EDL: an energy-aware semantic model for large-scale infrastructures

Zhu, H.; van der Veldt, K.; Grosso, P.; de Laat, C.

Universiteit van Amsterdam

System and Network Engineering

EDL: an energy-aware semantic model for large-scale infrastructures

Hao Zhu, Karel van der Veldt, Paola Grosso, Cees de Laat

June 6, 2015

Abstract

This report presents a semantic information model for large-scale computing infrastructures with energy awareness. The Energy Description Language (EDL) ontology reuses Infrastructure and Network Description Language (INDL), developed by our group, to describe the resources and network infrastructure that connects these resources. The EDL ontology itself focuses on conceptualizations and their representative terms for knowledge representation in the domain of energy monitoring. The concepts in EDL can support a varying range of power management scenarios such as reducing cost and lowering green house gas emission. Besides, EDL is flexible and low-cost for developing applications in large-scale infrastructure environment because of loose-coupling structure. Finally, the EDL ontology is OWL-based model, which facilitates ontology reuse and share by other developers. EDL has been adopted in our Energy Knowledge Base system.

SNE technical report SNE-UVA-2014-2

http://www.science.uva.nl/research/sne/reports/

1 Introduction

The state of art in power management researches lacks a deep understanding of the complicated infrastructure components and their state, their correlation and inter-dependency. For instance, which state and metrics resources have, which are useful for the power management, and how the state can be measured. This obscurity hinders the progress of this research field. The introduction of an ontology – describing the abstract concepts and the relationships in energy monitoring and energy measurement domain can help the research community achieve a better understanding of the available and useful knowledge for the power management.

Meanwhile, designing an effective, comprehensive ontology is a crucial step for knowledge interoperability and sharing in the context of power management across administrative infrastructure domains. Infrastructure operators tend to cooperate with each other to perform global power management. For example, National Research and Education Networks (NRENs) of Europe cooperate to look for a green routing path across different networks to transfer data in Mantychore Project. In this case, any one operator requires the infrastructure information and energy information from other domains. The ontology is the structural framework for organizing the information. Without an explicit ontology, it is hard to understand each other's request and exchange the information between domains considering that the information may be presented in different terms and formats. When the operators develop applications, like management systems or databases, they also need the ontologies to know how to properly manipulate data.

Therefore, we propose a semantic model that captures the intrinsic conceptual structure of infrastructures with energy concern. The EDL (Energy Description Language) ontology reuses Infrastructure and Network Language (INDL) [1] to describe resources and network infrastructure that connects these resources. INDL was modeled by our group. EDL itself focuses on the concepts and relationships of energy-aware monitoring and measurement, which are used for power management.

The structure of this report is as follows: First, we motivate our modeling approach using semantic web technology and introduce INDL in Sec. 2. Then we present the EDL model and its features in Sec. 3. Next we describe related work on energy monitoring. Finally we discuss the support of EDL for typical power management scenarios and conclude our paper in Sec. 5.

2 Semantic approach and INDL

Semantic web technologies provide a common framework that allows data to be shared and reused across applications. In the Semantic Web, any data resources can be described by a set of RDF triples, each one in a format:

{subject, predicate, object}. *Objects* and *subjects* represent resources to be described, and can be a data value or another resource. A *predicate* is a type of property relevant to this resource indicates the relationship between *subject* and *object* in a triple. RDF is a data model that is a standard to implement ontologies. It provides a basic vocabulary to construct triples. For example, a predefined property 'rdf:type' is used to specify the certain types of resource.

Ontologies are the information models in Semantic Web, which represent the knowledge in one domain as a set of concepts, relationship of these concepts and their property and type. RDF Schema (RDFS) and OWL are standards of adding semantics to describe ontologies using RDF. RDFS and OWL are designed for knowledge, and not data, means it is particularly concerned with semantics. RDFS provides mechanisms for describing groups of related resources and the relationships between these resources. For instance, it introduces the domains and ranges of properties. OWL further extends the RDFS vocabulary to help process the content of information such as inferring facts.

RDF, OWL and RDFS all use Uniform Resource Identifiers (URIs), which are universal global unique identifiers, to represent an element in a triple. The identifiers can link the concepts from different information sources. In addition, OWL features the separation between semantics and syntax. Ontologies implemented by different syntax can be compatibly combined.

XML schema is also a common way of representing informations model, but OWL ontologies are superior to XML schemas. Besides the support of global information identifiers, concepts defined by OWL ontology support the extension without loss of meaning. The RDF triples that use terms from other ontologies can even be added. However, adding new attributes to XML document might lead to ambiguity due that tree data model of XML is too flexible. The concepts defined by OWL ontologies has semantics in the context, and ontologies allow applications to infer new information from current information.

INDL and EDL are both OWL based ontology. So EDL can easily import INDL without name conflict, and other ontology designers can also easily use EDL. INDL imports Network Markup Language (NML) [2], which is a generic network description schema. The basic elements in NML are *Node*, *Port*, *Link* and *Topology*. *Node* is a device in the network, which can be hardware resources – router, switch or computer machine. *Node* connect the network throught its *Port*. The *Link* is connection between two *Ports*. *Port*, *Link* and *Node* consists of network topology. The most important feature of NML is that the Port can be logical concept, which is not corresponding to one physical interface. One physical interface can have two unidirectional NML port individuals to describe the traffic with different directions.

As Fig. 1 shows, INDL itself captures the concept of virtualization in computing infrastructures and describe *resource components*. INDL introduces the concept of virtual node to represent VM on one node. INDL concerns three types of resource components: *Memory Component*, *Processing Component* and *Storage Component*. It can depict the capability of data transmission and computing. An important difference from previous models is that INDL has loose coupling of virtualization, functionality and connectivity.

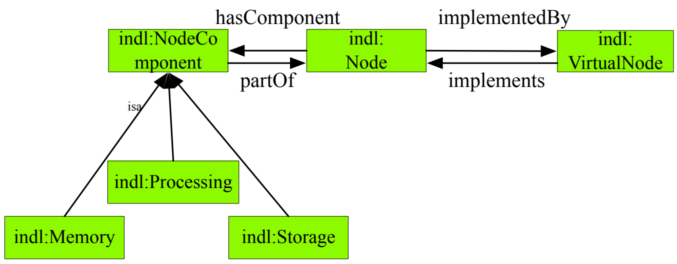

Figure 1: Infrastructure and Network Description Language

3 Energy Description Language - EDL

The goal of EDL[1] is to represent the energy monitoring objects and the energy-related state of resources. EDL links them to the *Node* class in INDL. The EDL model is shown in Fig. 2; it contains three main parts as following:

1. The *Green Metric* part defines the classes and properties to describe measurement data in different energy metrics.

2. The *Monitor Component* part describes the way of obtaining the measurement data of resources from sensors and the way of organizing measurement data in logs.

3. The *Characteristic* part is related to non-measurable state of computing or networking resources, e.g. energy source and energy efficient capabilities, which support the energy-aware resource discovery.

We describe these parts respectively in the following.

3.1 Green Metric

The *Metric* class contains the *Performance Metric* class and the *Green Metric* class. *Performance Metric* can be the general performance metrics such

[1]The EDL ontology file is available at: https://bitbucket.org/hzhu/edl

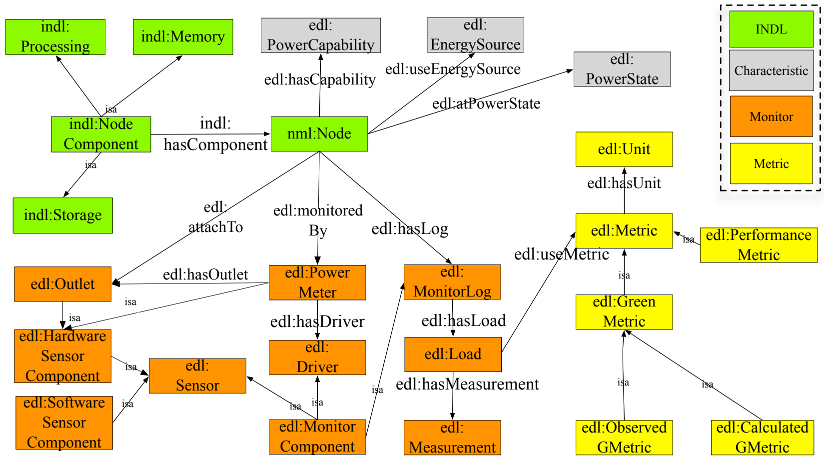

Figure 2: Energy Description Language imports INDL – INDL and three main parts in EDL. They are the Green Metric *classes and their properties, the* Monitor Component *classes and their properties and the energy characteristics classes and properties*

as throughput and utilization. Performance metrics qualify the capability of resources. There are some work on modeling performance metrics [3]. In this report, we focus on energy-aware metrics.

The *Green Metric* class represents the measurable state of resources in various metrics related to energy and sustainability. Energy management systems measure the current state in the green metrics to decide how to schedule tasks. The metrics they concern could be diverse. Some system may care GHG emission, while others may concern metrics about electricity cost of an infrastructure.

The Green Metric class defines two types of energy metrics, *Observed GMetric* and *Calculated GMetric*, distinguished by the way measurement data is obtained. Data in the former metrics is directly collected from sensors such as Power Distribution Units (PDUs), while data in the latter is usually obtained by numerical calculations of measurement data in observed green metrics and performance metrics.

In EDL, we predefined some common metrics classes. Besides these metrics, ontology developers can easily add more green metric individuals under green metric class. *Power Factor* is the ratio of real power to apparent power consumed. *Energy Consumption* and *Power Consumption* represent the total energy consumption in a period of time and real-time power consumption, respectively. They are the observed green metrics while *Energy Efficiency* and *Emission Efficiency* belong to *Calculated GMetric*.

Energy Efficiency is a measure of the rate of computation or transmission that can be processed by a computer for every watt of power consumed.

The Green500 List ranks supercomputers in the Top500 list using FLOPS per watt. Also, Energy Efficiency measures the number of operations or the bytes of data transmission for every joule of energy consumed. Comparably, Emission Efficiency measures the number of operations or the bytes of data transmission for every ton of GHG emission. *Calculated GMetric* also includes the metrics of the overall infrastructure like PUE. We also define the total amount of GHG emission and total amount of electricity costs.

Although the efficiency metrics seem more useful, absolute metrics e.g. Power Consumption is essential. Energy Efficiency can be improved by enhancing the performance even if resources continue to consume large amounts of absolute power. Resources in the idle state can not be adequately characterised by just efficiency but can be measured by Energy Consumption and Power Consumption.

Each metric instance is associated with a *Unit* instance according to its physical quantity. In many cases a numerical value alone cannot be understood without its unit type.

3.2 Energy-aware Characteristic Component

The *Energy Source* class defines the type of energy source used by the resources in infrastructures, *e.g.* wind, solar or thermal. Each energy source has a corresponding electricity price and GHG emissions rate; these can be used to calculate the total GHG emission of the resources in a period of time. With the description of energy sources, EDL has an awareness of environmental sustainability of resources.

The running state of a resource is determined by the *Power State* class. A management system should have the knowledge on whether the resource is in *Off*, *Sleep* or *Active* state. The *ratePower* property describes the thermal design power. The *Power Capability* class indicates which low power capability the resource has. Resources that are made up of embedded processors like Atom, Solid State Disk (SSD) storage and Energy Efficient Ethernet supporting IEEE 802.3az [4] have low power or energy efficient capability.

Based on the description of energy characteristics in EDL, applications can support resource discovery and request the resource with some features. For example, requests of resources with green energy or with low power processor or GPU can be supported.

3.3 Monitor Component

The most significant difference between energy monitoring and performance monitoring is that energy monitoring needs extra instrumentation devices to determine power or energy. The power is measured on these instrumentation devices instead of resources that consume power. For example, power value of a resource is retrieved from a PDU at the outlet. Therefore, the relationship between resources and instrumentation devices must be understood.

The relationship between resources and instrumentation devices is described by *Hardware Sensor Component* under the *Sensor* class. Sensors include hardware sensors and hardware sensors. The *Software Sensor Component* is software systems or tools which monitor the performance attributes that are not available from hardware sensors. The *Power Meter* class represents the instrumentation device for monitoring, usually a PDU. Each node can be monitored by a PDU, which usually has multiple modules. Each module includes multiple outlets that attach to different resources. Each specified *Outlet* is only responsible for providing the measurement data of the resource attached. The property *attachTo* describes one-one mapping between a resource and a outlet of the power meter. PDUs are differentiated by model or type, which feature an access address and APIs to collect measurement data. The *Driver* of PDU describes these information.

Besides the relationship of resources and power meters, *Monitor Component* describes the organization of measurement data. The *MonitorLog* is a collection of *Load* in different metrics with additional properties about sampling time interval as well as start time and end time of sampling. The measurement data in the same metric consists of a load. Each Measurement instance in a load represents one measurement that has an *metricValue* modeled by *datatype* properties with *xsd:double* type value. The measurement is sampled at the time point of *timestamp*.

3.4 Features of EDL

From the description of EDL in above sections, we can conclude its three features.

EDL supports a wide range of energy management scenarios. It defines an energy measurement log, which allows the representation of measurement data in different metrics at different sample time interval. These measurement data is necessary for various power managements. For example, the analysis of measurement data statistics can solve power models which are function of performance metrics. The models are used to predict power consumption of one scheduling policy. Besides of data support for general power management, we give two other examples to show how EDL works in terms of estimating agreements and resource discovery.

Energy budget accounts for a significant part of cost of operating infrastructures. On Clouds or Grids, the price that operators can charge depends on the performance they advertise; but the energy consumption levels of these resources are diverse. Clouds or Grids users can be encouraged to wait for utilizing the combination of energy efficient resources with low performance requirements, and at the meanwhile the operators refund part of profits from the cost of saving energy. But the users are worried about whether the amount of refund is real and is worth. They need to reach agreements with the operators.

EDL can describe the agreements, as there exists the definition of energy cost and capacity of resource components. The operators use the classes and properties in EDL to design an automatic mechanism for disseminating claims about resources capacity and energy cost in the agreements. The user sides have an explicit EDL model and are capable of understanding agreements. In this way, online agreements are easy and low-cost to reach.

EDL is able to support green resource discovery across different domains, which finds out the green network or computing resources. Although some infrastructures or part of resources have been using the green energy source, absent method of differentiating them from other infrastructures or resources limits the effect of power management. The terms such as the type of energy source of the resources and GHG emission rate generated are abstracted in EDL. It enables operators to search these information across domains in order to discover resources in terms of their energy and sustainability state.

EDL contributes to flexible and low-cost data manipulation in a large-scale e-infrastructures environment. In a large scale e-infrastructures environment, heterogeneous resources have drastically different energy-related state for different resource components, and their state and components are dynamically changing. For example, the servers dynamically switch the renewable solar energy to the brown energy in the evening; while the storage of servers is upgraded with energy-efficient SSDs. The description of the energy-related state of resources and the description for resource components in EDL are loose coupling. Because of this, the description of each heterogeneous resource can be diverse in the state and components. In addition, the update on resource state by applications doesn't influence resource components, vice versa. So updating the description of the large-scale infrastructures is flexible and low-cost.

Ontology is modeled by a semantic approach. EDL file follows the OWL standard, which is easily reused by other ontology developers.

The Energy Knowledge Base system (EKB) [5] which builds a distributed semantic database for energy monitoring of DAS-4 clusters has adopted EDL ontology. Data in EKB is used for power modelling and power management. EKB defines some common metrics such as CPU, memory utilization and power consumption at first. EKB associates the monitoring components with resources. After that, EKB accesses the PDUs and outlets through access information from drivers. EKB create one log for each resource and collects power and performance profiling information. At last, EKB inserts resource information and log information into database. With EDL, the EKB users can properly access and manipulate the measurement data.

4 Related Work

Within the area of energy monitoring in computing infrastructures there are a number of models which have been designed to capture energy concerns.

Daouadji et al. developed is an ontology-based resources description framework for resource allocation purpose with minimal CO2 emission. In the GSNONT ontology [6], each resource has an associated energy type, which is categorized into Green and Brown. Each type of energy source has a property, which qualifies the CO2 emission per energy unit. We find their model is mostly applied for green routing path selection.

The Common Information Model (CIM) [7] is a comprehensive model for ICT systems and devices. *Power Source* describes the output power of entities that produce power; *Power Supply* captures the capabilities of input voltage and frequency entities that supply power. The *Metric* details the measured value by *Sensor* classes. But the *Power Source* lacks the property that describes sustainability of source mentioned in GSNONT. Moreover, the *Sensor* class has no property about connection with resources monitored. When one sensor monitors multiple resources meanwhile, the information read out from the sensor is hard to correlate to correct resource. Given the complexity of the CIM, it's not easy to be used it by other ontology users or developers.

The Green Information Model (GIM) [8] is a parallel work with EDL. GIM includes the concepts of the power source, power supply and power monitoring. However, it only supports basic energy-related metrics such as power and current. In fact, the energy metrics that quality the resource can be numerous from different angles such as GHG emission, electricity cost and efficiency. This limits the range GIM can apply in.

The EMAN working group in IETF is working on an information model for power management of networks [9]. EMAN defines monitoring and control functions, but does not model important non-measurable state about the network resources, such as hardware capabilities or types of energy sources.

5 Conclusions

In this report, we have shown how we define EDL as an extension of INDL. EDL is a semantic model which defines abstract concepts and relationship of information about energy monitoring and energy measurement data. The information is the basis of diverse power managements in large-scale infrastructures.

EDL will keep evolving with its implementation in an infrastructure. Current EDL captures state of each resource, and sometimes it lead to the inefficient of information update. For example, each resource has one type of *Energy Source*, e.g. solar. When no solar at night, the energy source of each resource need to be switched. It leads to heavy cost on update of resource description in large-scale infrastructure. Therefore, for an infrastructure where all resources always use same energy source, current EDL can be improved by correlating *Energy Source* with the overall infrastructure rather than each resource. Moreover, EDL doesn't capture the information about how to control and manage resources. This is our future work.

References

[1] M. Ghijsen, J. van der Ham, P. Grosso, C. Dumitru, H. Zhu, Z. Zhao, and C. de Laat, "A semantic-web approach for modeling computing infrastructures," in *Computers and Electrical Engineering (To appear)*, 2013.

[2] J. J. van der Ham, F. Dijkstra, F. Travostino, H. M. Andree, and C. T. de Laat, "Using {RDF} to describe networks," *Future Generation Computer Systems*, vol. 22, no. 8, pp. 862 – 867, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X06000410

[3] G. Dobson, R. Lock, and I. Sommerville, "QoSOnt: a QoS Ontology for Service-Centric Systems," in *31st EUROMICRO Conference on Software Engineering and Advanced Applications*. Ieee, 2005, pp. 80–87. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1517730

[4] *IEEE Std 802.3az-2010 (Amendment to IEEE Std 802.3-2008)*, http://ieeexplore.ieee.org/servlet/opac?punumber=5621023, 2010.

[5] H. Zhu, K. Van Der Veldt, P. Grosso, X. Liao, and C. De Laat, "Ekb: Semantic information system for energy-aware monitoring in distributed infrastructures," in *Cloud and Green Computing (CGC), 2013 Third International Conference on*, 2013, pp. 60–67.

[6] A. Daouadji and K.-K. Nguyen, "Ontology-Based Resource Description and Discovery Framework for Low Carbon Grid Networks," in *2010 First IEEE International Conference on Smart Grid Communications*, Oct. 2010, pp. 477–482.

[7] (2013) DMTF Common Informantion Model(CIM). [Online]. Available: http://dmtf.org/standards/cim

[8] R. Carroll, A. Mackarel, and A. Pastrama, "Networks and services interconnection between Mantychore and GSN," Mantychore Project, Tech. Rep., 2013.

[9] (2013) Energy management (eman) model. [Online]. Available: https://datatracker.ietf.org/wg/eman/

Appendices

A UML representation of EDL components

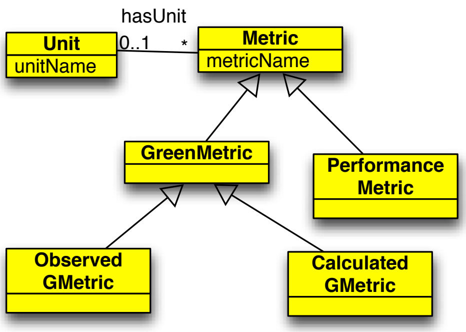

Figure 3: UML representation of Metric Component

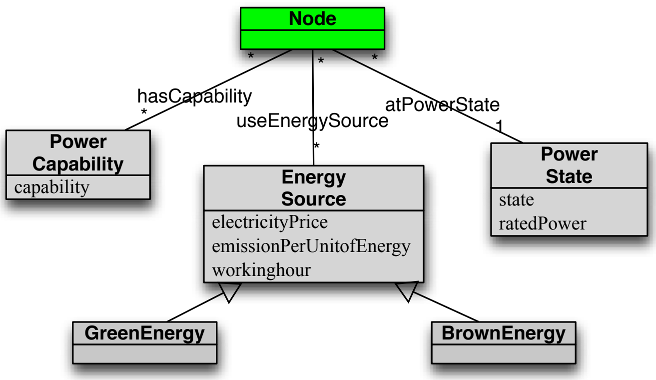

Figure 4: UML representation of Energy-aware Characteristics Component

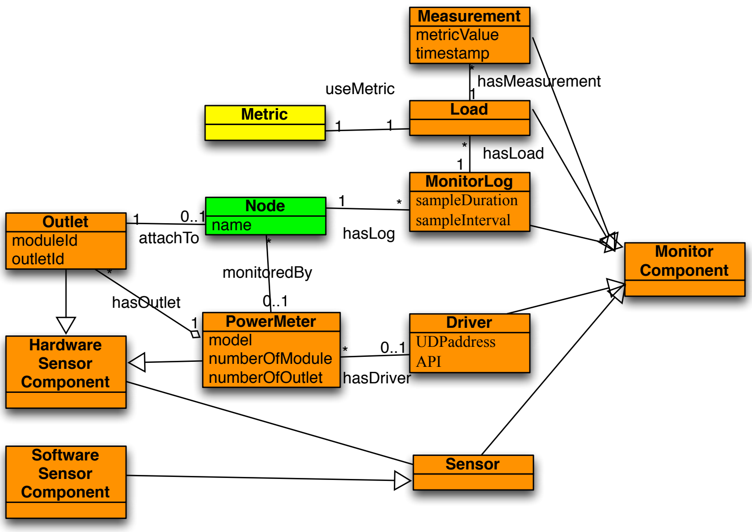

Figure 5: UML representation of Monitor Component