Information Repository

The need to support accretive information ingest and its sharing among thousands, potentially millions, of users means that the repository needs to be hosted on a scalable platform [24]. Initial estimates put the upper bound on the size of annual meter data for 1.4 Million Los Angeles consumers at 1 minute intervals at hundreds of Terabytes per year. In addition, a host of metadata and semantic information about the consumers is gathered in support of DR. We adopt a service oriented architecture deployed on Clouds to host the information repository, DR information processing and analysis tools, and information sharing endpoints.

19.2.3.1 Information Integration on Clouds

Clouds provide scalable resources in the form of Virtual Machines (VMs) and reliable storage services like queues, tables and blobs. These elastic resources with a pay-as-you-go model provide the flexibility needed for information acquisition, processing, storage, query and dissemination for large repositories like the ones we are considering for Smart Grids [24]. For example, it is possible to scale out clients on multiple VMs to access different online service providers for information in parallel. A number of commercial Cloud vendors and open source Cloud software fabrics exist, including Amazon AWS, Microsoft Azure, and Eucalyptus. There are, however, some issues that need to be resolved for Clouds to be used effectively for the Smart Grid information architecture.

One, current Clouds do not provide the ability to host large databases as a platform service. Microsoft SQL Azure[3] is one of the major providers of SQL server as a Cloud platform, but limits the database size to under 100GB which is

[2] www.oracle.com/us/industries/utilities

[3] www.microsoft.com/en-us/sqlazure/database.aspx

insufficient for hosting meter data. We will either have to consider a SQL database installed on a Cloud VM as an infrastructure service – with manually configured reliability features, or distribute the data into multiple databases, partitioned either by consumers or the type of information. Alternatively, a combination of public and private Clouds is possible.

In addition, the semantic features required by our architecture are not available out-of-the-box for datasets of the scale we need to process. Lastly, acquiring streaming data from millions of Smart Meters is expected to be a challenge when considering the network resource constraints. Our recent work has highlighted the absence of streaming as a Cloud storage service as problem for Smart Grids [25], and suggested algorithms to throttle bandwidth using application QoS needs [26].

19.2.3.2 Data Security and Privacy

Cyber-security has been identified as a key concern for Smart Grid adoption and is being examined in detail [27][28]. In addition to securing the communication between smart meters and the utility, it is also important to secure the actual information hosted in the information repository [26]. For particularly sensitive data, this may even amount to encrypting the data when storing it in a public Cloud to ensure the Cloud service provider themselves cannot access it. Splitting data across public and private Clouds is another approach to ensure sensitive data is kept within the confines of the utility's private Cloud, but introduces challenges when querying data that span the silos.

Another issue that is less understood but of heightened concern is one of information leakage as a result of integrating such diverse data [26]. *Data privacy* issues have been in the limelight for social networks and can severely compromise trust in the utility if measures are not put in place to clearly define information sharing and use policies that are enforced. While users are increasingly willing to post personal information in public, there is backlash when such knowledge is used by organizations to price commodities or in unconventional ways. Tracking the provenance of data used for demand forecasting and load curtailment, and distinguishing it from pricing policies is also necessary.

19.3 Semantic Complex Event Processing

New types of demand response applications locate patterns among a large class of realtime information to detect power consumption behaviors and predict usage in realtime. The requirement of timely processing and response to power usage situations makes complex event processing (CEP) an attractive component of a DR information system architecture. CEP deals with computation, transformation and pattern detection over large volumes of partially ordered events and messages [29]. An event is essentially a data object that represents something that occurs or changes the current state of affairs. In Smart Grid, continuous data from realtime information sources can be abstracted as events. These may be from sensors and appliances (*ThermostatChange* event, *ToasterOn* event), smart meters (*PowerUsage* event), weather phenomena (*HeatWave* event) or consumer activity

(*ClassScheduleChange* event, *HomeArrive* event). CEP correlates distributed events in order to detect meaningful event patterns or situations, thus supplying Smart Grid applications with an accumulated view of the incoming data, and allowing them to react to detected patterns.

Some limitations of current CEP systems impede their effective use for Smart Grids. Existing systems process events as plain data tuples. As such, complex event patterns can only be defined as a combination of attributes presented in event data [30]. Users have to know the details of event structures and sources, and define patterns over low level specifications. Also, most CEP systems only support well-defined and precise pattern specification and matching, without any leeway to relax pattern constraints. However, uncertainty is an intrinsic feature of real world cyber-physical applications, where potentially incomplete, imprecise and even incorrect data exist, but still need to be matched within certain bounds.

An effective CEP solution for DR needs to meet several requirements. First, it must be extensible to meet the organic growth of the Smart Grid information diversity with the provision to easily model, specify and identify new events, relationships, and event patterns. Second, it should be easy to use both by domain experts and by non-domain users like developers and consumers with limited power systems background. Third, it should support uncertainties in events and patterns. Naturally, the CEP system should also scale to large sizes and event rates.

Here, we describe our initial design of Semantic Complex Event Processing for DR that supports the above requirements [31].

19.3.1 Semantic Event Modeling

The state-of-the-art CEP systems process events as plain data tuples with structure. For example, power demand events may be modeled as *demand(appliance id, value, timestamp)*. As such, complex event patterns can only be defined as a combination of attributes presented in the event model. Data level specifications of event patterns are too complex for domain experts in an information rich space like Smart Grid.

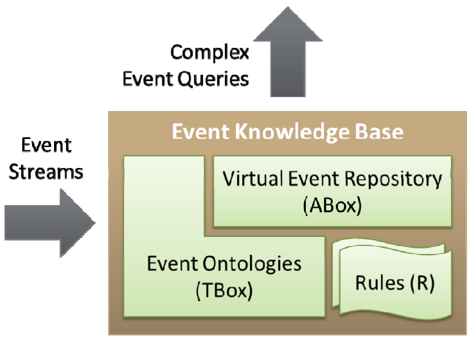

Fig. 19.3 Conceptual event model for CEP applications [1].

We propose an event model that captures the semantics of events, forming the foundation for expressive, high level complex event patterns [31]. By incorporating semantics, event data becomes meaningful information that enables automated mediation between domains and abstraction levels, and mapping goals to patterns.

We adopt a description logic based knowledge representation to model events (Figure 19.3). The event model captures the event types, type hierarchies (i.e., specialization and generalization), relations between events and other domain entities. Using description logic, a semantic event knowledge base K can be represented for the event model as:

$$\mathrm{K} = (\mathrm{T}, \mathrm{R}, \mathrm{A})$$

where T is the *TBox* which introduces the vocabulary (or *T*erminology) of events and domain concepts; A is the *ABox*, a virtual event repository that "stores" *A*ssertions about named individual events using the vocabulary; R is a *R*ule set that represents the correlations between events, and is used to derive virtual events into the event repository based on observed events [32].

The event model maps to an OWL schema, and instances about events and domain entities can be stored in a structured OWL representation. Event ontologies need to be organized in a modular and layered manner for easy extension. The top level event ontology captures concepts and relationships between events, such as the time the event occurs and which domain specific classes it corresponds to (e.g. *ThermostatChange*, *AirConditionerTurnedOn*). The second layer of the model contains the subjective domain ontologies introduced earlier in the information integration section. The lowest level of the event model relates to external ontologies, also introduced before. Connections between the event ontology and the domain ontologies are made using properties like "eventHappensTo" and "eventHappensAt", whose domain is an event and value is a domain object. For example, the "thermostat" has a "thermostatChange" happen to is a concept defined in the *appliance* domain ontology.

Semantic Web Rule Language (SWRL) can be applied to encode correlation rules between events (R in the event knowledge base). SWRL is a proposed rule language for Semantic Web grounded in horn logic. A SWRL rule consists of a condition (rule body) and a consequence (rule head), each of which consists of a set of atoms. Event correlation rules can be defined based on domain knowledge or based on analysis of historic data. Several simple correlation rules are showed below, which derive new demand response virtual events based on known events.

r1: (? e1 rdf: type event: classStart)(? e1 event: eventHappensAt ? time)
(? e1 event: eventHappensTo ? room) (? e2 rdf: type event: loadIncrease)
(? e2 event: eventHappensAt ? time)(? e2 event: eventHappensTo ? room)

r2: (? e1 rdf: type event: powerDistortion)(? e1 event: eventHappensAt ? time)
(? e1 event: eventHappensTo ? building)(? e2 rdf: type event: heavyDemand)
(? e2 event: eventHappensAt ? time)(? e2 event: eventHappensTo ? building)

19.3.2 Reasoning over Realtime Events

In our proposed CEP system, time critical situations are detected not only based on raw event data, but also upon additional knowledge such as semantics and correlations of events [31]. The performance of the reasoning technique is time sensitive. Current research into reasoning systems have improved performance over a static or slowly growing knowledge base, but reasoning over realtime events backed by ontologies and rules is a novel challenge.

A naive implementation of event reasoning would store the event history in a repository and perform inference over the event base as new events are observed. However, events may arrive at high frequency and are unbounded. Thus, optimizations are needed to improve its performance in terms of storage cost, latency and throughput.

Reasoning over realtime events can be parallelized to improve performance. Our prior work on parallel inferencing for OWL knowledge base using graph partitioning is applicable here [33]. In event reasoning, partitioning can be applied to both the historic event data and correlation rule set. Based on the relations to each other (edges in the semantic graph), events and rules can be grouped into partitions which are dispatched to different VMs for parallel processing. VMs communicate intermediate results for necessary synchronization. The goal of the partitioning algorithms is to balance load between different partitions by minimizing the edge-cut communications.

Another potential optimization includes using filters to drop duplicate or irrelevant events from high frequency inputs before they are processed. For realtime applications, we can restrict reasoning to a certain window which consists of a subset of recently observed events while previous information is ignored. This is meaningful for two reasons. First, ignoring older information allows us to save computing resources in terms of memory, storage and processing time, and response to important events in real time. Further, most event processing applications assume that older events eventually become irrelevant. Applying appropriate time windows will help significantly reduce the size of rules and data to consider in the inference process.

19.3.3 Semantic Mining for Energy Use Patterns

Meaningful energy use patterns for complex event processing can be defined either based on domain knowledge or from an analysis of historic data [31]. For example, event patterns that represent power consumption trends can be simply expressed by using a sequence of monotonically decreasing/increasing AMI meter readings. However, identifying event patterns from a large corpus of energy data is non-trivial. Moreover, energy use patterns may also constantly evolve over time. For example, the event patterns that predict the demand of a building can vary due to changes of owners, renovation, and so on. There is a need to enhance CEP with pattern mining to assist with automatic specification of novel patterns.

Existing data mining techniques can be applied to extract static energy use patterns. The most relevant approaches are association rule learning and sequential data pattern mining. Algorithms for mining association rules from relational data have been well studied [34, 35]. The central idea is to find all rule patterns whose confidence and support are above corresponding thresholds. Some studies [36, 37] contribute to the mining of sequential patterns in temporal datasets. Most of these methods are based on searching algorithms using a heuristic which states the fact that any super-pattern of an infrequent pattern cannot be frequent.

Typically, data mining is done off-line. However there are examples in which event processing and mining run concurrently. In these cases, we can use a target event to constantly monitor the change of event patterns, for example, using meter readings to monitor patterns that predict peak power demand. When the event pattern deviates significantly from the target, it triggers the pattern mining process. For Smart Grid applications, mining dynamic energy use patterns from a complete set of historic data is not necessary as older events become less relevant when the energy use behaviors evolve. Data mining algorithms can be applied to the event data in a fixed time window to identify events frequently occurring with the target events.

19.4 Related Work

19.4.1 Semantic Information Integration

Semantic information integration is an active research area that has been studied in various research and application domains, including databases [38] [39], web services [40] [41], eHealthCare [6] [42], oil industry [43] and transportation [8]. The general objective is to facilitate interoperability of information systems and share information sources that are often heterogeneous and distributed [31].

Contemporary approaches for semantic interoperability can be classified into two categories. The first is the so-called Brute-force Data Conversions (BF) [44], which directly implements all necessary data transformations manually. This approach may require a large number of transformation agreements to be hardcoded that are difficult to maintain [31]. Another approach is Global Standardization, in which different information systems agree on a uniform standard. This causes the semantic differences to disappear and there is no need for data transforms between components. Unfortunately, such standardization is usually infeasible for many domains, for example, because of organizational and operational reasons.

Semantic Web provides an extensible framework that allows information to be shared and reused across application and domain boundaries using ontologies. The shortcomings of the traditional approaches can be overcome by declaratively describing data semantics using ontologies and separating knowledge representation from data transform implementation [44]. A widely adopted approach is to allow information sources to describe their vocabularies of information independently using ontologies. Inter-ontology mappings and reasoning services are then applied to identify semantically corresponding terms of component ontologies, e.g., which terms are semantically equal or similar.

Numerous research projects utilize ontologies to represent data semantics and facilitate information integration. For example, [42] describes the use of Semantic Web technologies for sharing knowledge in healthcare. It combines relational databases and ontologies to extract knowledge that was not explicitly declared within the database. An ontology representation of the UMLS (Unified Medical Language System) represents the basic medical concepts, and mappings and inference over the semantic knowledge are done to query and update heterogeneous databases. [8] developed a unifying traffic modeling and simulation framework into which more focused simulators can be integrated. A domain ontology was used as the common modeling language and data exchange model for integrated simulation.

Information processing has started to get more sophisticated at utilities in recent times. There have been efforts to display grid-level information integration through a consumer portal or a home energy monitoring system. These commonly take the form of a simple chart or histogram of energy usage over various time periods, such as Google's PowerMeter [45], the Pulse energy management software [46], and AgileWaves [47]. While this approach is promising, they do not (yet) support the ability to incorporate external applications.

Efforts have been taken to leverage standards [48, 49] for Smart Grid participants to integrate components, and span energy information from the “micro” (i.e. the power domain) to the “macro” (i.e. multiple relevant domains and users). The existing Smart Grid standards span multiple domains from electric power generation, electrical appliances to information technology, and are designed by a number of organizations: IEC, EPRI, NIST, W3C and others. In the last decade, the power industry has made efforts in creating a common information model (CIM) to resolve semantic inconsistency between the different standards. IEC 61970 and IEC 61968 series standards [50] define data exchange specifications of CIM so that the interoperability between various applications can be achieved.

There has been recent work on developing scalable, semantic-level Smart Grid information integration framework. In [51], the authors propose a shared ontology model to provide common semantics for Smart Grid applications. The ontology captures concepts governed by business semantics, engineering and scientific principles by transforming existing standards, such as CIM, to a uniform conceptual model. To make semantic modeling accessible to domain experts, they also developed the Semantic Application Design Language (SADL), a controlled-English language with an associated environment for building semantic models.

19.4.2 Complex Event Processing

Complex event processing deals with detecting real-time situations, represented as event patterns, from an event loud. CEP originated from the RAPIDE simulation research project in 1993 [52][31]. Recently, CEP has received attention in the research community due to its applicability to domains such as financial services

[53, 54], health care [55] and RFID data management [56]. Research prototypes and commercial systems, such as ruleCore [57], Oracle CEP Server [58] and Esper [59], are available.

Snoop [60] is a proposed event specification language for active databases. In active database system, updates to the database are treated as events. The authors observe that the detection of database event patterns leads to monotonically increasing storage overhead as previous occurrences of events cannot be deleted. To overcome this problem, they introduce the event selection and consumption policies, i.e. so-called parameter contexts for precisely restricting event occurrences. The notion of event selection and consumption are also important for Smart Grid event processing systems. For example, redundancy will often be present in the high frequency event data from meters and appliances. The effect of duplicate events from the same meter or sensor can be subsumed by the most recent event.

Cayuga [61] is a high performance CEP system designed to efficiently support a large number of concurrent complex event subscriptions. In Cayuga, event streams are infinite sequences of relational tuples with interval-based timestamps. It defines an expressive event algebra that contains six operators: selection, projection, renaming, union, conditional sequence and iteration. Event algebra expressions are detected by non-deterministic finite automata (NFA). We believe the event algebra generalized in Cayuga is necessary but not sufficient to support Smart Grid applications. It can be used to specify rigid accurate patterns but is unable to express event patterns that incorporate semantics and uncertainties of data. Recently, parallelized and distributed complex event processing has received attention in literature [64, 65] and are of relevance to our architecture.

19.5 Conclusion

The information system architecture we have described for demand response optimization in the Los Angeles Smart Grid project is intended to transform the way utilities and consumers treat energy management. By proposing an information rich environment with semantically meaningful data integrated from diverse sources, our architecture enables advanced analytical tools and algorithms to effectively and efficiently forecast energy load and identify load curtailment response. It also opens the door for other rich Smart Grid applications to be developed for desktop and mobile platforms that access the Cloud-hosted information repository.

The project is currently in the first year of a 3 year research and development cycle. We are in the process of prototyping several components of this architecture and evaluating within the USC campus micro-grid testbed, and subsequently scaling it to the LA DWP service area [2]. This software architecture will evolve during this period, informed by further research and evaluation. However, it establishes the advent of an informatics approach to Smart Grids that will lead to automated, intelligent and sustainable energy management.

Acknowledgement

This material is based upon work supported by the Department of Energy under Award Number DE-OE0000192and the Los Angeles Department of Water and Power. The authors would like to thank Carol Fern from the USC Facilities and Management Services, Julie Albright from the USC Annenberg School of Communication, Don Paul from the USC Energy Institute, and Aditya Sharma and Khaled Salem from the LA DWP for discussions and contributions.

Disclaimer: This report was prepared as an account of work sponsored by an agency of the United States Government. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

[1] Electric power industry overview (2007), http://www.eia.doe.gov/electricity/page/prim2/toc2.html

[2] Simmhan, Y., Aman, S., Cao, B., Giakkoupis, M., Kumbhare, A., Zhou, Q., Paul, D., Fern, C., Sharma, A., Prasanna, V.: An Informatics Approach to Demand Response Optimization in Smart Grids, Technical Report, University of Southern California (2011)

[3] LD WPD 2010 integrated resources plan, Sec N.5.Resource Adequacy Gap Analysis (2010)

[4] Existing capacity by energy source (2009), http://www.eia.doe.gov/cneaf/electricity/epa/epat1p2.html

[5] W3C Semantic Web Activity, http://www.w3.org/2001/sw

[6] Niekerk, J.C., Griffiths, K.: Advancing Health Care Management with the Semantic Web. In: Proc. of IEEE BroadCom 2008 (2008)

[7] Taswell, C.: Doors to the semantic web and grid with a portal for biomedical computing. IEEE Transactions on Information Technology in Biomedi-cine 12(2) (2008)

[8] Zhou, Q., Bakshi, A., Soma, R., Prasanna, V.K.: Towards an Integrated Model-ing and Simulation Framework for Freight Transportation in Metropolitan Areas. In: Proc. of IEEE International Conference on Information Reuse and Integration (2008)

[9] Semantic Web Ontology Language (OWL), http://www.w3.org/TR/owl-ref

[10] Finamore, E.P.: Integrated Outage Management: Leveraging Utility System Assets Including GIS and AMR for Optimum Outage Response. Electric Energy Magazine (2004)

[11] Aman, S., Simmhan, Y., Prasanna, V.K.: Towards Modeling and Prediction of Energy Consumption for a Campus Micro-grid. Under Submission (2011)

[12] McMorran, A. W.: An Introduction to IEC 61970-301 & 61968-11:The Common Information Model, Technical Report, University of Strathclyde (2007)

[13] Crapo, A., et al.: Overcoming Challenges Using the CIM as a Semantic Model for Energy Applications. Grid-Interop Forum, GridWise Architecture Council (2010)

[14] ANSI/ASHRAE 135-2008/ISO 16484-5 BACnet – A Data Communication Protocol for Building Automation and Control Networks

[15] Semantic Web for Earth and EnvironmentalTechnology (SWEET 2.1), NASA Jet Propulsion Lab (2010), `http://sweet.jpl.nasa.gov`

[16] NextGen Network Enabled Weather (NNEW): Data Models and Formats, University Corporation for Atmospheric Research, UCAR (2011), `http://wiki.ucar.edu/display/NNEWD/The+NNEW+Wiki`

[17] Zhou, Q., Bakshi, A., Prasanna, V.K., Soma, R.: Towards an Integrated Modeling and Simulation Framework for Freight Transportation in metropolitan Areas. In: Proc. of IEEE International Conference on Information Reuse and Integration, Las Vegas (2008)

[18] Halpin, H., Tuffield, M.: A Standards-based, Open and Privacy-aware Social Web. W3C Incubator Group Report (December 2010)

[19] Mika, P.: Ontologies are us: A unified model of social networks and semantics. Journal of Web Semantics 5(1) (2007)

[20] Internet Calendaring and Scheduling Core Object Specification (iCalendar)

[21] Alex, B., Grover, C., Haddow, B., Kabadjov, M., Klein, E., Matthews, M., Tobin, R., Wang, X.: Automating Curation Using a Natural Language Process-ing Pipeline. Genome Biology 9(Suppl 2):S1 (2008)

[22] William, J.C., Vieweg, S.: Travis Rood and Martha Palmer, Twitter in mass emergency: what NLP techniques can contribute. In: Workshop on Computational Linguistics in a World of Social Media, WSA (2010)

[23] Kireyev, K., Palen, L., Anderson, K.: Applications of Topics Models to Analysis of Disaster-Related Twitter Data. In: NIPS Workshop on Applications for Topic Models: Text and Beyond (2009)

[24] Simmhan, Y., Giakkoupis, M., Cao, B., Prasanna, V.K.: On Using Cloud Plat-forms in a Software Architecture for Smart Energy Grids. In: IEEE InternationalConference on Cloud Computing, CloudCom (2010)

[25] Zinn, D., Hart, Q., McPhillips, T., Ludascher, B., Simmhan, Y., Giakkoupis, M., Prasanna, V.K.: Towards Reliable, Performant Workflows for Streaming-Applications on Cloud Platforms. In: IEEE/ACM International Symposium on Cluster Computing and the Grid, CCGrid (May 2011)

[26] Simmhan, Y., Kumbhare, A., Cao, B., Prasanna, V.K.: An Analysis of Security and Privacy Issues in Smart Grid Software Architectures on Clouds. In: Workshop on Scientific Cloud Computing, ScienceCloud (2011)

[27] Burr, M.T.: Smart-grid security: Intelligent power grids present vexing cyber security problems. Public Utilities Fortnightly, 43 (January 2008)

[28] Echols, M., Sorebo, G.: Protecting Your Smart Grid. Transmission & Distribu-tion World, Penton Media (July 2010)

[29] Luckham, D., Schulte, R. (eds.): Event processing glossary- version 1.1.,Technical Report, Event Processing Technical Society (July 2008)

[30] Dong, L., Carlos, P., John, D.: Semantic enabled complex event language for business process monitoring. In: International Workshop on Semantic Business Process Management (2009)

[31] Zhou, Q., Simmhan, Y., Prasanna, V.K.: Towards an Inexact Semantic Com-plex Event Processing Framework for Demand Response. Under Submission (2011)

[32] Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.: The Description LogicHandbook: Theory, Implementation and Applications. Cambridge University Press, Cambridge (2002)

[33] Soma, R., Prasanna, V.K.: Parallel Inferencing for OWLKnowledge Bases. In: International Conference on Parallel Processing (2008)

[34] Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: International Conference on Very Large Data Bases, VLDB (1994)

[35] Zaki, M.: Scalable Algorithms for Association Mining. IEEETransactions on Knowledge and Data Engineering 12(3) (2000)

[36] Agrawal, R., Srikant, R.: Mining Sequential Patterns. In: International Confer-ence on DataEngineering, ICDE (1995)

[37] Yang, J., Wang, W., Yu, P.S.: Mining Asynchronous Periodic Patterns in Time Series Data. IEEE Transactions on Knowledge and Data Engineering 15(3) (2003)

[38] Kim, W., Seo, J.: Classifying schematic and data heterogeneity in multidatabasesystems. IEEEComputer 24(12), 12–18 (1991)

[39] Pottingerand, R., Bernstein, P.: Merging models based on given correspon-dences. In: International Conference on Very Large Databases, VLDB (2003)

[40] Lastra, J.L.M., Delamer, I.M.: Semantic web services in factory automation: fundamental insights and research roadmap. IEEE Transactions on Industrial Informatics 2(1) (2006)

[41] Acuna, C.J., Marcos, E., Gomez, J.M., Bussler, C.: Toward Web portals inte-gration through semantic Web services. In: International Conference on Next Generation Web Services Practices (2005)

[42] Nardon, F.B., Moura, L.A.: Knowledge sharing and information integration in healthcare using ontologies and deductive databases. Medinfo, 62–66 (2004)

[43] Soma, R., Bakshi, A., Prasanna, V.K.: A Semantic Framework for Integrated Asset Management in Smart Oilfields. In: IEEE InternationalConference on Cluster Computing and the Grid (2007)

[44] Gannon, T., Madnick, S., Moulton, A., Siegel, M., Sabbouh, M., Zhu, H.: Se-mantic Information Integration in the Large: Adaptability, Extensibility, and Scalability of the Context Mediation Approach, Technical Report, CISL# 2005-04. MIT, Cambridge, MA (2005)

[45] Google PowerMeter, `http://www.google.com/powermeter`

[46] Pulse energy management software, `http://www.pulseenergy.com`

[47] AgileWaves, `http://www.agilewaves.com`

[48] NIST Framework and Roadmap for Smart Grid Interoperability Standards, `http://www.nist.gov/smartgrid`

[49] Becker, D., Falk, H., Gillerman, J., Mauser, S., Podmore, R., Schneberger, L.: Standards-based approach integrates utility applications. IEEE Computer Ap-plications in Power 13 (2000)

[50] IECStandards, `http://www.iec.ch`

[51] Crapo, A., Wang, X., Lizzi, J., Larson, R.: The semantically enabled Smart Grid. In: The semantically enabled Smart Grid. GridWise Forum (2009)

[52] Luckham, D.C., Frasca, B.: Complex Event Processing in Distributed Systems, Technical Report (1998)

[53] Adi, A., Botzer, D., Nechushtai, G., Sharon, G.: Complex Event Processing for Financial Services. In: IEEE Services Computing Workshops (2006)

[54] Magid, Y., Adi, A., Barnea, M., Botzer, D., Rabinovich, E.: Application gen-eration framework for real-time complex event processing. In: IEEE International Computer Software and Applications Conference (2008)

[55] Churcher, G.E., Foley, J.: Applying and extending sensor web enablement to a telecare sensor network architecture. In: International ICST Conference on COMmunication System softWAre and middlewaRE (2009)

[56] Wu, E., Diao, Y., Rizvi, S.: High-performance complex event processing over streams. In: ACM SIGMOD international conference on Management of data (2006)
[57] RuleCore, `http://www.rulecore.com`
[58] Oracle, CEP Engine, `http://www.oracle.com/technetwork/middleware/complex-event-processing/overview/index.html`
[59] Esper, `http://esper.codehaus.org`
[60] Chakravarthy, S., Mishra, D.: Snoop: an expressive event specification lan-guage for active databases. Data & Knowledge Engineering 14, 1–26 (1994)
[61] Demers, A.J., Gehrke, J., Panda, B., Riedewald, M., Sharma, V., White, W.M.: Cayuga: A general purpose event monitoring system. In: Conference on Inno-vative Data Systems Research (CIDR), pp. 412–422 (2007)
[62] Gyllstrom, D., Wu, E., Chae, H.-J., Diao, Y., Stahlberg, P., Anderson, G.: SASE: Complex event processing over streams. In: Conference on Innovative Data Systems Research, CIDR (2007)
[63] Cherniack, M., Balakrishnan, H., Balazinska, M., Carney, D., Cetintemel, U., Xing, Y., Zdonik, S.B.: Scalable distributed stream processing. In: Conference on Innovative Data Systems Research, CIDR (2003)
[64] Brenna, L., Gehrke0, J., Hong, M., Johansen, D.: Distributed event stream processing with non-deterministic finite automata. In: ACM International Conference on Distributed Event-Based Systems, DEBS (2009)
[65] Akdere, M., Cetintemel, U., Tatbul, N.: Plan-based complex event detection across distributed sources. In: VLDB Endowment (2008)