HBNA: A New Type of Data Center Network Topology

Rui Wang

Department of Computer Science, Jinan University, Guangzhou, 510632, China.

Abstract

With the popularization of cloud computing, the amount of tasks processed by data centers has increased dramatically in recent years, and the situation of sudden loads has become increasingly serious. Once the sudden load of the data center at a certain moment exceeds the processing threshold of the data center, it will cause a sharp increase in service processing delay and reduce the user's service quality. Therefore, how to quickly increase the number of data center servers in response to predictable sudden loads has become a challenging problem. This paper proposes a new data center network architecture based on HyperCube network, called HBNA. HBNA is constructed by using m-port switches and single-port servers. The HBNA architecture has excellent expansion performance, allowing the data center to expand rapidly and gradually when needed. Simulation experiments and performance results show that the throughput of HBNA is comparable to Fat-Tree, BCube and DCell. In addition, the analysis results show that, compared with the latest data center network architecture, HBNA achieves a good balance between diameter, equal width, incremental scalability, cost and energy consumption.

Keywords

Data Center; Bursts; Network Topology.

1. Introduction

With the explosive growth of data volume, data centers have gradually become the core of cloud computing. The construction of a data center network needs to consider the network structure and related network protocols at the same time, so as to connect thousands or even hundreds of thousands of servers, storage devices and network devices in a network. At the same time, the construction of the data center network should consume lower equipment costs, and provide a higher and balanced network capacity, a structure that is easy to follow-up expansion, and a strong fault tolerance.

The data center network (Data Center Network, DCN) architecture can be divided into two categories: server-centric and switch-centric[1]. In a server-centric design, the server acts as both a server and a network repeater. DCell, BCube and FiConn structures fall into this category. The switch-centric network structure usually consists of a multi-layer switch tree connected to the server, and the Fat-Tree structure belongs to this category.

FiConn, DCell and BCube structures are recursively defined data center network structures, indicating that the structure of dimension n is constructed by the structure of dimension (n-1). In these structures, the number of servers increases significantly as the dimensionality increases. The recursively defined network structure cannot achieve fine-grained expansion, so servers cannot be gradually added to the data center network according to requirements, and the incremental scalability is poor.

In order to achieve the higher progressive scalability of the data center network structure, many scholars[2-4] have studied incomplete HyperCubes. The n-dimensional hypercube is constructed from two (n-1)-dimensional HyperCubes. A complete n-dimensional hypercube can only connect 2n vertices accurately, which greatly limits the progressive scalability of the structure. Therefore, many studies have proposed a variant hypercube topology with greater flexibility, which is called an incomplete hypercube. An incomplete hypercube is composed of several smaller HyperCubes. An incomplete hypercube can contain any number of vertices, and its main topological properties are similar to those of a complete hypercube. In addition, incomplete HyperCubes can be gradually expanded by adding HyperCubes of different dimensions.

In this chapter, we propose a new data center network architecture HBNA (HyperCube Based Network Architecture) based on HyperCube. HBNA is constructed by multi-port commercial switches and single-port servers. Because single-port NICs and commercial switches are inexpensive, the construction cost of HBNA is extremely low. The diameter of the HBNA structure is n+1, and the divided width is N/2n, where N is the number of servers and n is the dimension of the structure. This means that the HBNA structure has a smaller diameter and a higher bisecting width. In addition, in order to achieve higher progressive scalability, this chapter introduces three incomplete HBNA structures in detail. The incomplete HBNA structure allows the data center to add the number of servers as needed without changing its network topology attributes. The data center network structure has a high gradual scalability, which allows the data center to more easily increase the number of servers when responding to predictable sudden loads, so as to better provide users with services.

2. Relate Work

In this section, we will introduce four common data center network architectures, including Fat-Tree, DCell, FiConn and BCube. All these architectures are constructed through commercial switches, and many scholars have studied them. In all these four DCN (Data communication network) architectures, Fat-Tree is switch-centric, and the other three architectures are server-centric.

2.1 Fat-tree structure

Fat-tree[5] is a hierarchy-based architecture. As shown in Figure 1, the switch in Fat-tree is divided into three levels, including the edge layer, the convergence layer, and the core layer. The Fat-tree architecture constructed by n-port switches has n Pods. Each Pod contains n/2 switches at the edge layer and the aggregation layer. At the edge layer, each switch uses n/2 ports to connect to n/2 servers, and uses the remaining n/2 ports to connect to n/2 switches at the aggregation layer. There are (n/2)^2 switches in the core layer, and each switch is connected to a Pod through a port. Therefore, the Fat-Tree architecture consists of (n/2)^2+n^2 switches, which supports (n/2)^2*n servers in total. The scalability of Fat-Tree is limited by the number of switch ports. Therefore, if you need to expand Fat-Tree and make full use of existing switches, you must replace the existing switches with more ports.

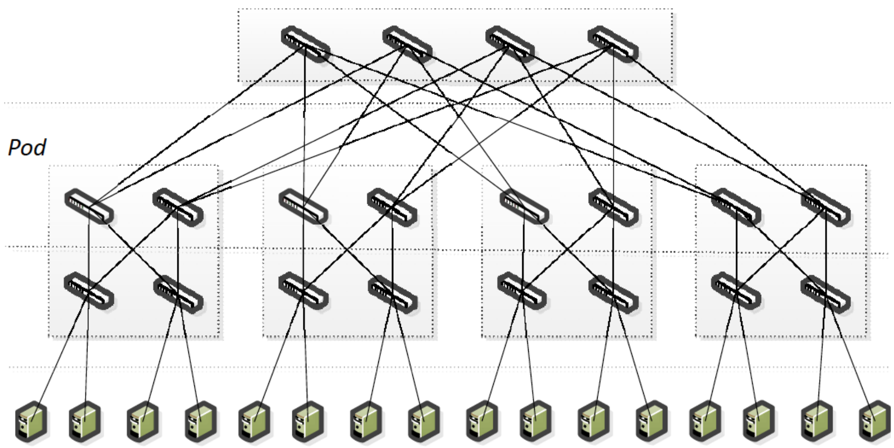

Fig. 1 Fat-tree structure with 4-port switch

2.2 DCell structure

DCell[6] structure is defined in a recursive way. $DCell_0$ contains n servers and a commercial n-port switch, which is only used to connect servers. DCell1 includes n+1 DCell0. Similarly, DCell2 includes (n+1)×n+1 DCell1.

If DCellk has n servers, DCellk+1 can be constructed by using n+1 DCellk. The t servers in DCellk are respectively connected to other t DCellk. Figure 2 shows the topology of DCell1 with n=4. The DCell architecture is highly scalable and can be extended to a very large scale. However, the scalability of the DCell is limited by the server port, which means that if some new switches and servers are added to the DCell, some other ports must be added to the server to establish a connection. In addition, the progressive scalability of DCell is very poor. Once the DCell architecture is completed, it is difficult to add new servers to the architecture without changing the original architecture.

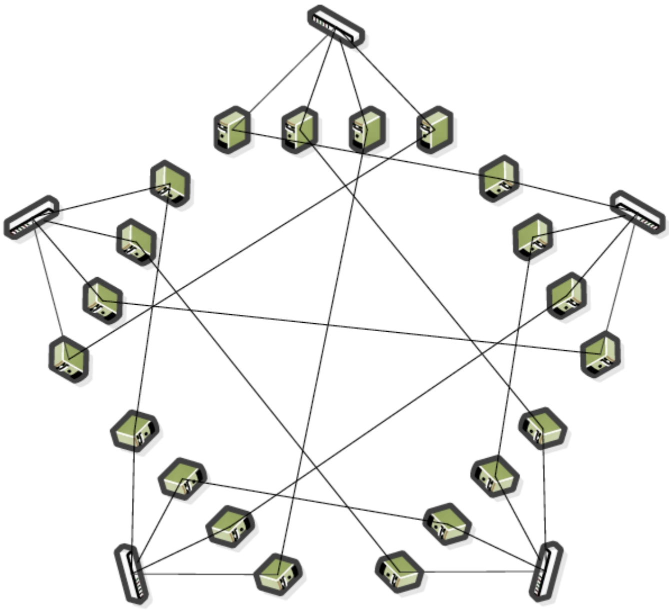

Fig. 2 DCell1 structure with 4-port switch

2.3 FiConn structure

FiConn[7] is constructed using n-port switches and 2-port servers, where n is an even number. Similar to DCell, the FiConn network topology can also be constructed in a recursive manner. In FiConn0, each server is connected to the switch through a port. If $FiConn_{k-1}$ contains b spare ports, then b/2+1 $FiConn_{k-1}$ can be used to construct the FiConnk topology. In the FiConnk structure, each FiConnk-1 connects to other b/2 FiConnk-1 through b/2 servers with available backup ports. Figure 3 shows the FiConn1 structure with n=4. The FiConn structure has good scalability, and there is no need to add other server ports or switch ports during the expansion process. However, its progressive scalability is poor. In order to solve this problem, some scholars have proposed an incomplete FiConn. By using a small amount of complete FiConnk-1 to construct an incomplete FiConnk, and FiConnk-1 is fully connected, this has led to The bisecting width of the incomplete FiConn is very small.

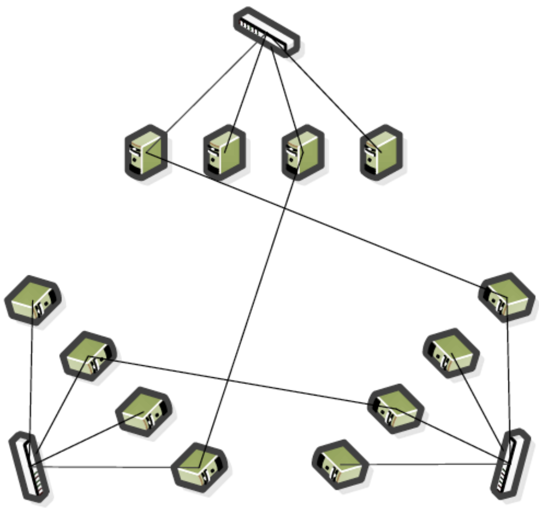

Fig. 3 FiConn1 structure with 4-port switch

2.4 BCube structure

The BCube system structure[8] is also defined in a recursive manner. BCube0 is composed of n servers and an n-port switch. In general, BCubek consists of nk n-port switches and n BCubek-1 units. Each server in BCubek has k+1 ports. Figure 4 illustrates BCube1 with n=4. The BCube topology is suitable for building data centers containing up to thousands of servers. However, building a data center based on the BCube topology will incur high switch and wiring costs.

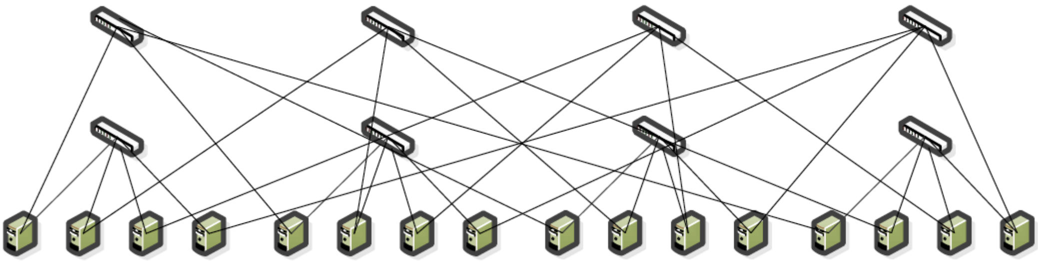

Fig. 4 BCube1 structure with 4-port switch

Due to the characteristics of each structure, different structures have their own advantages and corresponding disadvantages. Enterprises will face the dilemma of cost, performance, and scalability when building data centers. Therefore, it is necessary to improve the existing network architecture.

3. Literature References

In order to make the data center have a better performance in response to the predicted sudden load, we first need to solve the problem of gradual scalability of the data center. The HBNA (HyperCube Based Network Architecture) structure designed in this paper is mainly composed of multi-port commercial switches and single-port servers. This structure can build a large-scale and scalable data center network. It is also economical because of the use of commercial switches and single-port servers. Very impressive. The HBNA structure is based on HyperCube. The structure diagram is shown in Figure 5.

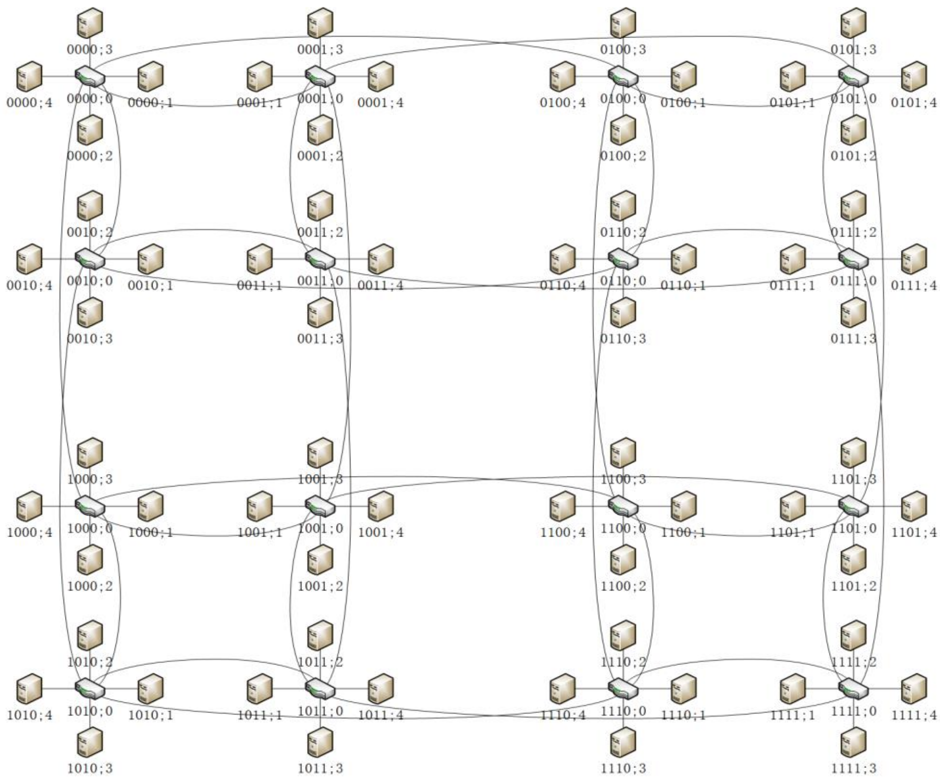

Fig. 5 Schematic diagram of HBNA4 structure

Structure definition: 1) The structure is constructed based on HyperCube; 2) The definition of n-dimensional HBNA points and edges is shown in Table 1.

Table 1. Definition of $HBNA_n$ structure points and edges

	Definition statement
Point definition	$(x_{n-1} \cdots x_0)$ $x_i \in \{0,1\},\ 0 \le i < n, y \in [1,n]$
Switch Point	$(x_{n-1} \cdots x_0; 0)$
Server Point	$(x_{n-1} \cdots x_0; y)$
Edge definition	$\left((x_{n-1} \cdots x_0),(x_{n-1} \cdots x_{y+1}\overline{x_y}x_{y-1} \cdots x_0)\right)$ $x_i \in \{0,1\}, 0 \le y < n$
Switch and server	$\left((x_{n-1} \cdots x_0; 0),(x_{n-1} \cdots x_0; y)\right)$
Switch and switch	$\left((x_{n-1} \cdots x_0; 0),(x_{n-1} \cdots \overline{x_i} \cdots x_0; 0)\right)(i \in [0, n-1])$

In Table 2, we compare and analyze the Fat-tree, DCell, FiConn, and BCube structures of the new data center topologies that are currently studied in the HBNA structure. For the convenience of expression in the table, we use m to represent the number of servers in the data center, k to represent the number of switches used, and n to represent the dimension of the topology. Diameter is a measure of data center network communication delay. The smaller the diameter, the lower the data center communication delay. Since the diameter of the DCell and FiConn structures is related to the actual situation, what we show in the table is the upper bound of the diameter of this structure. The equalized bandwidth represents the minimum number of connections that need to be deleted when the network topology is divided into two equal parts. The larger equalized bandwidth means that the network topology has higher network capacity and stronger fault tolerance.

The construction cost of the data center network depends on the number of servers, switches and links we use. We compared the cost and energy consumption of the HBNA structure with other data center network topologies through simulation experiments. In order to make a fair comparison, we assume that the five compared structures all support the same number of servers (m) in the data center and use the same type of switches (k ports).

Table 2. Comparison of the properties of different data center topologies

	Fat-Tree	DCell	BCube	FiConn	HBNA
Diameter	$2\log_2 m$	$\le 2^{n+1}-1$	$n+1$	$\le 2^{n+1}-1$	$n+1$
Bisection Width	$\frac{m}{2}$	$\ge 2\log_k m - 1$	$\log_k m$	$\ge 2\log_k m - 1$	$\frac{m}{2k}$
Number of Switches	$\frac{5m}{k}$	$\frac{m}{k}$	$\frac{(n+1)m}{k}$	$\frac{m}{k}$	$\frac{2m}{k}$
number of ports	≤ 2	$n+1$	$n+1$	≤ 2	1
Scalability limited by server ports	No	Yes	Yes	No	No
Scalability limited by switch ports	Yes	No	No	No	Yes
Incremental Scalability	Good	Poor	/	Poor	Good

4. Throughput analysis

We use the stream-level simulator mtCloudSim[9] to evaluate the throughput of different data center network structures. Throughput is a more important performance index of the data center network structure. The data center network topology with high throughput has stronger processing ability to deal with complex services[10]. Ideally, an excellent network topology should provide the shortest possible delay and the largest possible throughput. In this experiment, the number of servers is 4096, and the link data transmission rate is 1Gbps. As shown in Figure 6, the highest throughput of the five network structures is between 250Gbps and 300Gbps, which proves that our HBNA structure is not at a disadvantage in terms of throughput. At the same time, Fat-Tree, DCell, BCube and HBNA structures can transmit all data within 150 seconds. Experimental results show that the HBNA structure achieves better performance in terms of throughput.

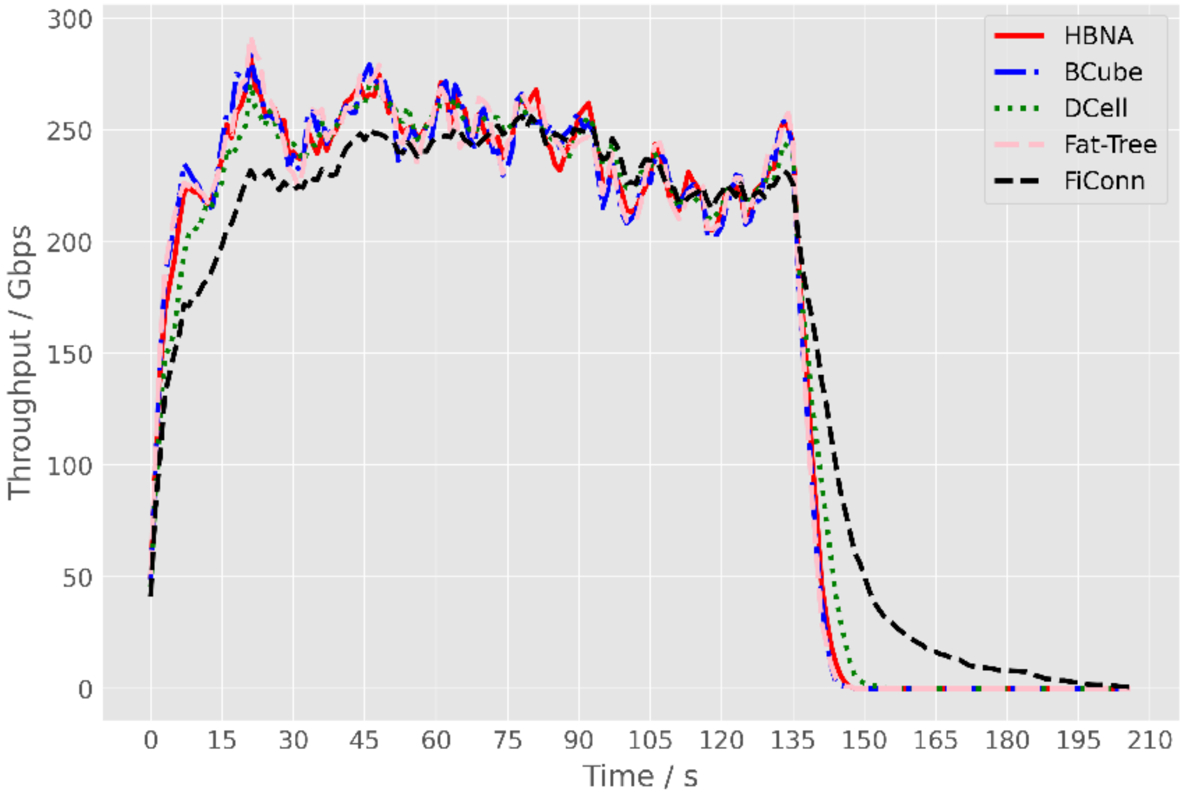

Fig. 6 Comparison of throughput between different network structures

5. Conclusion

In order to deal with predictable sudden loads, such as Taobao's Double Eleven and JD's June 18th, in this chapter, we propose a new type of data center network topology, HBNA, with high progressive scalability. The HBNA structure is constructed based on HyperCube, with excellent gradual scalability, and the number of servers can be gradually increased according to demand. At the same time, the HBNA structure uses single-port servers and multi-port commercial switches, saving a lot of hardware costs.

We compared the HBNA structure with other new data center topologies (Fat-tree, DCell, FiConn, and BCube structures) that are currently being extensively studied. The analysis results show that the HBNA structure and FiConn structure have better expansion capabilities than other structures. The experimental results show that the HBNA structure has higher throughput and lower delay like other structures, and the hardware cost of the DCell structure and the FiConn structure is greater than that of the HBNA structure. Through the analysis of the performance-price ratio, we can find that the data center network topology using HBNA structure can achieve higher cost performance than the other four structures. Therefore, the HBNA structure can become one of the best-performing topologies that can be selected for today's large-scale data center construction.

References

[1] Zhang Z, Deng Y, Min G, et al. HSDC: A highly scalable data center network architecture for greater incremental scalability[J]. IEEE Transactions on Parallel and Distributed Systems, 2018, 30(5): 1105-1119.

[2] H. P. Katseff, "Incomplete hypercubes," IEEE Transactions on Computers, vol. 37, no. 5, pp. 604–608, 1988.

[3] N.-F. Tzeng, "Structural properties of incomplete hypercube computers," in Distributed Computing Systems, 1990. Proceedings., 10th International Conference on. IEEE, 1990, pp. 262–269.

[4] Sen, A. Sengupta, and S. Bandyopadhyay, "On some topological properties of hypercube, incomplete hypercube and supercube," in Parallel Processing Symposium, 1993., Proceedings of Seventh International. IEEE, 1993, pp. 636–642.

[5] M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data center network architecture," in ACM SIGCOMM Computer Communication Review, vol. 38, no. 4. ACM, 2008, pp. 63–74.

[6] C. Guo, H. Wu, K. Tan, L. Shi, Y. Zhang, and S. Lu, "Dcell: a scalable and fault-tolerant network structure for data centers," in ACM SIGCOMM Computer Communication Review, vol. 38, no. 4. ACM, 2008, pp. 75– 86.

[7] D. Li, C. Guo, H. Wu, K. Tan, Y. Zhang, S. Lu, and J. Wu, "Scalable and cost-effective interconnection of data-center servers using dual server ports," IEEE/ACM Transactions on Networking, vol. 19, no. 1, pp. 102– 114, 2011.

[8] GUO C, LU G, LI D, et al. BCube: a high performance, server-centric network architecture for modular data centers[C]. Proceedings of the ACM SIGCOMM 2009 conference on Data communication, F, 2009.

[9] J. Xie and Y. Deng, "mtcloudsim: A flow-level network simulator for multi-tenant cloud," in Proceedings of the 22th IEEE International Conference on Parallel and Distributed Systems, 2016.

[10] XIE J, DENG Y, MIN G, et al. An incrementally scalable and cost-efficient interconnection structure for data centers [J]. IEEE Transactions on Parallel and Distributed Systems, 2016, 28(6): 1578-92.