Complex Systems Smart Home Security Studies based Big Data Analytics

Zhao Yan

Beijing Polytechnic, Beijing 100176,China
tozhaoyan@sina.com

Abstract

Intelligent home system, due to smart appliances and home access points, communication method used copy, and its electric load and the impact brought about by economic considerations, it makes the intelligent home control and power delivery system has become very complex, complicated accord characteristics of the network; this complex network, its security and Lupin performance becomes extremely important; simply consider the performance of the network topology and other network anomalies can only be expected short-term, however, a large area of network failure systemic, primarily business-related, for this problem, according to the data herein smart home business, a new web-based method of stabilizing the complex network theory.

Keywords*: Smart home business, large data, complex systems, network cascading failure, robustness*

1 Introduction

Residential users and business users class service is a major component of China's electricity users, which accounted for 80% of the number of users more than the total number of electricity users. Distinguished from industrial users, etc., which is more complex composition of two types of users, electricity behavioral characteristics, the law is more diverse. Therefore, for these two types of research-side electricity users to carry out, it is the key to the whole network users of electricity grid orderly and intelligent. The commercial complex electricity, district public side electricity and residential electricity in the concentrated area is typical of these two types of sophisticated users. This study focused on residents of the region for electricity, intelligent electrical equipment installed in the residential side of the first premise, smart power to carry out related key technologies. Under the impetus of networking technology, today, network appliances, such as automatic washing machines, digital microwave, digital TV sets, refrigerators and other intelligent computer has become invisible place, digital, intelligence families have become a reality, various "things appliances "can add remote intelligent control functions through the network, even when electricity abnormal, timely automatic alarm via mobile phones and other mobile terminals. Things realization of "all things" and "efficiency, energy saving, safety, environmental protection" and "tube, control, camp" integration, so that anything in life can become "feeling, thinking," a variety of "no from the eyes "in the form to enter people's daily living space.

With the construction of the smart grid, will effectively promote household electricity monitoring system for real-time monitoring and management of smart appliances, automatic meter reading, automatic payments and other features that make the user's life easier. Second, the smart grid can access a small home wind and rooftop photovoltaic power generation equipment, development and promotion of renewable energy sources, to make life more carbon. In addition, under the support of the smart grid, users can set up household electricity regulatory services platform to help users reduce electricity costs between the community of users to share energy-saving programs to provide more

services, not only to promote energy-efficient economy, and help improve people's quality of life. We currently live in an era of data explosion, a report IDC forecasts, from 2005 to 2020, the global data volume will increase 300 times, increased from 130EB 40000EB. From the early commercial, financial and other fields, and later the transportation, medical, energy and other areas, in May 2011, McKinsey & Company released "Big Data: The Next Frontier, competitiveness, innovation and productivity," describes big data research value. Big Data research has attracted the attention of experts and scholars at home and abroad, has become a research hotspot.

In the energy sector, smart grid is seen as one of the important areas of big data applications. With the power of information technology to promote smart substation, smart meters, on-line monitoring system, on-site mobile maintenance systems, measurement and control systems, and the integration of a large number of professional services in various information management system and gradually built the application, the power industry generated a lot of structure diverse, complex sources, the huge size of the data, which are widely used in power system planning and operation, asset management, and electricity market management and end-user services, smart grid data generating electricity in large power system, transmission, substation , distribution, electricity and scheduling each production process. Currently in power system simulation, data transmission equipment state detection, advanced metering infrastructure data analysis, short-term power load forecasting has been carried out big data research. Big Data analysis based on cloud computing technology has a more complete reference architecture and software, and has been applied in some industries. However, these systems are mostly from the Internet industry, its design, implementation and features are fully reflects the needs and characteristics of big data and Internet services. When applied to the Internet outside of the industry, a common practice in large data characteristics of the industry, the basis for full and detailed business needs analysis, for the limitations of existing systems, custom development and optimization. For large data analysis constructed before a large-scale real estate smart power means cloud platform, a former real estate means for large-scale intelligent power and data mining methods cloud platform structure, this section will be discussed in detail how to use cloud computing technology to solve analysis of the difficulties encountered in the system before the installation of large power real estate data. To analyze from a storage mechanism, computer system and job scheduling aspects.

Resource allocation and task scheduling is the key to cloud computing, the computing resources used to mask and heterogeneous nature of complex cloud environment, cloud computing resources of each node unified management and allocation, enable resources to effectively share, work together to improve the effective utilization of resources. Meanwhile, according to the needs of users, a reasonable allocation of appropriate resources to clients to achieve unified business expenses and profits. Rational resource allocation and task scheduling strategy can not only efficient use of computing resources in the cloud environment, it can also give full play to the advantages of heterogeneous resources, highlighting the timeliness and geographical characteristics of resources. Furthermore, through the efficient allocation of resources, but also for resources such as resource potential failure, node failures, power outages and other unexpected computer problems make timely feedback, in order to better respond to and improve the robustness of the accident and fault tolerance system , quality of service to further improve the system. In the cloud computing environment, a user submits a request to the cloud service on demand. To increase the throughput of the system, the system will first break down tasks users submit. Division of tasks is to try to eliminate the overhead caused by interactive communication between servers, reducing overall running time of decency. General requirements as much as possible so that the maximum parallelism, Correlation between tasks to a minimum.

However, according to a large data mining Cloud computing, associate degree and the extent of the impact of business tasks are relatively close, so very easy to produce a cascade of large-scale system failure, so be sure to analyze large data according to the data service , according to the principle of self-organized criticality be analyzed in order to effectively prevent a wide range of related degree related fault caused; those nodes distributed in space has a metric among the networks network is. Usually a misconception is that the concept of space network and mixed network plane slippery. Flat network nodes embedded in the space of the same, but even that defines the network edge node can not except in the presence of the cross. The Space Network is no such restriction, we can only say that is a special case of a flat network space network[1-3].

Since the network node space is embedded in the space has a metric attribute among the introduction of spatial factors will affect the structure of the network, and thus influence behavior on the network, which makes the space network will show some of the traditional networks are not the same characteristic. Generally, spatial network often has the following four characteristics:

1. attribute node having a position; node spatial network has spatial coordinates, that is, for a network user, if he knows the coordinates and coordinate their own network in a node, then he can determine the position of that node. Space Network location property can bring a good navigation performance, that is, although the user may not be aware of the overall topology of the network, but they can often be more effort to shuttle in the network. Road network is a good example of life experience tells us that usually only understand the general location of the destination, then we can determine quickly go over there, how to go, and in most cases we can achieve relatively smooth destination.

2. Even with a side length attribute; even the shortest physical length of the side of the space network is the straight line distance between two vertices connected edges, but in reality the network due to geographical factors and other restrictions, even space network edge is often curved. And even the length of the edge is related to the cost of even the edges, such as smart home operation costs and network connection costs. Often, even the price is proportional to the length of the side edges even. At the same time, the structure of the space network is often closely related to the length of the properties of the network connected to the edge, for example, research and others show that in the space network design for network transmission efficiency with different tendencies construction costs will result in generating a spatial networks of different topologies . Also worth noting is that statistical studies show that in recent years, even the edge length distribution network is often a lot of physical space follows a power law distribution, which means that in the space network, even the most minor side length, but there are also a small amount across even larger scale long-range side. Research shows that the length of the spatial distribution of the network for many networks behavior (such as diffusion behavior of the network in substance) has an important impact. For a power-law form even space length distribution network side, can the concept of topological space dimension to promote, the definition of the concept of the space dimension of the network. The multilayer composite network is shared between the network nodes of the network a set of layers, each layer of one side of the expression subnet-wide information network. For example, a social network is a kind of typical multi-layer networks, each network layer subnets expression between people in a particular type of connection information. For example, according to the relationship between people, social networking can be abstracted layer by family relationships, friendships layer, the relationship between students and other multi-layer composite network; in accordance with the communication tool and can be abstracted by layer, layer, layer of micro-channel and other multi-layer networks. As another example, the classical physics of particle diffusion reaction (network, depending on the diffusion rate of diffusion channels or different particles can be abstracted by a different layer of the composite particle transport network. Again, real-time communication network traffic flow dynamics, according to time communication occurring, different time periods can be

abstracted from the network layer of the composite ground, wherein each layer of the network based on the communication records generated on behalf of a certain time period of the communication network. the multilayer composite network through multi-layer polymeric (the way to get converged networks, but converged networks will lose a lot of useful information. for example, in the polymerization of social networks, and even the side there is a link between people can be characterized, but can not distinguish between specific types of links in the polymerization communication network, even the presence of the edge can be characterized by the physical link between the router and the router, but can not reflect the differences between the time periods of traffic routers. in addition, the network layers of the composite web structure may exist, differences in behavior. Accordingly, the use of more layer composite network to express certain complex system that can contain more useful information to carry out a more refined study.

These complex network theory to guide the research from the theoretical space of complex networks, this theory with large data mining, this paper proposes a smart home system security solutions.

2. Related Works

2.1 Space Complex Networks

Spatial Analysis on the current network for the analysis of the real space network focused on the transport network and infrastructure networks are more common, important and relatively easy to obtain data space network. Intelligent traffic analysis study, aviation networks are more abundant. The so-called aviation network refers to the airport or city) as the nodes communicate as a node connecting flights resulting network edge. Early aviation research network focused on the static characteristics of network analysis, such as statistical analysis of characteristic path length, distribution and other aviation network. The results show that aviation network is a typical small-world networks. Another important finding is that the performance of the distribution network for aviation power law function and exponential function multiplied form, it is intuitively when the value is small, it is expressed as a power law curve decline, while the larger value , then had an exponential decline curve, indicating that the network does not appear very large value of the node, this phenomenon is due to the volume of space factors make a node (ie the airport) can not accommodate too many flights. Follow-up studies on the basis of the introduction of the airport on the geographical coordinates provided some additional statistical properties[4-6]. Ground bus, subway, train networks are another study based on more space transport network. Such networks different from other types of network space to the line as the basic unit constituting the network, so it can be referred to as line traffic network. Study of such networks need to establish a special network representation. In order not to lose information, generally two types of networks to represent them, are the network and the network. In the network, two nodes (sites) exists if and only if at least one line adjacent nodes connected to both sides; in the network, two nodes (sites) exist side simply connected two nodes at least reach each other by a line. Figure is an example diagram of these two networks. In the framework of the network and the network, a lot of analysis on line traffic network has been carried out, such as: et al and others were statistically analyzed static network statistical characteristics between India and China's railway network; et al studied the Chinese railway network the degree of correlation between the neighboring site. On line traffic analysis study based on a network of networks are mostly based on the traditional network of statistical indicators, such study based on network traffic display line networks with small world, while the network and network nodes tend to have exponential or power-law distribution of the form. In addition to these studies, there have been some relatively new, in-depth analysis and research, more representative example: the actual passenger, who use the data to

extract the maximum passenger Seoul subway network minimum spanning tree, and accordingly the network analysis et al. from the perspective of study based on the hierarchical European rail network, given the structure of the physical layer of the network traffic flow for logic layer has an important influence, as well as spatial factors such traditional between　centrality can not be good conclusions describe the actual network traffic flow. Analysis of other space transport network also includes the maritime shipping network, a public bicycle network and other network analysis. For example, the use of public bicycles, who also recorded by a number of cities around the world to establish their own public bike network (where node is the public bike rental sites, and for those more frequent public bicycles to borrow the car the car side of the site to establish a connection, analysis display public bicycle network has obvious characteristics of the community.

Network infrastructure to provide space for network substance information transfer service to the community. Urban road network, resources and energy transmission networks, network routers belong to the typical network infrastructure. Analysis of such networks in addition to the traditional network contains a large study based on statistical analysis (for example, for the distribution of the network, beyond the statistical analysis of the clustering coefficient, characteristic path length, but also include many of the study based on the actual situation of the network close related analytical study based on, for example:., who used a mobile phone location data of a large amount of data over time, the establishment of road network traffic flows demand matrix row two typical American city, and accordingly carry out these two cities as well as the distribution of traffic flow traffic congestion analysis study based on respect, given a section of traffic jams often stems from a small city conclusions specific regional traffic flowing lines; Youn et al Dresden road network were analyzed and found by network connected side (ie road) area surrounding the generated mesh approximate power law distribution, Carvelhol and others using the actual traffic flow data on the global three metropolitan road network traffic flow analysis, the individual is still the only convenient, lack traffic guide road network Conclusion and easy network congestion, the road can not be efficient use of resources and other issues; Lammer et al European gas transmission pipeline network space network analysis showing that the network has a highly efficient transmission performance, and better able to withstand a certain random damage range impact on the global pipeline to bring natural gas transmission[7-8].

2.2 Big Data Research Situation

Big data services major role in business planning, government decision-making and discovery of new knowledge played on such big data is becoming an important strategic resource, by the universal concern of the government, industry and academia. The current status of the study, the researchers seek opportunities cloud computing and big data integration in various fields, aimed for the big issues in the field of data processing, cloud-based solutions to solve problems in the field of big data processing. For example, researchers from the meteorological scientific point of view, analysis of meteorological data analysis big challenges facing off against the challenges proposed to build and deploy cloud environment weather big data analysis services IBM researchers fine from big data analysis, data display model Angle[9-10], through the analysis of enterprise financial data to dig out financial fraud model tW. Jens, who will explain how to use Amazon ・ EC2, NERSC multiple cloud platforms Magellan and Future Grid peer collaboration to achieve NASA discovered extraterrestrial "project scientific computing tasks. In the actual study, the fusion of cloud and big data services mainly in two ways: (1) the migration of existing applications to the cloud service to perform, or to develop and deploy large data services in the cloud; (2) study of cloud-based big data processing techniques and tools on the one hand, in recent years. with the development of computing

technology and 5 gradually mature, from both cost and performance reasons, more and more enterprises and spy organizations choose to move applications to the cloud environment in the implementation, and to develop and deploy large data services in the cloud. for example, Facebook researchers in their research work describes how Facebook to its internal application migration to hybrid cloud executed; Wu et al. discuss how technology-networking application migration to the cloud in its research clever work by utilizing cloud rich resources and on-demand calls, pay-reducing operation and maintenance cost advantages W applications; Zhang et al., from the perspective of large data processing efficiency Gao, an optimization model is designed for geographically distributed data resources efficiently migrate to the cloud, efficient processing of large data .Facebook, Twitter and other companies choose to use cloud infrastructure as a large data services by using MapReduce and Hadoop and other big data technologies and tools to achieve PB-level data processing PW. Another aspects of cloud computing as a large data processing technical support, spawned a series of cloud-based big data processing techniques and tools. At present, the leading big data processing techniques and tools are mainly Internet companies such as Google, Microsoft, Amazon and Ali cloud, which mainly provides large data storage and analysis services to the enterprise by combining data with external data analysis to help companies get the data value, for example, Google provides users Google · Big Query products to support massive data analysis; Microsoft by Windows · Azure · MarketPlacet62 Azure-based big data analysis services; Amazon in its provides users with big data analysis services; Ali cloud also for large data processing needs of users, providing a large range of data services solutions. · · ·

Currently, the cloud environment-related research services focused on large data or how to present large data applications migrate to the cloud to perform pounds, W reduce the cost of building large data services or how clever research had some big data processing techniques or tools to design more efficient large data processing algorithms in practical applications, from data services to build large angle, the current lack of a common data service application model large study, to provide technical information to quickly build big data services. Specifically, under the cloud of a large common mode data service applications, data and services to be considered large build process multiple techniques, including: how to collect data and services required to build large large data resources; how to choose a cloud of existing resources services to meet the large demand for data services, storage or computing and other IT resources; and how to design large data analysis algorithms to achieve big data processing, in order to complete the implementation and deployment of large data services. Therefore, how to design the cloud environment of large common mode data service applications, data and services to efficiently build large significance.

3. The Proposed Scheme

Based on the characteristics of smart home, smart home gives the analysis of the spatial complexity of the network, according to the space complex network analysis proposed based big data analysis algorithms suitable for this basis and in accordance with the algorithm proposed scheduling of cloud computing , thereby obtaining a smart home system to prevent cascading failures of network security solutions based on big data analysis algorithms.

3.1. Fuzzy Clustering Smart Home based Data Analysis Methods

FCM algorithm is a clustering algorithm based on partition, so it is thought to be divided between the greatest similarity with a family of objects, and the minimum degree of similarity between the different clusters. Do it on the basis of the mean ordinary C above a certain improvement in ordinary C-means algorithm for data classification is rigid, and FCM is a flexible fuzzy division by introducing membership functions to represent

each data belonging to different level categories, the basic steps of soft data partitioning algorithm is as follows: first to estimate the central data clustering, and then constantly adjusts the cluster centers according to the actual situation, the data to be set for each point to the center of each cluster the termination condition is met when the minimum distance and or, the algorithm stops.

Let n samples of data collection is characterized by a vector of samples $X = \{x_1, x_2, \cdots x_n\} (\subset R^m), x_k = (x_{k1}, x_{k2}, \cdots x_{km})^T (\in R^m)$; the number of categories to $c(2 \le c \le n)$ the data sample is divided; clustering prototype of class i is $p_i = (p_{i1}, p_{i2}, \cdots p_{im})^T (\in R^m)$, constitutes a cluster prototype matrix $P = (p_1, p_2, \cdots p_c)(\in R^{m \times c})$; the membership matrix $U = (u_{ik})_{e \times n} (\in R^{e \times n})$, which u_{ik} shows a sample x_k for clustering prototype degree of membership, and p_i, $u_{ik} = [0,1]$ based on the fuzzy clustering analysis of the objective function can be expressed as:

$$\begin{cases} \min J_b(U,P) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^b (d_{ik})^2, b \in [1, \infty) \\ s.t. \sum_{i=1}^{c} u_{ik} = 1, k = 1, 2, \cdots, n \end{cases}$$

（1）

Where: The blur degree index b, is a flexible parameter control algorithm, if b is too large, the clustering effect is poor, on the contrary, close to the traditional method of C-means algorithm, the range of literature is usually described $[1.5, 2.5]$, d_{ik} shows a sample with dissimilarity measure between x_k cluster p_i prototype is usually taken as the Euclidean distance, ie

$$d_{ik} = \sqrt{\sum_{j=1}^{m} (x_{kj} - p_{ij})^2}$$

（2）

(1) parameter settings: In the study coordinated optimal control area commercial complex electrical control system, usually on their electrical equipment in commercial complex electrical control influence is divided into several levels, that is the key node, an important node and isolated nodes, etc. Therefore, the degree of importance for different nodes, we take the number of categories; and accordingly take the blur degree index; iteration termination threshold; iteration counter; to arrive clustering prototype $p^{(t)}$;

(2) calculate the corresponding partition matrix $U^{(t)}$: for $\forall i, k$, if $\exists d_{ik}^{(t)} > 0$, there

$$\mu_{ik}^{(t)} = \left\{ \sum_{j=1}^{c} \left[\left(\frac{d_{ik}^{(t)}}{d_{jk}^{(t)}} \right)^{\frac{2}{b-1}} \right] \right\}^{-1}$$

（3）

If there are i and r, such that there are $d_{ir}^{(t)} = 0$, $u_{ir}^{(t)} = 1$ and $j \ne r, \mu_{ij}^{(t)} = 0$ on;

(3) clustering prototype matrix $P^{(t+1)}$, to update:

$$p_i^{(t+1)} = \frac{\sum_{k=1}^{n} (\mu_{ik}^{(t)})^b x_k}{\sum_{k=1}^{n} (\mu_{ik}^{(t)})^b}$$

（4）

(4) If $\left\| P^{(t)} - P^{(t+1)} \right\| < \varepsilon$ the algorithm stops and output divided matrix U and cluster prototype P, otherwise orders $t = t + 1$, steering (2). $\| \ \|$ F- matrix norm,

$$\left\| P^{(t)} - P^{(t+1)} \right\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{c} (p_{ij}^{(t)} - p_{ij}^{(t+1)})^2}$$

$$（5）$$

3.2 Cloud Computing Architecture

Cloud computing can be viewed simply as a system service, including service providers and the use of two main tenants. Cloud computing operating system providers, content distribution services provided by the network; tenants to customize their individual needs to the service provider according to demand, and based on permission given by the service provider to access the cloud computing system. Currently, a number of cloud services providers provide monitoring and display modules, such as Salesforce, to show users real time system health. However, the data is more than a simple display log data, for the purposes of the tenants is difficult to directly guide the selection of their behavior. Therefore, we think we should increase the model analysis module monitoring module, so that the whole cloud computing truly dynamic, observable and controllable service system. Tenants and providers can get real-time indicators of its property interest and take appropriate measures to ensure that cloud computing systems are becoming eco-efficient operation of the system.

After analysis of cloud computing, cloud computing architecture can be divided into four layers: the physical layer, resource layer architecture, development platform layer, application layer. 4-layer structure of cloud computing architecture, including four-tier service, can be UI interface to Web Services provided to the user, all services have a reliable, secure, scalable, on-demand services, and economic characteristics. In addition, the table also lists the five typical market goods and services

Physical layer: refers to the geographical distribution of different local resources in the country, providing local support resources. Local resources can be counted a computing resources, storage resources, sensors, servers, networks and other local resources. The physical layer is the underlying infrastructure for cloud computing. The physical layer is responsible for operation, management, maintenance and upgrading of physical resources to provide Haas (Hardware as a Service) services to a huge IT needs of large enterprises.

Resource layer architecture: distributed applications _L refers to the core layer of deployment of distributed resources to provide basic services. The basic services of the distributed resource layer include infrastructure services (IaaS), Data as a Service (DaaS), e-commerce as a Service (CaaS). Where infrastructure services (IaaS) is a distributed computing services, to provide a flexible, efficient, high-intensity computing services, infrastructure services (IaaS) primarily through virtual technology. Data as a Service (DaaS) is a distributed storage service to provide reliable, secure, high-capacity, easy data storage services. E-commerce as a Service (CaaS) is a network communications services, to provide a reliable, secure network communications services.

Plat layer: refers to a variety of cloud application through the API developers computing programming environments, while also providing extended application developers, load balancing, support diverse service authorization, email, user interface. PaaS to accelerate the deployment of application services, application support service extensions.

Application Layer: is provided by the development platform development environment and market demand, we developed a variety of applications. Application provider is responsible for software development, testing, operation, maintenance, upgrades, to provide users with safe, reliable service.

Cloud computing is now more commonly used way is Hadoop HDFS distributed storage, the advantages of this approach is that you can in the form of distributed storage of large amounts of data quickly and securely stored and can not only ensure the safety of storage, but also to achieve data sharing. This is very beneficial things smart home system from home network to the network and then to implement community WAN, between this application and smart home applications, among home, can be achieved between the various communities so as to form a data sharing smart home data storage center.

Smart home system will deploy cloud services on a virtual machine instance of the application server and Hadoop's HDFS systems connected in this way is to chosen Hadoop HDFS file system is mounted on the virtual machine as a virtual machine's storage directory, using technology is webDAV technology. webDAV (web-based Distributed Authoring and versioning, web-based Distributed Authoring and Versioning) is based on HTTP 1.1 is a communication protocol. It does this by choice GET, POST, HEAD, and several other HTTP standard way to add some new methods become extensions of HTTP 1.1 protocol, so that applications can be written directly to the file on the Web Server, and write files when you can add files lock (locking), after writing a file to unlock (unlock), you can also support made to the file version control. Based webDAV can achieve a powerful content management system or configuration management system, is used on Hadoop hdfs-webdav, access form client ==> hdfs-webdav proxy ==> hadoop namenode server.

Root.dat now stored in the system has been registered user name and the corresponding user file block profile IP address, the system can use this file user registration, authentication and access to documents after logging block description IP address of the node file. Node.dat file is maintained throughout the cloud service system all child nodes of the IP address, port, the largest space, the remaining space and other information, the client through the entire document can obtain information about the cluster, enabling a direct connection to each child node complete file storage and computing capabilities, which greatly improves the data file transfer rate, reducing the load on the managed node.

Specific segment information for each file and stored by the system with the user's user name (Username) as the file name of the file block description files stored therein a node, IP address of this node can be found in the file Root.dat . Exists in the system between the different roles of the two types of data transfer: information data and command data, the message data is real-time monitoring data and home information system needed to complete the task, such as the description of the system and a variety of documents and other information, such data relatively large amount of data; the command data CMD and child nodes are managed through the command data to determine their own next step is to complete the work. As a result of the migration policy to calculate storage, the system appears in the file transfer is relatively small, thus greatly improving the efficiency of the system.

1, the flow of the document distributed storage

File storage system during the first connection by the client to the managed node reads Root.dat file data, check if there is the presence of the user, such as the presence of the user data to obtain an IP address where the node block file; by reading the file Node.dat from the list of IP address management node read child nodes, based on the above information to complete data segmentation function also start multiple threads to connect each child node data are stored on each child node, last updated Username table to prepare for the new access find the distribution of documents. Username files will be stored on a node, a node management node assigned to the user based on the distribution of existing Username files stored Username IP address information file, the file name is the user name, because the user name in the system it is unique, so each user's Username is unique and will not cause confusion.

4. Simulation Results and Analysis

This paper uses the Matlab language to realize, in the Intel 64 bit processors and 8 G memory on the machine running, the Windows 7 operating system. Researchers have proposed a number of different evaluation criteria to evaluate complex network data mining standards, such as

Table 1. The Testing Software Environment

Data network	nodes	edge	Snapshot number
Job site topology	4873	343303	20
home	251	14057	20
Stadium	2359	642959	28
Business zone	58186	160224	10

This section analyzes the cloud computing environment, a large data mining convergence speed and network security analysis, Figure 1 shows a place of work in time for the network under different number of snapshots T. Can be found in its run time is not proportional to the number of snapshots, on the contrary, it is inversely proportional to the number of snapshots.

Figure 1 shows the running time of the job site network under different snapshot numbers T. It can be found that the running time is not proportional to the number of snapshots, in contrast, it is inversely proportional to the number of snapshots.
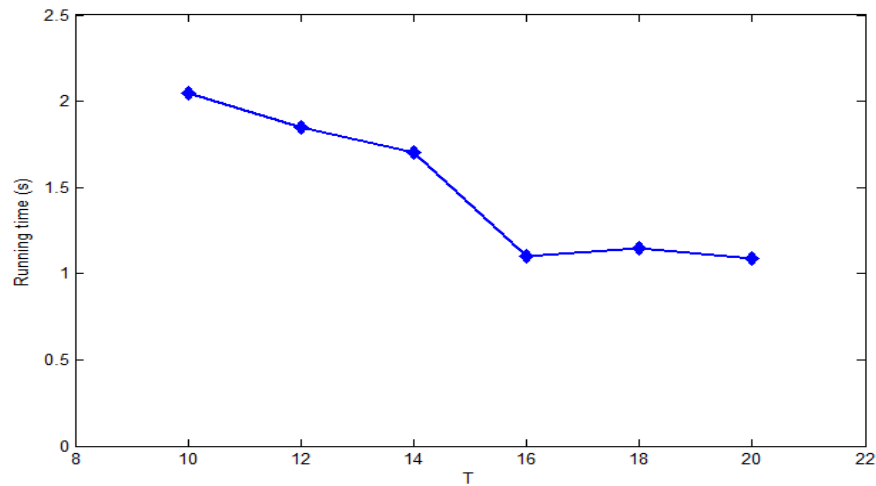

Figure 1. Running tTme

As can be seen from Figure 1, the time cost of the algorithm of single machine mode can be increased sharply with the increase of sample data, and the time cost is increasing slowly, which is more obvious with the increase of data quantity. It can be seen that, with the increase of the number of nodes, the running time is not increasing, but it tends to be steady.
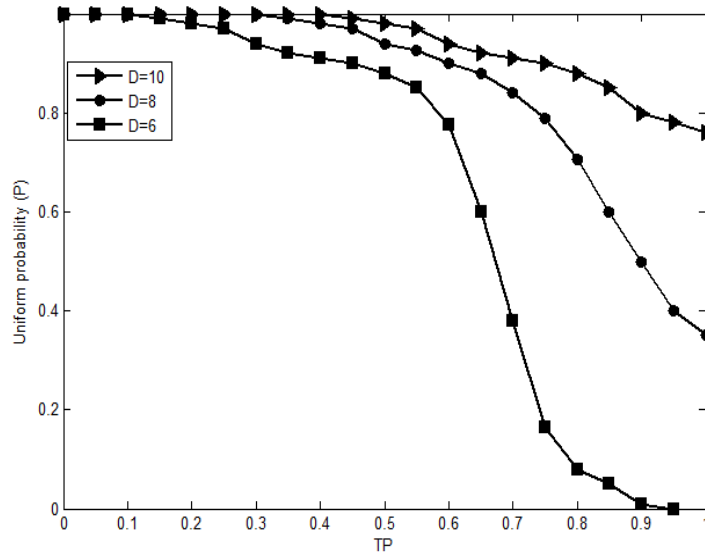
Figure 2. Unified Probability affects by TP Change

The TP value of the variable clustering non-scale network increases, the average maximum number of points will be reduced, the clustering coefficient has an obstacle effect. The main reason is that the higher clustering coefficient implies that the individual neighbor relationship is more closely related to the formation of local individual.

5. Conclusion

Based on the smart home features, space utilization theory of complex networks, according to the characteristics of cloud computing, the use of large data clustering method to decouple the smart home system, thus avoiding the network is large chain system failure risk simulation shows the degree of convergence of the algorithm, and the algorithm can effectively avoid the probability of failure of a large area.

References

[1] Dörfler F, Bullo F. Synchronization in complex networks of phase oscillators: A survey[J]. Automatica, 2014, 50(6): 1539-1564.

[2] Yuan Z, Zhao C, Di Z, et al. Exact controllability of complex networks[J]. Nature communications, 2013, 4.

[3] Yan G, Ren J, Lai Y C, et al. Controlling complex networks: How much energy is needed?[J]. Physical review letters, 2012, 108(21): 218703.

[4] A. Miyaji and M. S. Rahman, ''Privacy-preserving data mining:A game-theoretic approach,'' in Proc. 25th Data Appl. Security Privacy,2011, pp. 186–200.

[5] X. Ge, L. Yan, J. Zhu, and W. Shi, ''Privacy-preserving distributed association rule mining based on the secret sharing technique,'' in Proc. 2nd Int. Conf. Softw. Eng. Data Mining (SEDM), Jun. 2010,pp. 345–350.

[6] N. R. Nanavati and D. C. Jinwala, ''A novel privacy preserving game theoretic repeated rational secret sharing scheme for distributed data mining,'' vol. 91, 2013. [Online].Available:http://www.researchgate.net/publication/256765823_A_NOVEL_PRIVACY_PRESE RVING_ GAME_THEORETIC_REPEATED_RATIONAL_ SECRET_SHARING_SCHEME_FOR_DISTRIBUTED_DATA_MINING

[7] M. Halkidi and I. Koutsopoulos, ''A game theoretic framework for data privacy preservation in recommender systems,'' in Proc. Mach. Learn.Knowl. Discovery Databases, 2011, pp. 629–644.

[8] S. Ioannidis and P. Loiseau, ''Linear regression as a non-cooperative game,'' in Proc. Web Internet Econ., 2013, pp. 277–290.

[9] S. L. Chakravarthy, V. V. Kumari, and C. Sarojini, ''A coalitional game theoretic mechanism for privacy preserving publishingbased on k-anonymity,'' Proc. Technol., vol. 6, pp. 889–896,2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2212017312006536

[10] R. Nix and M. Kantarciouglu, ''Incentive compatible privacy-preservingdistributed classification,'' IEEE Trans. Dependable Secure Comput.,vol. 9, no. 4, pp. 451–462, Jul. 2012.

Author
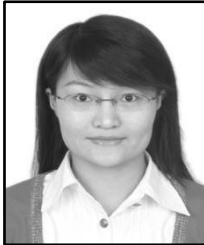
Zhao Yan. Zhao Yan received the Bachelor degree in Computer Science and Technology from Qiqihar University, and the Master degree.in Computer Application Technology from Harbin University of Science and Technology. She is currently researching on Software Technology, Mobile Internet, Big Data Application.