Cyber Physical Power System Modeling and Simulation based on Graph Computing

Jiangpeng Dai¹, Zhangxin Zhou², Ziyan Yao¹, Ming Ni³, Guangyi Liu¹, Renchang Dai¹ and Zhiwei Wang¹

¹Global Energy Interconnection Research Institute North America
San Jose, CA, USA

²Department of Electrical and Computer Engineering
Texas A&M University
College Station, TX, USA

³NARI Technology Co. Ltd
Nanjing, Jiangsu, China

Abstract—As an extension of Cyber Physical System, Cyber Physical Power System (CPPS) is to describe a communication, computing, control and physical coupled system for the power grid. While the computational complexity increases, it is not the best option to utilize relational database and matrix computation to implement real-time processing and analysis in CPPS. The graph computing has demonstrated its capabilities in managing large-scale data and executing the distributed parallel computation algorithms. This paper firstly uses graph database and graph computing to model and simulate the CPPS. The three-layer framework of CPPS is implemented with the graph database. The efficiency of information collection and command transmission is analyzed based on graph computing. The simulation shows that the time latency's calculation is significantly simplified by using graph computing which converts matrix computation problem to a graph traversal problem.

Keywords—Cyber Physical Power System, Graph Database, Graph Computing.

I. INTRODUCTION

The electric grid is the world’s most complex interconnection engineering system. Its complexity is continuously increasing as renewable energy, electric vehicles, energy storage and advanced measurement devices are integrated into the system. On the one hand, a lot of measurement devices and communication network give us more opportunities to estimate the grid condition, predict and control the power system. On the other hand, the understanding, modeling, and protection of the system need to be further investigated. The Cyber Physical System (CPS) was coined in 2006 by researchers to describe a communication, computing, control and physical coupled system[1]. The smart grid is one typical Cyber Physical System so that a concept of Cyber Physical Power System (CPPS) is raised. It focuses on how to model the communication network in power grid system and assess its impact on the power grid. Different approaches have been proposed to understand, to shape and to co-simulate the CPPS (e.g. [2-6]). However, the sophisticated modeling and computation will bring the control center new challenges.

The relational database is widely applied in the power grid control center to store and process the data. However, due to the enormous amount of data and sophisticated modeling in CPS study, the traditional relational database may not be the best option to implement real-time analytics and parallel computation applications.

The graph computing has demonstrated its potent capabilities in managing large-scale data and implementing the distributed parallel computation algorithms by replacing matrix computation [7]. In numerical calculation, an element \( a_{ij} \neq 0 (i \neq j) \) in matrix \( A \) is equivalent to one edge existing between vertex \( v_i \) and \( v_j \), while \( a_{ij} (i \neq j) \) refers to the attribute on edge, and \( a_{ij} (i = j) \) refers to the attribute of \( v_i \). Compared with matrix computation, the efficiency of graph computing relies on the merits as follows.

1) It manages mass data with internal dependence by the structure of graph. The relationship within data sets can be easily fetched through an edge rather than by matrix operations.

2) It combines data structure and processing. While graph data is stored in the specific structure of graph, the graph data process equals to graph structure analysis. The typical form is called a graph traversal.

3) It realizes real-time data update by dynamically modifying states of vertex and edge in the graph, much easier than in matrix.

4) It naturally supports BSP parallel computing model [8-9], which is mainly used to deal with the data synchronization and calculation iteration problems of multi-computing nodes in the large-scale network. Inspired by BSP model, the Pregel is proposed as a graph processing framework and used to solve parallelizable problems by the way quite similar to MapReduce.

Therefore, the vertex and edge in the graph can play both roles of the storage unit and the parallel computation unit. The communication between vertices in a graph is realized by exchanging information through edges. Local computation in vertices and cooperation through edges are utilizing in the graph computing to achieve large-scale parallel computing more efficiently.

In this paper, we firstly model the CPPS using the graph database and simulate the running of CPPS based on graph computing. Section II introduces architecture of CPPS by discussing the physical layer framework, the cyber layer framework, the interaction layer framework and CIM/E Modeling. Section III demonstrates CPPS modeling based on graph computing and the advantages of doing so. Section IV
describes the graph computing-based simulation for CPPS. Section V concludes.

II. ARCHITECTURE OF CYBER PHYSICAL POWER SYSTEM

Conventionally, the Cyber Physical Power System consists of three layers, i.e., power grid physical layer, power grid cyber layer and CPS interaction layer. Each layer comprises of different objects and achieves different goals. We describe them as follows.

A. Power Grid Physical Layer

The physical layer of power grid contains the vast amount of equipment, such as generator, load, energy storages, power electronics devices, smart appliances and transmission lines. In traditional Energy Management System, two basic models are describing the physical connection: node breaker model and bus branch model. The node breaker model is a full-topology model to represent the specific actual physical components, such as circuit breakers, disconnectors, etc. The bus branch model is an abstract and aggregated model to describe the substation configuration and connection by applying the topology processing to the node breaker model.

B. Power Grid Cyber Layer

The cyber layer of the power grid consists of many information functions, such as measurement, computation, control, decision making, and management. This layer can be regarded as a complex embedded system doing data acquisition, computation, and analytics for the power grid. Through this layer, the “what now” and “what if” questions of power grid conditions are analyzed, and the operation and control commands will be created.

C. CPS Interaction Layer

The CPS interaction layer mainly contains a private or public network composed of multi-type communication devices. It builds up the coupling relationship between the physical layer and cyber layer. The information collection and command distribution between these two layers are accomplished by the communication network in the interaction layer. Meanwhile, the communication in the interaction layer is modeled with consideration of delay and disruption.

D. Framework of Cyber Physical Power System

Corresponding to the hierarchy in CPPS, we introduce a three-layer framework for CPPS modeling, which consists of the physical layer, information layer, and cyber physical coupling layer, as shown in Fig. 1.

In this framework, the cyber physical coupling layer is a crucial factor, since it performs as a bridge connecting the other two layers. This layer is composed of secondary devices and plenty of communication nodes, which forms a coupling network. The secondary device can collect and send information, or provide control services by real-time data analysis, depending on what kind of substation it is affiliated with. The communication nodes are designed to realize information transmission.

E. CIM/E Modeling

EPRI developed common information model for representing power system components. It is used as an open standard for exchange data between Energy Management System vendors. The power system model is exchanged using CIM/XML format. In 2011, State Grid Corporation of China (SGCC) developed a model exchange format CIM/E, which has smaller size and simpler form [10]. Due to the smaller size of CIM/E standard data, the efficiency and speed can be increased during online study and data exchange.
significantly reduce size as there are a huge amount of terminals. We model the components in CIM/E with graph database, in which the data size is significantly reduced compared with the modeling in S.A. Khaparde’s work [11].

III. GRAPH COMPUTING-BASED MODELING METHOD FOR CYBER PHYSICAL POWER SYSTEM

A. CPPS Modeling based on Graph Computing

1) Modeling CIM/E Data with Graph Database

Modeling the power grid equipment in CIM/E into graph database is the first and most fundamental step to apply graph computing to provide a solution for power system problems.

![Fig. 4 Substation Representation in One-line Diagram and CIM/E](image)

We take a substation as an example. Fig. 4(a) shows the model of the one-line diagram of a substation, which consists of the bus bar, circuit breaker, disconnector, load, and generator. Fig. 4(b) shows a CIM/E representation of this substation.

![Fig. 5 Modeling of Substation in CIM/E with Objects Described by Vertices](image)

When the CIM/E standard data is modeled with graph database, all the component objects are represented by vertices in this method, as shown in Fig. 5. This technique can give the operation engineers the straightforward visualization of the substation and provide them with easy access to do the data management and graph model manipulation.

To reduce the vertex set, we can also model the circuit breaker and disconnector by edge. All objects are defined using the graph concept $G(V, E)$. The objects that do not have on and off status are denoted as vertices. Circuit breaker and disconnector, which have on and off state, are denoted as edges. Both the attributes of vertex and edge can be accessed through graph database, as shown in Fig. 6.

![Fig. 6 Modeling of Substation in CIM/E with Objects Described by Vertices and Edges](image)

2) Modeling Cyber Physical Coupling Layer in CPPS with Graph Database

The cyber physical coupling layer bridges the physical and the cyber layer in CPPS. We utilize graph to model the coupling relationship and realize the interconnected CPPS, as shown in Fig. 7. The vertices represent physical devices in physical layer and communication devices in the cyber layer, respectively. The edges among them represent the interaction from one device to another, and hence describe the function of coupling physical and cyber systems. The graph $G(V, E)$ intuitively expresses the complex network system in cyber physical coupling layer.

![Fig. 7 Interconnected CPPS Model](image)

With the attribute of vertices and edges in a graph database, the graph model of cyber physical coupling layer can efficiently store all the necessary information in CPPS
simulation, especially for the interaction between different layers. Section IV introduces the details in the simulation of CPPS using the graph computing technique.

3) Software Package to Convert CIM/E into Graph Database

The basic CIM/E files contain 15 tables, which denote base voltage, substation, bus bar, AC line, generator, transformer, load, compensator_P, compensator_S, converter, DC line, island, topo node, breaker, and disconnector. We developed Python package to convert this file to the loading files used by graph database. The nd number in CIM/E represents the number for connectivity node. We scan all the files and find all the nd columns. Then merge all nds in one csv file and delete the duplicate information. This new file will be the vertex definition of the connectivity node.

For the vertex definition, the critical information to extract is the two connectivity information of an object. For all the conducting equipment that is defined in CIM/E, the neighbors are connectivity node. The bus bar, generator, load, and compensator_P are particular kinds of connectivity nodes.

Each kind of object will be defined as one type of vertex, which means they have the same schema for it. We save all those objects information on the vertex. For the edge, it depends on the vertices information which is connected to this edge. Once the category of two vertices is fixed, the category of the edge is defined. After we have all the loading files of vertices and edges, we can write a query to print the model in visualization window.

B. Advantages of Graph Database and Graph Computing in Real-time CPPS Study

Due to the graph structure and interconnection of the power grid cyber physical system, the graph database is well-suited for its modeling, computation, analytics, and visualization. The typical advantages are: 1) efficient data storage for online application; 2) flexible CPS modeling in graph database and data management; 3) robust to do distributed graph computation; 4) seamless connection to the visualization platform.

1) Efficient data storage in online application: The data is stored in a graph structure, which is similar to the CPPS connection structure. So the query time to do the traversal will be significantly reduced, which provides a tool to solve CPPS online problems.

2) Flexible CPPS modeling in the graph database and data management: The traditional relational database has to do One-to-Many, Many-to-one, and Many-to-Many procedure to explore the CPPS data. The attributes of the model in a graph database, which are attached to the native node, can be directly changed and added.

3) Robust tool to do distributed graph computation: In a graph database, the vertices are both data storage units and dynamic computation units. We could write a query to simultaneously get easy access to neighbor's information by exchanging information through edges and do the parallel computation using graph computation engine. Achieving the goal of real-time energy management is critical. Using the graph computing, the complicated matrix calculation in CPPS information processing becomes the graph traversal problem, which query and compute through weighted route.

4) Seamless connection to the power grid visualization: The graph database also gives us a right tool to do the CPPS visualization in the graph manner, which is the natural structure of the CPPS. This tool will benefit system operators to do online monitoring and decisions.

IV. GRAPH COMPUTING-BASED SIMULATION FOR CYBER PHYSICAL POWER SYSTEM

A. Simulation Scenario

The Cyber Physical Power System simulation based on graph computing aims to do the analysis of impact coming from both of physical layer and information layer, and to reveal the interaction relationship between these two layers, in a simple and intuitive method. Since the different types of vertices and edges in graph present devices in different layers and the relationship between each two, respectively, the paths in the graph show the way how collected information and control information data flow in the CPPS. The attribute value along the paths indicate the performance of interaction among physical and information devices, such as time delay, interruption probability, and disruption probability. By traversal or query in the graph, we can simulate the running of CPPS, especially find the relationship of impacts from different layers in the system.

B. Representation Method

In the simulation of three-layer CPPS architecture, there are different kinds of correlations inside the cyber physical coupling layer, such as the one between two communication nodes, and the one between secondary device and communication node. Also, there are several types of correlations linking two layers, such as the one between physical and secondary devices, or the one between secondary device and information system. All these correlations represent the characteristics of each part in CPPS and can be described with the graph model.

The correlation between two communication nodes is mainly determined by the topological structure. Its direct impacts in the CPPS include time delay, information interruption, and disruption. With graph model, we can use two vertices and one edge between them to represent two communication nodes and the status of their interaction. The attribute \((T_c, P_i, P_d)\) on that edge describes the performance of information transmission. \(T_c\), \(P_i\), and \(P_d\) refer to time delay, interruption probability, and disruption probability, respectively. Since the graph model intuitively represents the connection state of communication nodes, there is no edge between two vertices if these two communication nodes do not share a direct connection. It greatly simplifies the subsequent computational complexity, because only effective data is kept.

In the CPPS, the correlation between secondary device and communication node is used to describe the process of information uploading or command sending. Similarly, we use
two types of vertices to represent secondary device and communication node, respectively and use one edge and its attribute \((T_{sc}, P_e)\) to describe the characteristics of the correlation. \(T_{sc}\) and \(P_e\) refer to time delay and bit error probability of information passing from the secondary device to the communication node. As before, the edge in graph only exists if there is a connection between secondary device and communication node. There is no redundancy.

Furthermore, the correlation between physical and secondary devices describes the process of information collection and command execution in the physical layer. The correlation between secondary device and information system describes the process of information collection and command formation in information layer. We also use vertices, edges and related attributes in the graph to indicate specific characteristics in information processing and transmission.

C. Implementation of Graph Computing

Graph computing platform is an integration of graph database and graph computing engine, as shown in fig. 8. It also exposes specific APIs to simplify graph programming.

![Graph Computing Platform](image)

The graph database has an online management system with Create, Read, Update, and Delete methods. It stores graph data consistent with the graph data model and is optimized for transactional performance (transactional integrity and operational availability). The graph database consists of the underlying storage and the processing engine which is different from computing engine. The underlying storage utilizes a graph data model to represent and query large sets of highly connected data. The processing engine allows efficient execution of typical graph operations, such as single or multi-step graph traversal. It also brings the native graph processing capabilities, supporting intuitive format for the underlying data and index-free query, which lead to more straightforward application design and lower development costs.

The graph computing engine is built on the top of the graph database management system. It is primarily used for graph data mining, such as offline analysis of graph data in bulk. However, it provides OLAP if combined with a graph database. The graph computing engine is optimized and tunable for scanning and processing large amounts of relational information in batches by running global graph computational algorithms against large datasets. It can also update graph data in real time through the interaction with the graph database.

D. Case Study

We take the information collection and command transmission efficiency analysis as an example. The object is to calculate the time latency of collected information uploading process from the physical layer to information layer.

![Fig. 9 Simulation of Security and Stability Control System based on Graph Model](image)

We abstract a security and stability control system (SSCS) of a real UHVAC/DC hybrid system into Cyber Physical Power System and simulate it with graph model as shown in Fig. 9. There are four types of components in this system, i.e., main station, slave station, execution station and communication node. Each station connects to a communication node. Every communication node also connects to another node if there is a communication route between them. The information transmission is realized through these connections inside the system.

In the security and stability control system, the sequence of execution is described as below. One execution station (ES) sends a fault message to a slave station (SS) to which the ES is affiliated, and then the slave station (SS) send it to the main station (MS). The main station (MS) identifies N-2 fault and sends control commands to related execution stations (ES). From the perspective of information transmission, the secondary devices collect and process measured data from execution station, then upload them to information processing system in the main station through communication node network.

The time latency of information uploading in this execution sequence is composed of several segments, including the time cost of information collection and processing in each execution station, the transmission time between execution station and communication node, and the time latency between secondary device and information system inside the main station. All these time latency data are attached to the edges in the graph. We can get rid of complicated matrix computation, by using graph traversal method to find the shortest weighted path, which refers to the most efficient information transmission.
route. We can also calculate the total time latency by sum all the time cost attributes on every edge along this path.

Fig. 10 Time Latency Data from SSCS

We use a set of time latency data from a real security and stability control system [12], as shown in Fig. 10, to simulate the collected information uploading process from one specific execution station (whose id is 4 in this case) to the main station. $C$ matrix refers to the time latency between every two communication nodes. $C_{ij}$ is the time delay from node$i$ to node$j$. $P$ tuple refers to the time latency of information collection in execution station. While $S$ tuple refers to the processing latency in execution station, $P_i$ and $S_i$ denote the values in station$i$. By the graph traversal, we get four possible paths from ES-4 to MS. While one of these paths has the least time latency. Fig. 11 shows the simulation result. We accomplish the equivalent time latency calculation by graph computing, but more simply and intuitively.

Fig. 11 Information Uploading Paths from ES to MS

V. CONCLUSION

As a complex Cyber Physical Power System (CPPS), the smart grid usually utilizes relational database and matrix computation to implement real-time processing and analysis. However, it is not the best option when the computational complexity is increasing in CPPS. The graph computing has demonstrated its capabilities in managing large-scale data and implementing the distributed parallel computation algorithms. This paper firstly uses graph database and graph computing to model and simulate the CPPS. The three-layer framework of CPPS is successfully implemented with the graph database. The information collection and command transmission efficiency is analyzed based on graph computing. The time latency’s calculation is greatly simplified by using graph computing which converts matrix computation problem to a graph traversal problem. We hope this graph computing method will bridge the gap of computation challenge brought forth by Cyber Physical Power System modeling and matrix-based analysis.

ACKNOWLEDGMENT

This work was supported by State Grid Corporation of China technology project SGSHXT00JFJSI700138.

REFERENCES