

A Survey of Data Center Network Topology Structure

Mingyao Zhao¹, Zhijie Han³, Xiaoyu Du^{1,2*}

¹ School of Computer and Information Engineering, Henan University, Henan, China

² Henan Province Engineering Research Center of Spatial Information Processing, Henan University, Henan, China

³ School of Software, Henan University, Henan, China

zhaomingyao1997@163.com, hanzhijie@126.com, dxy@henu.edu.cn

Abstract—In the new application mode, data center network (DCN) is faced with many challenges. Firstly, this paper introduces the requirements of modern data center network, such as network scalability, network fault tolerance, network bandwidth, equipment overhead, etc. Then it introduces the characteristics of the traditional data center network architecture and divides the data center network into two types: switch-centered and server-centered. Then, the representative structures of the two types of structures are reviewed in detail, and the above requirements are compared. Finally, a summary of the paper is made and the future development direction is pointed out.

Keywords—Data center network; Structural features; Topology; Scalability; Fault tolerance

I INTRODUCTION

Data center network is a network within the data center that connects a large number of servers through high-speed links and switches. Data center is the center of data transmission, computing and storage, and it concentrates various hardware and software resources and key business systems. Traditional data center networks are usually tree-like in structure [1]. The total bandwidth of such networks is limited by the available bandwidth of the root nodes, which become the performance bottleneck of the system and require expensive high-end switches to build as the network grows in size. In recent years, with the extensive application of virtual simulation technology and the development of cloud computing and other emerging application modes, the data center network has been developing rapidly, and the data center network is undergoing profound changes in the aspects of composition, structure, function, scale and application mode [2][3]. A modern data center can often be tens of thousands or even millions of servers. In order to adapt to the new requirements of data center network development, many new network structures have been proposed successively, such as Fat-Tree [4], Clos [5], DCell [6], BCube [7], FiConn [8], etc. Different from the traditional data center network, these new networks generally have some new features, such as providing multi-path connection and sharing network resources among multiple users. It is of great significance to study and analyze the characteristics of data center network to improve the performance of data center network and to design a reasonable resource management scheme and system [9].

This paper analyzes and compares data center network requirements, data center network classification and various data center network structure performance analysis, so as to comprehensively understand the characteristics of data center network.

II DATA CENTER NETWORK REQUIREMENTS

As data centers gather a large number of services, higher requirements are put forward for data center network, and the performance of data center network will directly affect the operation effect of data center. Compared with traditional network, data center network has higher requirements in the following aspects:

1) Network scalability: With the growth of data center business, the scale of data center will continue to expand, because this requires the network can accommodate more routers and switches, and easy to expand, the addition of equipment can not affect the performance of existing services.

2) Network fault tolerance: The topology of data center network requires to ensure that there are multiple parallel paths between different nodes, reduce single point failure of link, make the network have better fault-tolerant performance, and ensure service quality.

3) Network bandwidth: Due to the huge amount of data transmission, the data center requires the network to have high bandwidth to meet the demand of high throughput transmission. In addition, when there is a large number of sudden business, the network needs to automatically achieve business diversion.

4) The equipment cost: The scale of the data center network is large, so the construction of the data center needs to fully consider the cost issue.

III DATA CENTER NETWORK CLASSIFICATION

A. The switch-centric data center network

The switch-centered data center network is one in which switches undertake all routing addressing and forwarding functions in the whole data center.

1) Tree structure

At present, most of the data center is according to many of the root tree structure to deploy, tree structure is the most typical switch centered network structure, it is usually

divided into two levels or three levels to deploy, in more traditional root tree structure, servers are connected to the switches on the edge level, boundary level switches to connect to the aggregation level switch or directly connected to the core level switches, the aggregation level switches connected to several core level switches, then through the core level switches connected to the Internet. The figure below shows a traditional network of data centers with three levels and multiple trees. The structure of multiple root trees is characterized by clear hierarchy and easy deployment, but it has a fatal disadvantage: the entire network is very dependent on the switches in the core level, and once a switch in the core level fails, thousands of servers will be affected, Therefore, the data center network with traditional multi tree structure must use large high-end commercial switches as the switches of the aggregation level and the core level to ensure the stable operation of the entire network as much as possible, which makes the cost of the entire network expensive and complex. The tree structure as shown in Figure 1.

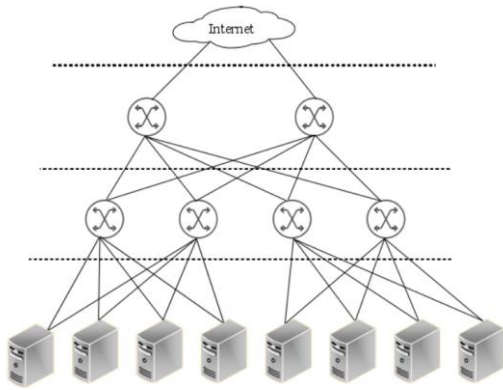


Figure 1. Tree structure

2) Fat-Tree structure

The Fat-Tree [4] structure is a typical improvement of tree structure. In fat tree structure, the network topology is still divided into three levels, namely core level, junction level and access level. However, different from the tree structure, the fat tree structure stipulates that an intermediate node can have multiple parent nodes, thus increasing the number of links between the upper and lower switches and between the aggregation level switch and the core switch. In addition, switches at the sink level and access level are divided into several different domains, and full connection can be realized between switch devices at different levels in each domain. The topology of fat tree uses two routing tables for two-level routing, and each switch has fixed coding rules. From the Figure 2 , Fat-Tree structure will be divided into k aggregation level switches and edge switches a pod, each within the pod is $k/2$ level aggregation switches and $k/2$ level edge switches, each switch port number is k , each level edge switches $k/2$ a switches directly connected with the aggregation level, the aggregation level of each switch $k/2$ a switch directly connected with core level, the benefits of such connections are greatly enhance the robustness of the entire network, regardless of any of the aggregation level and core level switch malfunctioned, There are other paths you can take. However, although Fat-Tree structure provides

multiple parallel paths to choose from, the characteristics of the Tree structure determine that the communication efficiency of the whole network will be affected if the traffic of multiple paths in the network is not balanced, which indicates that Fat-Tree structure cannot well support multicast communication. The Fat-Tree structure as shown in Figure 2.

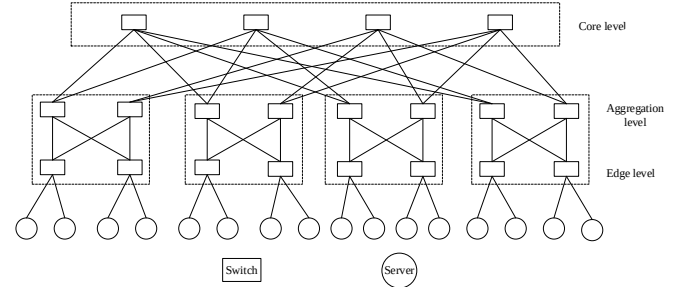


Figure 2. Fat-Tree structure

3) Clos structure

Since the traditional shortcomings of root tree structure, researchers have been trying to transform, one of which is Clos [5] network a kind of representative, it is on the basis of traditional multiple root tree structure increased the number of core level switches, make the next level of a level switches and switch are connected in order to increase the robustness of the entire network. The Clos structure as shown in Figure 3.

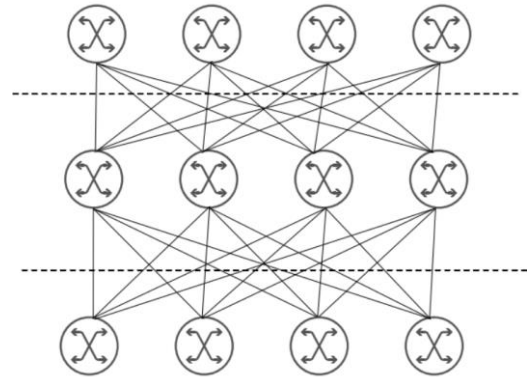


Figure 3. Clos structure

B. The server-centric data center network

In the server-centered data center network, the server undertakes the main interconnection and routing functions, while the switch only undertakes the simple vertical and horizontal switching functions. The advantage of this is that all switches in the network can use the cheap commodity switch, which reduces the construction cost of the network in the source.

1) DCell structure

The DCell [6] structure is a server-centered topology, in which the server has the same data forwarding function as the switch. DCell topology builds large-scale networks recursively through low-end branch switches and multi-port servers. In DCell topology, there are two link connection modes, that is, server is connected to server, and switch is connected to server. There is no interconnection between switches. DCell₀ is the basic unit for building a DCell

topology, with each cell acting as a node and acting as the basic unit of the next level structure, which ensures that the connections at each level are a complete graph. K represents the number of levels of DCell topology, and N represents the number of interfaces of switches in DCell₀. A very small K and N can accommodate a large number of servers, thus ensuring the high scalability of the network. DCell is composed of 5 DCell₀ networks through full connection, and the number of DCell₀ network servers in the figure is 4, usually in the number of servers in DCell₀ network is no more than 8, so only the 8-port commodity switch with single-port bandwidth between 1Gbps and 10Gbps can meet the needs of the whole network construction in DCell network. The DCell structure as shown in Figure 4.

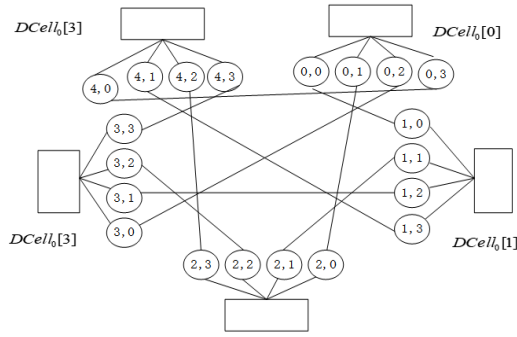


Figure 4. DCell structure

2) BCube structure

BCube [7] is a modular version of DCell that is connected in a hierarchy other than BCube₀ by a micro switch. It differs from DCell in the number of cells per level. If DCell has N DCell _{$k-1$} in the k level, then there are $n+1$ DCell _{k} in the $k+1$ level. While BCube has the same number of units n at any level, it is easy to conclude that BCube _{k} has $nk+1$ servers (n is the number of servers in BCube₀). The BCube structure as shown in Figure 5. The BCube structure ensures that the switch is only connected to the server, and the switch will never be directly connected to each other, which enables us to regard the switch in the BCube network as a vertical and horizontal virtual node connecting to the adjacent servers, and the transfer of traffic is mainly done by the server. So if use 8-port small switch to build the network, can accommodate 4096 servers in a BCube network. Therefore, BCube network achieves the goal of using low-end commodity switch to build the data center by placing routing function on the server.

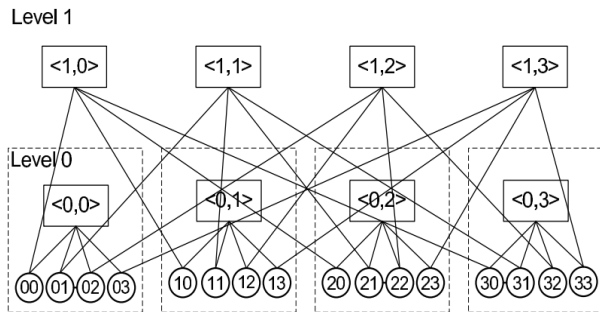


Figure 5. BCube₁ structure

3) FiConn structure

FiConn [8] is also a recursive definition of network structure, it and DCell design standards are very similar, but they differ in that FiConn structure within the server only two ports, and structure DCell _{k} server port number is $k+1$, that means that FiConn structure node degree is always in 2, and FiConn link cost is smaller than DCell. The FiConn topology, in which switching devices and servers are capable of forwarding, was also proposed by Microsoft staff. It is also a server-centric topology. Modern commercial server devices typically have two Ethernet ports, one for network connections and one as a standby port. The core of FiConn topology is to realize network interconnection by using standby server ports, and to eliminate large-scale high-performance switching equipment while ensuring good network performance, thus saving interconnection cost. FiConn adopts a recursive approach, and the high-level FiConn structure is constructed from the low-level FiConn structure. The low-level FiConn uses half of the available standby terminals to achieve interconnection and form a mesh structure. The FiConn structure as shown in Figure 6.

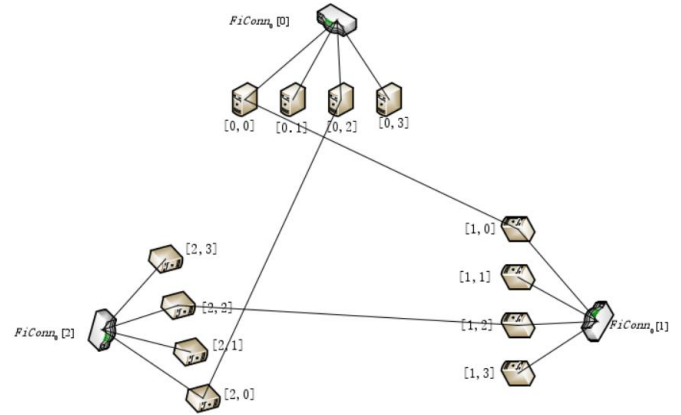


Figure 6. FiConn structure

4) Snow structure

Snow structure [10] also is a kind of network structure of recursive definition, it with a level 0 Snow Snow₀ as the basic unit for network extension, a level 0 Snow Snow₀ is composed of n servers connected to a small switches, and according to the data center of the relevant literature, n values are usually between 3 ~ 8, figure 7 is a Snow₀ structure diagram, figure in the server is not real link between the dotted line, but to expand structure and definition of a virtual connection. When expanding from Snow₀ to Snow₁, first break the three virtual connections of Snow₀ and connect one Snow₀ respectively. When connected to the newly added Snow₀ of the switches and the original Snow₀ connected to the server, then the connection becomes a real connection, we call this the virtual link into a real link connection for the real connection, real connection not only emphasizes the real existence of link, more important is emphasized the connection state transition from virtual to real, so Snow₀ inside the connection between the server and the switch is not called it a real connection, Figure 8 is a structure diagram of Snow₁.

Figure 8 shows that instead of a full $Snow_0$ being added to the broken virtual connection, there is a missing virtual connection $Snow_0$, which is defined as a Cell to distinguish it from $Snow_0$. There are 6 virtual connections and 6 real connections in $Snow_1$. When expanding from $Snow_1$ to $Snow_2$, break all virtual connections and real connections and add a Cell at the break point, and so on. When expanding from $Snow_{k-1}$ to $Snow_k$, it is necessary to break all virtual connections and real connections in $Snow_{k-1}$ and add a Cell at the break point.

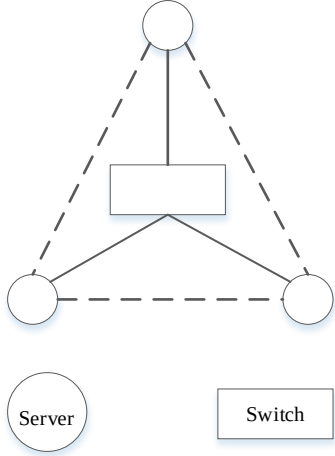


Figure 7. $Snow_0$ structure

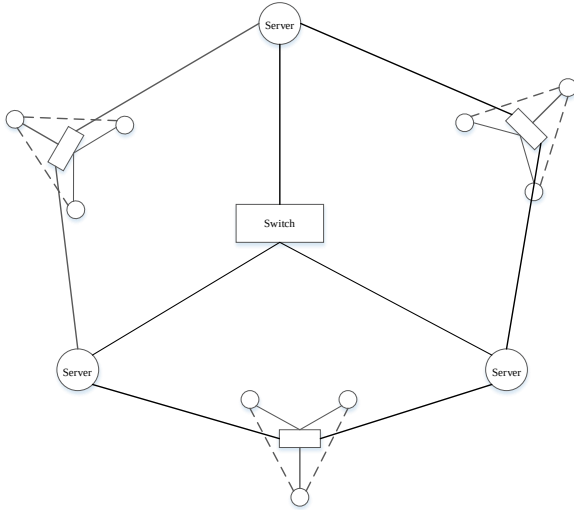


Figure 8. $Snow_1$ structure

5) ExCCC-DCN structure

ExCCC-DCN [11] is for Exchanged Hyper Cube network. The ExCCC(Exchanged Cube-Connected Cycles) network is a Exchanged topology structure proposed in 2017. It is for Exchanged hypercube EH (Exchanged Hyper Cube). EH structure is a topological structure proposed in 2005, and EH structure is a deformation of Hypercube. According to the relevant knowledge of graph theory, a physical network can be abstracted into a graph for representation. The rule is that a point is used to represent a node in the network and an edge is used to represent a link in the network. Considering that communication in the network is always bidirectional, undirected graph is generally used to represent the structure

of the network. Then EH can be defined as $EH(s, t) = \{(V, E) \mid s \geq 1, t \geq 1\}$, where s and t represent the scale of EH network, V represents the set of nodes, and E represents the set of edges. Figure 9 is the structure diagram of an EH (1,2).

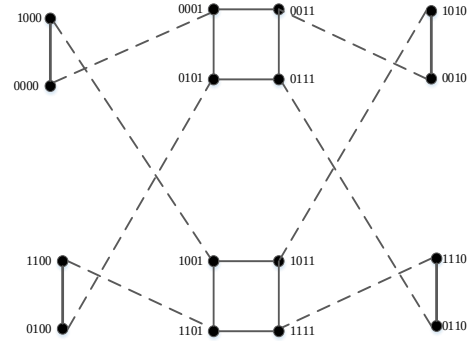


Figure 9. EH(1,2) structure

The rule of EH expansion into ExCCC is to replace each point in EH with a ring, thus greatly increasing the capacity of the entire network. Similarly, ExCCC network can also be defined by an undirected graph, but since each node in $EH(s, T)$ is replaced by a ring when the EH network is converted into ExCCC network, for an $s + t$ dimensional EH network, its corresponding ExCCC network is an $s + t + 1$ dimensional network. From the definition of nodes, it can be seen that in Ex CCC network, the address of each node is composed of three parts, and the corresponding length of each part is respectively s , t and 1. Moreover, because the address of the node is represented by binary number, each bit of node address has only two values, namely 0 and 1.

The E set of edges in ExCCC contains three types of edges, which are cycle-edge, Cube-Edge and Exchanged edge respectively. Figure 10 is a structural diagram of ExCCC(1,2). The side represented by the thin solid line in the figure is for cycle-edge, the side represented by the thick solid line is for cube-edge, and the side represented by the dashed line is for exchanged-edge.

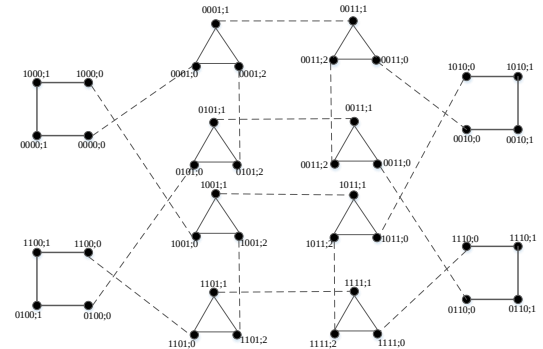


Figure 10. ExCCC (1,2) structure

6) BCCC structure

The network structure is a recursively defined, new server centric network topology. The remarkable advantages of the structure are good scalability, small network diameter, and high server port utilization [12]-[15]. BCCC is a recursively defined structure composed of switches and servers. First, connecting n servers to a switch with n port is called an

element. A BCCC (n, 0) is connected to n switches through one element, and a BCCC (n, k) is connected to n^k elements through n BCCC (n, k-1). First, need to index n BCCC (n, k-1) from 0 to $n-1$. Each server in the structure is represented as $a_{k+1}a_k a_{k-1} \dots a_0$, $a_0 \in [0, k-1]$, $a_i \in [0, n-1]$, $1 \leq i \leq k+1$. In addition to n BCCC (n, k-1), n^k elements required to build BCCC (n, k) contain n^{k+1} servers. These servers can also be represented as $a_{k+1}a_k a_{k-1} \dots a_0$, $a_i \in [0, n-1]$, $1 \leq i \leq k+1$. According to the node code, $A = a_{k+1}a_k a_{k-1} \dots a_0$ and $A' = a'_{k+1}a'_k a'_{k-1} \dots a'_0$ belong to the same element and $a_{k+1} \neq a'_{k+1}$ and $a_i = a'_i$, $\forall i$, $0 \leq i \leq k$. In general, build a BCCC (n, k), which requires $(k+1)n^{k+1}$ dual port servers, $(k+1)n^k$ n-port switches, and n^{k+1} (k+1) switches. The BCCC structure as shown in Figure 11.

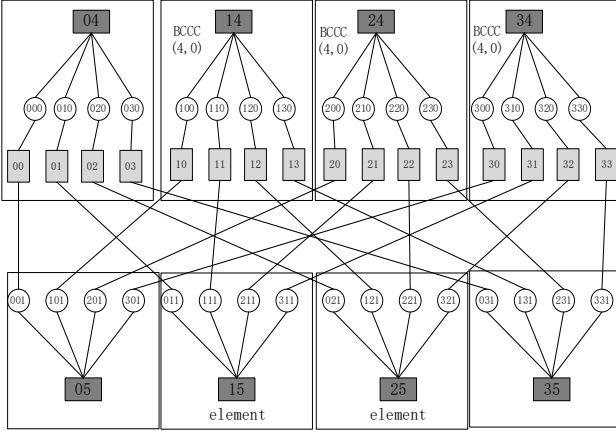


Figure 11. BCCC structure

IV PERFORMANCE ANALYSIS OF VARIOUS DATA CENTER NETWORK STRUCTURES

A. Network scalability:

The Tree topology structure adopts vertical way to achieve expansion, network expansion capacity is limited by the number of ports of high-level network equipment, expansion capacity is limited; The topology of fat tree can support more servers by means of horizontal expansion, so as to meet the expansion requirements of data center. DCell adopts recursive method, which can support large-scale servers and is convenient for expansion. BCube also adopts recursive method, which can carry out modular construction and has good network expansibility [16]. FiConn also uses recursive method to construct the network topology, and the network scalability is good. The snowflake structure also adopts recursive topology construction, which can reach the number of millions of servers in the later period and can support the construction of large and expensive servers. It can be seen that the snowflake structure's network scalability is very good. The number of Cycle-Edge is the largest among all the edges of ExCCC-DCN, so it can be seen that ExCCC-DCN is very suitable for building a large-scale data center network. Moreover, in the whole ExCCC-DCN, the port

number of each switch is a relatively small constant, and each server only needs two ports, so the scalability is very good.

B. Network fault tolerance:

The Tree structure network has poor fault-tolerant performance. When the network link or node fails, it is easy to separate the network into independent subnets, resulting in network paralysis and deterioration. The fat tree network at all levels has multiple redundant transmission paths and no single point of failure, and has good fault tolerance performance. However, the internal fault tolerance performance of Pod is poor, and it is more sensitive to the fault of the underlying switching equipment [17]. DCell adopts recursive definition, decentralization and complete interconnection, etc., which can support the fault tolerance of routing better. The connection mode and recursive structure of BCube enable the modular construction of data center and good implementation. BCube performs better than DCell in the case of incomplete structure and has the best fault-tolerant performance. The FiConn network has multiple parallel links between different servers, so the fault tolerance of the network is very good [18]. The connection mode and recursive structure of the snowflake structure enable the data center to be modular, with multiple redundant transmission paths, and the routing can easily calculate the path size and length, with strong fault tolerance. ExCCC-DCN has very low construction cost and energy consumption.

C. Network bandwidth:

The traffic distribution of tree topology is not uniform, and the core nodes are easy to become the bottleneck of network performance. With the expansion of network scale, the fat tree structure has higher throughput and can achieve load balancing between multiple links in the core level to avoid network performance bottleneck. DCell using recursive definition, decentralization, fully connected, etc., can well support routing fault tolerance, and provide better than tree structure and fat tree topology aggregation of bandwidth, but the DCell topological defects is uneven distribution of flow, the bottom flow rate on the link on the high side, hierarchical full connection and server implementation forward capability will affect throughput and network latency [19],[20]. BCube and FiConn network has a large bandwidth and can meet the demand of high throughput transmission of data. The traffic distribution of snowflake structure is uneven, and the traffic of snowflake structure at the lower level is higher, which will affect the throughput and network delay. ExCCC-DCN has better fault tolerance.

D. The equipment cost:

The high performance requirement of treetop topology device results in high device overhead. The network equipment of fat tree topology is mostly commercial equipment, so the cost of network equipment can be greatly reduced. DCell networks replace high performance switches with long chains, which leads to increased link overhead. FiConn makes full use of the standby port of the server and saves some device overhead. The advantage of snowflake structure is that the micro-switch replaces the expensive

switch, so the cost is low. ExCCC-DCN has very low construction cost and energy consumption [21-23].

E. The managing complexity:

The traditional tree topology structure is simple and easy to implement. Topology rules and symmetry of fat tree structure are beneficial to network cabling and automatic configuration. The interconnection between each level of DCell topology makes the routing complexity too high, which is not conducive to automatic configuration, management and engineering implementation. DCell network replaces high performance switches with long links and increases link overhead. There are many switches in BCube, so the cost of routing layout and deployment in BCube is relatively high. The complexity of FiConn topology wiring is high. If the underlying server and switch fail, the level of the snowflake structure may directly reduce one level, which will reduce the number of servers by a lot. The underlying switch and server are relatively difficult to manage, with high management complexity and relatively high cost. However, ExCCC-DCN has a complex structure, and its wired link deployment mode also has some shortcomings.

V CONCLUSION

This paper first introduces the status quo of data center network, then introduces the requirements of modern data center network from several aspects, then introduces the basic characteristics of different types of data center network, and compares the structure from different aspects. Through analysis, it is found that the two types of schemes have their own advantages and disadvantages. It is hoped that subsequent researchers can propose more new structures with high fault tolerance and scalability to promote the development of data center networks. This paper is of great learning significance for the people who are first exposed to the data center network.

ACKNOWLEDGMENT

This work was supported by Special project for key R & D and promotion of Henan Province (222102210052, 222102210007, 222102210062, 222102210272, 212102210079).

REFERENCES

- [1] Muhammad M, Alyas T, Ahmad F, et al. An analysis of security challenges and their perspective solutions for cloud computing and IoT[J]. EAI Endorsed Transactions on Scalable Information Systems, 2020, 8(30).
- [2] Cao Z, Chen Z, Zheng Y. Express Delivery Handling Model Based on Monte Carlo Simulation and Linear Programming in the Background of Cloud Computing[C]//Journal of Physics: Conference Series. IOP Publishing, 2021, 1915(4): 042088.
- [3] Zhao D. Status of Cloud Computing in Data Intensive Data Center[C]//Journal of Physics: Conference Series. IOP Publishing, 2021, 1881(3): 032089.
- [4] Guo Z, Yang Y. Multicast Fat-Tree Data Center Networks with Bounded Link Oversubscription[C]. Infocom, IEEE, 2013.
- [5] Ganesh L, Ganesh L, Ganesh L, et al. Dynamo: facebook's data center-wide power management system[C]. Acm, International Symposium on Computer Architecture. 2016:469-480. I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, p. 271-350.
- [6] Guo C. DCell: A scalable and fault-tolerant network structure for data centers[J]. ACM SIGCOMM Computer Communication Review, 2008, 38(4):75-86.
- [7] Guo C, Lu G, Li D, et al. BCube: A High Performance, Server-centric Network Architecture for Modular Data Centers[J]. Acm Sigcomm Computer Communication Review, 2009, 39(4):63-74.
- [8] Li D, Guo C, Wu H, et al. FiConn: Using Backup Port for Server Interconnection in Data Centers[C]. IEEE INFOCOM 2009. IEEE, 2009M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.
- [9] Guo C, Lu G, Wang H J, et al. Secondnet: a data center network virtualization architecture with bandwidth guarantees[C]//Proceedings of the 6th International Conference. 2010: 1-12.
- [10] Xiaoqian Liu, Shoubao Yang, Liangmin Guo, et al. Snow structure: A new data center network structure[J]. Journal of Computer Science, 2011, 34(1):76-86.
- [11] Zhang Z, Deng Y, Min G, et al. ExCCC-DCN: A highly scalable, cost-effective and energy-efficient data center structure[J]. IEEE Transactions on Parallel and Distributed Systems, 2016, 28(4): 1046-1060.
- [12] Li Z, Guo Z, Yang Y. BCCC: An expandable network for data centers[C]//Proceedings of the tenth ACM/IEEE symposium on Architectures for networking and communications systems. 2014: 77-88.
- [13] Huang L, Jia Q, Wang X, et al. Pcube: Improving power efficiency in data center networks[C]//2011 IEEE 4th International Conference on Cloud Computing. IEEE, 2011: 65-72.
- [14] Li Z, Yang Y. Permutation generation for routing in BCube connected crossbars[C]//2015 IEEE International Conference on Communications (ICC). IEEE, 2015: 5460-5465.
- [15] Li D, Li Y, Wu J, et al. ESM: Efficient and scalable data center multicast routing[J]. IEEE/ACM Transactions on Networking, 2011, 20(3): 944-955.
- [16] Greenberg A, Lahiri P, Maltz D A, et al. Towards a next generation data center architecture: scalability and commoditization[C]//Proceedings of the ACM workshop on Programmable routers for extensible services of tomorrow. 2008: 57-62.
- [17] Ma X, Hu C, Chen K, et al. Error tolerant address configuration for data center networks with malfunctioning devices[C]//2012 IEEE 32nd International Conference on Distributed Computing Systems. IEEE, 2012: 708-717.
- [18] Li D, Guo C, Wu H, et al. FiConn: Using backup port for server interconnection in data centers[C]//IEEE INFOCOM 2009. IEEE, 2009: 2276-2285.
- [19] Raiciu C, Barre S, Pluntke C, et al. Improving datacenter performance and robustness with multipath TCP[J]. ACM SIGCOMM Computer Communication Review, 2011, 41(4): 266-277.
- [20] Chen Y, Jain S, Adhikari V K, et al. A first look at inter-data center traffic characteristics via yahoo! datasets[C]//2011 Proceedings IEEE INFOCOM. IEEE, 2011: 1620-1628.
- [21] Guo C, Wu H, Tan K, et al. Dcell: a scalable and fault-tolerant network structure for data centers[C]//Proceedings of the ACM SIGCOMM 2008 conference on Data communication. 2008: 75-86.
- [22] Al-Fares M, Loukissas A, Vahdat A. A scalable, commodity data center network architecture[J]. ACM SIGCOMM computer communication review, 2008, 38(4): 63-74.
- [23] Kandula S, Sengupta S, Greenberg A, et al. The nature of data center traffic: measurements & analysis[C]//Proceedings of the 9th ACM SIGCOMM conference on Internet measurement. 2009: 202-208.



Mingyao Zhao received her master's degree in Computer Application Technology from Henan University, Kaifeng Henan. Her main research directions is data center network.



Zhijie Han received the M.S. and Ph.D. degrees computer science from Henan University and Soochow University in December, in 2006 and 2009, respectively. He is currently a vice professor in the school of computer and information engineering, Henan University. His research interests include parallel and distributed computing, cloud computing, and big data.



Xiaoyu Du received her master degree in Applied Mathematics from Henan University, Henan Kaifeng, and PhD degree in Information Network from Nanjing University of Post and Telecommunication, Jiangsu, Nanjing. She is currently working as a teacher in the Henan University. Her main research interests are data center network structure, multicast and wireless sensor networks localization and coverage.