EXTENDED DOUBLE STAR: A MASSIVE PARALLEL BIG DATA NETWORK

Sadashiba Pati[1], Dr. Vinay Singh[2], Dr. Nibedita Adhikari[3]

Article History: Received: 20.02.2022 **Revised:** 21.04.2022 **Accepted:** 05.06.2022

Abstract

The digital philosophy claims that all physical course of our environment are forms of computation or information processing at the most fundamental level. This concept gives rise to the concept of Big data. As a human nature we all need thing to be fast always. For faster processing we have to rely on parallel processing. This research proposes a new parallel computing architecture called as Extended Double Star (EDS) for implementing the big data framework. The different topological parameters are derived for EDS and compared with the parent networks. Next efficient routing and broadcasting algorithms are proposed to show the efficient message passing using the network controllers. The link complexity of EDS network is found to be least among all extended topologies. While packing the highest number of nodes the EDS network retains all original features of the base network but results in faster routing and broadcasting.

Keywords: Interconnection, Message passing, Routing, Star graph

[1]Usha Martin University, Ranchi, Jharkhand, India.
[2]Usha Martin University, Ranchi, Jharkhand, India.
[3]Utkal University, Vani Vihar, Bhubaneswar, Odisha, India.

Email: [1]sadashibapati1@gmail.com, [2]vinaysinghuma@gmail.com, [3]nibedita.cs@utkaluniversity.ac.in

DOI: 10.53555/ecb/2022.11.12.347

1. INTRODUCTION

Both data volumes and processing speeds have been on exponentially rising trajectories since the onset of digital age. Big data plays an important role to store and process huge amount of both structured and unstructured data. The big data systems collect data from various sources, whether internal and external, and optimized the data for retrieval for many purposes. Primarily big data system is designed to gather business insights and allows businesses to integrate their data, manage it, and analyze it at many levels. Here the data volume is growing exponentially with time. The size of collected data is so large and complex that none of the traditional data management tools can store it or process it efficiently. More clearly, big data needs efficient and reliable platforms to deal with ever growing variety, volume of data that exists with more velocity. Message passing or communication between the data gathering and storing nodes is also becoming very crucial. Researchers agree at one point that the big data systems need massive computation and faster communication [1, 2, 4 & 5]. Big data systems contain a central data lake to store and process large amounts of complex structure and unstructured data. There is a need to provide a foundation layer to create opportunities for service providers to enhance quality of service for the end user. Here, quality of service refers to easy and faster data access Big data requires big performance on a network infrastructure which means the network needs to be resilient, consistent, and have some form of application awareness. Ideal big data network architecture must be designed with a distributed architecture in mind in order to deal with the availability of distributed resources all the time while simultaneously working in parallel [17]. Discrepancy between the explosive growth rate in data volume and the improvement trends in processing and memory access speeds necessitates that parallel processing be applied to the handling of extremely large data sets[17]. The main problem encountered here is huge amount and variety of data involved in various activities. It needs several computations and communications simultaneously. All of these need to be done with less time delay. Hence there is a need for multi computer distributed system which will support faster computation and communication. Parallel processing distributes the data across different nodes and these nodes operate on the data in parallel which involves message passing. In big data system, the map-reduce frame work involves huge message passing and faster computation.

A Message passing communication can easily supplement the distributed shared memory model. Each communicating entity has its own message send/receive unit. The messages are not stocked on the communications link, but rather at the senders/receivers at the end nodes. But in case of, shared memory communication can be seen as a memory block used as a communication device, in which all the data are kept in the communication link/memory. Applications in which processing units function relatively independently, are accepted easily as candidates for message passing communication. The message passing in massive parallel computing system has been a potential area of research for many decades. The big data concept has also emphasized on the faster communication among the processing units. As such the parallel interconnection network is the backbone of big data systems [17]. Thus more emphasis is always given to faster data access while storing and/or retrieval from the data store as well as during processing.

In the parallel systems where routing and broadcasting are faster and low cost, those will be a better candidate for implementation purposes. A parallel computing system consists of multiple processing units connected via some interconnection network plus the software needed to make the processing unit work together. The interconnection network is the backbone of any parallel computing system [3 and 4]. Nowadays massively parallel computing is at the background of every scenario, including big data. In the big data concept, large problems are divided into smaller parts, which are solved simultaneously. It helps to complete the work faster than serial processing. Multiple instruction multiple data (MIMD) computer architectures are the most popular concept in this regard.As the Parallel Interconnection Networks (PIN) are behind any parallel computing system, so many researchers are experimenting with their structure [3]. Major PINs are Hypercube [4], Crossed Cube, Star graph [5], Extended Hypercube, Extended Star and Double star are to name a few.

Recently, the use of cluster computing with thousands or more number of nodes has been used profusely to cater the need of Big data. As packing density in a cluster increases, there is a rapid growth in the complexity of the communication subsystem. It may cause delay in message passing intervals over the interconnection and it is a stern performance issue in the execution of parallel programs[1 &2]. To increase the packing density of a parallel system several variations of the interconnection exist. Mostly they are derived from the Hyper cubes or star networks among all existing structures [4,6,9]. The popular variations are derived either by addition of nodes or by alteration of links or by both. In some interconnections a network controller node is added to the base network to improve message communication.

The Extended Hypercube (EH) is based on such scheme [8]. Other such variations are extended varietal hypercube (EVHC) [10], Extended crossed cube (ECC) [12], Extended star (ES) [13 & 14] etc. In the EH, EVHC and ECC there are exactly same number of nodes, but variation is realized only in the diameter. In extended star network there is a little modification in total number of nodes similar to the variation in star as compared to the hypercube. But in Extended double star the packing density is very high as it is designed with an aim to become a promising candidate for the Big data system. To cater to the needs of massive parallel system the Extended Double star network (EDS) is proposed here based on the double star network topology [16].

The current work proposes an interconnection network architecture that is called as Extended Double Star (EDS) for implementing the big data framework. The details like construction, topological parameters and routing algorithm, broadcasting algorithm need to be discussed and compared with the existing networks. The paper is organized as follows: Section two presents the background of the work followed by the proposed topology and different topological parameters in Section three. Then the Section four presents the performance analysis of the EDS network. Then Section five presents the routing and broadcasting algorithms. Complexity comparison for EDS network are discussed in the Section six. Lastly concluding remarks are presented in Section seven.

2. BACKGROUND

A.HYPER CUBE

A Hyper Cube is a PIN where each node is having node degree n. In hyper cube the node degree and dimension are same [4]. Message passing to every node is easy with message density 1.

B.STAR GRAPH

The star graph shown in Fig.1 is the n dimensional star also called as n-Star. In this symmetric structure there are $n!$ number of nodes and $\frac{n!(n-1)}{2}$ number of edges. Vertices in an n-star has $(n-1)$ incident edges. Several variation of star graph exist namely; Star cube[6], Generalized Star Cube [7], Extended star [13], and Star Crossed Cube [11].

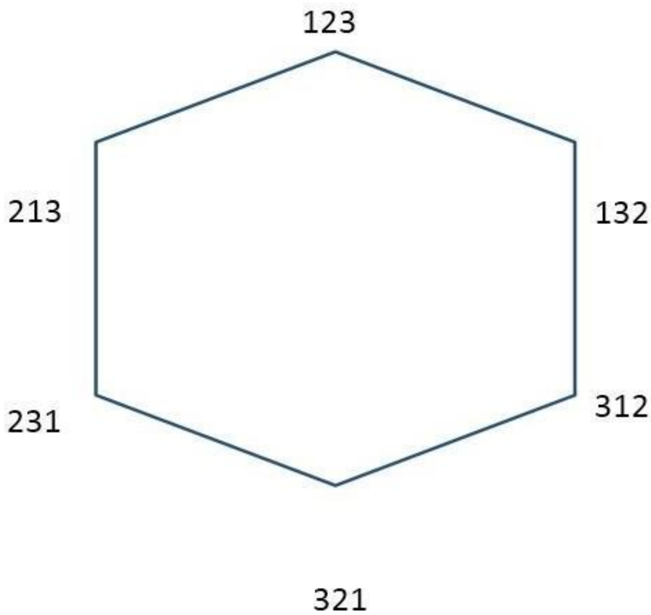


Figure 1: Star Graph of Dimension 3

The node addresses are permutation of 1, 2,3……n. For example in 3-star there are six nodes and node addresses are 123, 213, 312, 132, 231, 321 respectively using permutation of 3 bits(1,2 & 3).

C. EXTENDED STAR (ES)

Basically the ES is a hierarchical topology with basic module as n-star graph and one network controller which helps the processing elements solely concentrate on computational work [7, 11]. A ES consisting of $(n!)^k + \frac{(n!)^k - 1}{(n!) - 1}$ number of nodes and having edges is $\frac{1}{2}(n+1)! + \frac{(n!)^k - 1}{(n!) - 1}$. ES retains the positive features of the n-star at different levels of hierarchy and has some additional advantages like reduced diameter, constant degree of node, better performance, isolation of devices and benefits from centralization. The diameter of ES is about two-third of the diameter of star, thus, it reduces the cost of travelling in the network.

D. DOUBLE STAR

The double star graph as shown in Fig.2 is derived from the star graph. In DS graph the number of nodes is $2n!$ and edges of the graph is $E = n2 - n + n!$. The degree of n- dimensional DS(n) is n. Hence it is regular and recursive in nature [15 and 16]. The node address in DS has two pairs V(x, y), where x is a binary bit representing inner ring or outer ring and y is the n bit permutation denoting the star address part. The star part of the address helps to establish the connection between the consecutive basic modules when constructing higher dimension graphs.

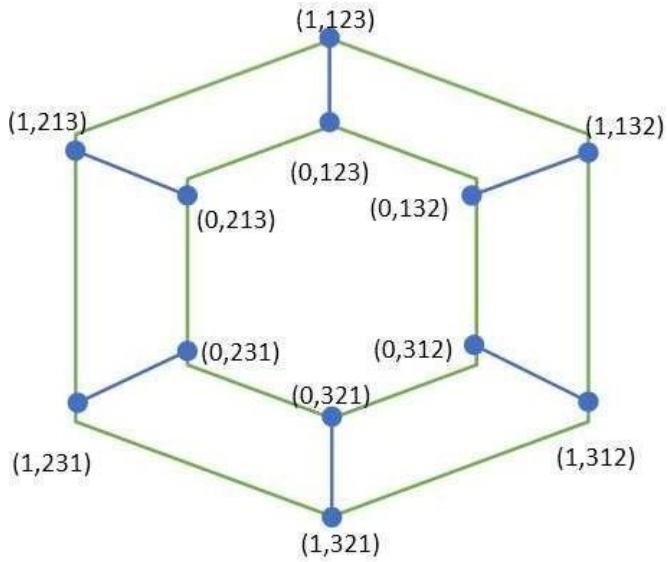


Figure 2: Double Star of Dimension 3

3. PROPOSED TOPOLOGY

This section describes the proposed EDS network topology. It is a derived network based on the Double star (DS) network. EDS is having more number of nodes than star graph of same dimension. Similar to Double star network, the nodes of the EDS network have two significant parts in the node addressing. The nodes of outer ring of DS network are connected to a network controller. To reduce the link complexity the nodes of outer ring are connected to the NC. There is no direct connection from inner ring to the NC. With double the number of nodes computation and message passing is also doubled. Hence the concept of extension is applied to DS topology which results in the proposed EDS network and will support faster routing. The node addressing concept need brief description. For convenience brief explanation about structure and basic topological parameter are presented below.

A. Construction and Addressing

Basically, the EDS is a hierarchical topology with basic module as DS graph and one NC which helps the processing element to solely concentrate on computational work. The EDS has two parameters namely n and k, where n is the dimension of DS and k represent the level of the NC. The EDS consisting of a DS and one network controller forms the basic module EDS (3,1) as shown in Fig. 3. The EDS (3,2) is shown in Fig 4. The movement from outer ring to inner ring and vice versa takes place through leaf edge. In Fig. 3 the node addressing for basic module is done using the star notation that is the permutation of n bits along with a binary bit for designating the ring structure.

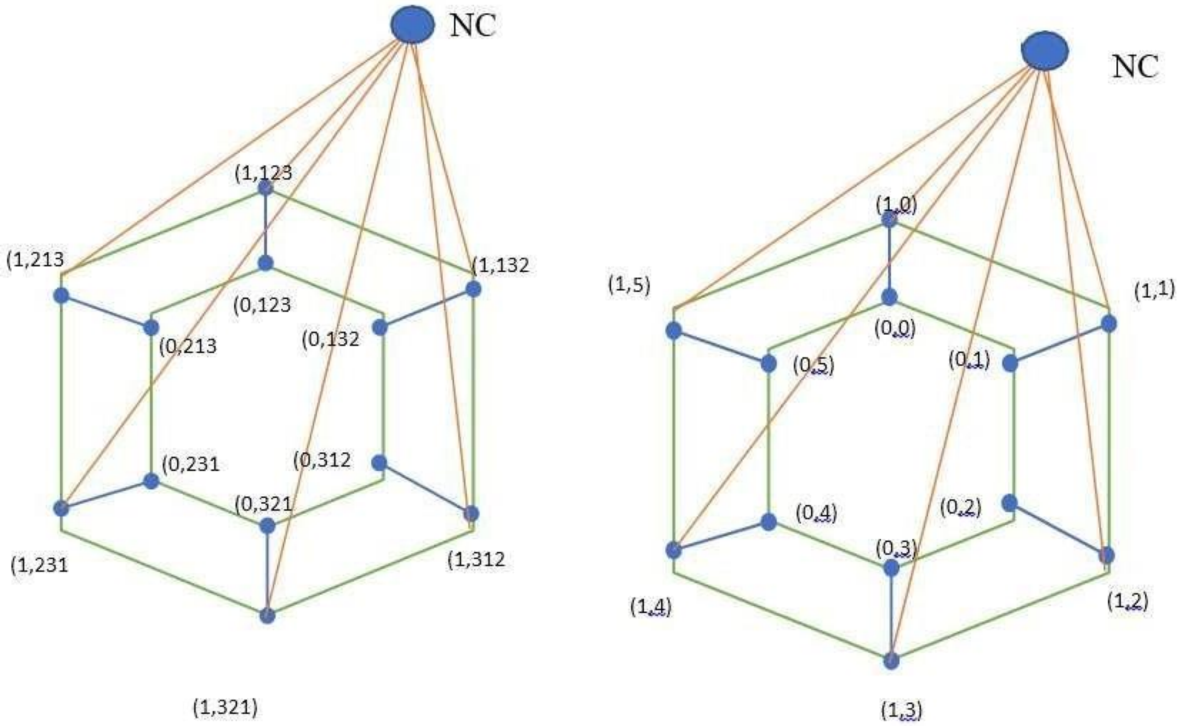
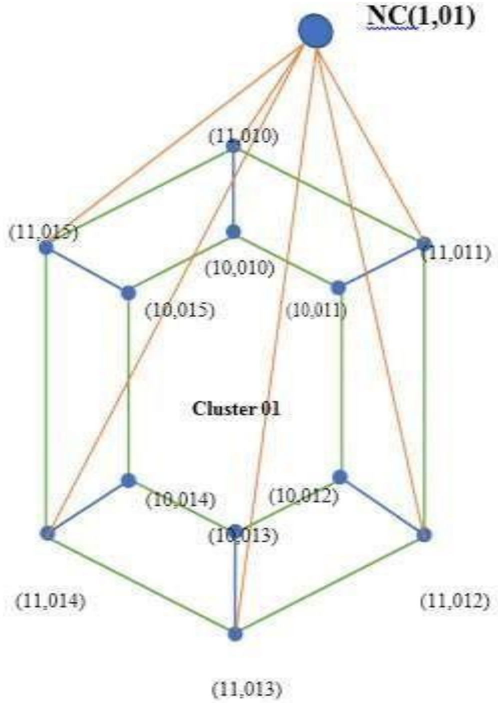
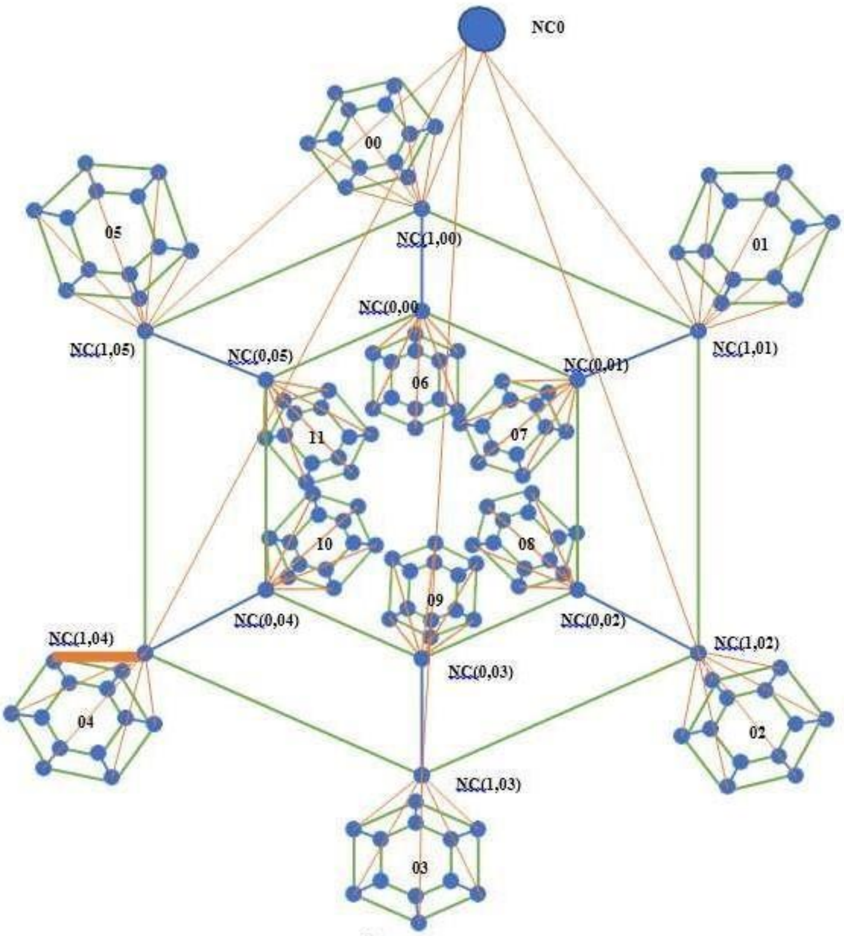


Figure3(a): EDS (3,1), (b): EDS (3,1) with Alternate Addressing

Next in Fig. 3 (b) the alternate address of nodes are shown with decimal numbering of six nodes starting from 0 to 5 with one binary bit for ring identification. Here 0 is for inner ring and 1 for outer ring. The address-labelling scheme is a basic step to construct the multi- computing system as it helps in designing the routing algorithm. The basic building block for an EDS is a DS containing $2n$ number of nodes. The node address has two parts x and y. Thex part denotes the level of NC and position of NC (inner ring and outer ring). The y part denotes Cluster address and position of node in the cluster. The address of NC proceeds the addressing of node. The EDS (3,2) is shown in fig 4(a) and 4(b). The outer ring contains six clusters (00) to (05). Then the inner ring contains 6 clusters addresses starting from (06) to (11). Single cluster (01) is shown in Fig 4(a).



(A)



(b)

Figure 4(a): Cluster 01 of EDS(3,2), (b): EDS(3,2) with NC Addressing

B. Topological Properties

The different topological properties of EDS network are as follows:

Theorem-1: The total number of nodes in EDS (n,k) is $P = (2n!)^k + \frac{(2n!)^k-1}{2(n!)-1}$.

Proof: For double star network the total number of nodes is $2n!$. In DS network it uses two-star graphs in the inner and outer rings, so the total number of nodes is $P = n! + n! = 2n!$. Next the basic module is connected if double

is $\frac{(2n!)^k - 1}{2(n!)-1}$ star fashion toand construct the total the number next dimension of PEs is graphgiven similarby $(2n \text{ to!})^k$ the . Therefore, Extended $P = (2n!)^k + \frac{(2n!)^k - 1}{2(n!)-1}$ Star network. The total number. of NCs

Theorem 2: The degree of EDS(n,k) network is $(n! + n + 1)$.

The topological parameters of the EDS network are discussed along with the parent networks below in Table 1 for better understanding.

Proof: In an EDS (n,k) , each PEs directly connected to n neighboring PEs of the same cluster and to one NC at the next higher level. In the EDS there are 2 rings:

1. Inner ring ($EDS3$)
2. Outer ring ($EDS3$) and a NC

Inner ring ($EDS3$) has connection to 2 neighbor PEs, so the degree is (n - 1).

Outer ring ($EDS3$) has connected to 2 neighbor PEs and NC, so the degree is (n). From inner ring one connection exist to connect the outer ring.

So finally the degree of nodes in ($EDSn$) is (n + 1).

The degree of NC other than the top most NC is ($n! + n + 1$) and topmost NC is only n! .

Theorem 3: The total number of edges E in EDS (n,k) network is $E = 2n(n-1) + n!\frac{(2n!)^k-1}{2(n!)-1}$

Proof: The basic building block, there are n! number of links in both inner and outer rings and n! connections to connect both rings. So the edges in both rings is $\frac{(n-1)}{2} + \frac{(n-1)}{2}$. In order to

connect the two graphs, it uses the leaf edges. So the edges of that node is $E = \frac{(n-1)}{2} + \frac{(n-1)}{2} + n! = n2 - n + n!$.

But in case of EDS(n,k) an extra extended node which is called NC connected to the both inner and outer ring.

So the total links $E = 2(n-1) + n!\frac{(2n!)^k-1}{2(n!)-1}$

In the next higher level, there are 2n! number of clusters with 24 links each again connected to both inner and outer rings containing NCs at level 01 and n! connections to NC at level 0.

Thus, the total number of links is $E = 2n\,(n-1) + n!\,\frac{(2n!)^k-1}{2(n!)-1}$.

Theorem 4: The diameter of the EDS(n,k) network is $Dg = \lfloor \frac{3}{2}(n-1) \rfloor + 2k$.

Proof: In extended DS the basic structure is a double star in inner ring as well as outer ring connected and a network controller at next higher level, so the distance between any two farthest nodes is when one node belongs to outer ring and the other exits in the inner ring depending upon the level of NC.

The cost of travelling in DS network is $\lfloor \frac{3}{2}(n-1) \rfloor + 1$.

Next travelling from one level of NC to next higher or lower level the number of edges involved is ($2k$-1) . Hence, the diameter of the EDS network will be,

$$Dg = \lfloor \tfrac{3}{2}(n-1) \rfloor + 1 + 2k - 1 = \lfloor \tfrac{3}{2}(n-1) \rfloor + 2k.$$

Theorem 5: The cost of EDS(n,k) is $C = [\lfloor \frac{3}{2}(n-1) \rfloor + 2k] \times (n! + n + 1)$.

Proof: The cost of any parallel interconnection network is the product of its degree and diameter. So the cost of EDS is defined as follows,

$C = DgEDS(n,k) \times DegreeEDS(n,k)$

$= [\lfloor \frac{3}{2}(n-1) \rfloor + 2k] \times (n! + n + 1)$

Table 1: Comparison of Topological Properties of EDS Network

Parameter	n-star	Extended Star	Double Star	EDS
Nodes	n!	$n!^k + \frac{n!^k - 1}{n! - 1}$	2n!	$(2n!)^k + \left(\frac{2n!^k - 1}{2n! - 1}\right)$
Links	$n! \times (\frac{n-1}{2})$	$\frac{1}{2}(n+1)! \times \frac{n!^k - 1}{n! - 1}$	$n^2 - n + n!$	$2n(n-1) + n!\left(\frac{2n!^k - 1}{2n! - 1}\right)$
Degree	n-1	$n! + n$	n	$(n! + n + 1)$
Diameter	$\left\lfloor \frac{3}{2}(n-1) \right\rfloor$	$\left\lfloor \frac{3}{2}(n-1) \right\rfloor + 2(k-1)$	$\left\lfloor \frac{3}{2}(n-1) \right\rfloor + 1$	$\left\lfloor \frac{3}{2}(n-1) \right\rfloor + 2k$
Cost	$(n-1) \times \left\lfloor \frac{3}{2}(n-1) \right\rfloor$	$(n! + n)\left\lfloor \frac{3}{2}(n-1) \right\rfloor + 2(k-1))$	$n \times \left\lfloor \frac{3}{2}(n-1) \right\rfloor + 1$	$\left\{\left\lfloor \frac{3}{2}(n-1) \right\rfloor + 2k\right\} \times (n! + n + 1)$
Fault Diameter	$\left\lfloor \frac{3}{2}(n-1) \right\rfloor - 1$	$\left\lfloor \frac{3}{2}(n-1) \right\rfloor + 2k$	$\left\lfloor \frac{3}{2}(n-1) \right\rfloor$	$\left\lfloor \frac{3}{2}(n-1) \right\rfloor + 2k$-1

3. PERFORMANCE ANALYSIS OF EDS

The performance analysis is done in terms of fault tolerance, node, diameter, degree and cost. All comparisons are done against network dimension. The parent networks namely the n-star, double star and extended star graph.

A. Fault Tolerance

The EDS network is a hierarchical structure with DS network as the basic buildingblock where the inner ring behaves like the backup node for outer ring. Also the NC provides additional path for faster routing or message passing with failure of one node or any node in the network. For all the PEs, the node degree is (n+1) , sothe EDS can tolerate up to n faults. Thus the EDS network become more robust as compared to the n-star as well as Extended star.

B. Fault diameter

In any PIN, the node or link failure results in a deformed network and the resulting diameter is called as the Fault Diameter (Df) derived from $Dg$by deleting at most *f* number of vertices. By principle the Df need to be close to Dg.

Theorem 6: The fault diameter of EDS(n,k) is $Df = \lfloor \frac{3}{2}(n-1) \rfloor + 2k$-1

Proof: From Theorem 4, it is clear that the $Dg = \lfloor \frac{3}{2}(n-1) \rfloor + 2k$. If there is one node at fault in the outer ring then, removing that particular node, we get the fault diameter as $Df = \lfloor \frac{3}{2}(n-1) \rfloor + 2k$-1.

C. Cost effectiveness Factor

While calculating the cost of a multiprocessor network, along with the cost of processing elements, the cost of the communication link is also consider. In EDS based network the number of links is a function of number of processors. The cost effectiveness factor takes this into account and gives more insight to the performance of multiprocessor system.

Theorem 7: The cost effectiveness factor of EDS(n,k) is CEF(p) =

$$\frac{1}{1+\rho \times \left[\frac{2n(n-1)}{p} + n!\left(1-\frac{(2n!)^k}{p}\right)\right]}$$

Proof: In general the number of links is a function of the number of nodes that is E= $f(p)$.

The total number of processors is: $p = (2n!)^k + \frac{(2n!)^k-1}{2n!-1}$

And the total number of links in EDS $= 2n\,(n-1) + n!\,\frac{(2n!)^k-1}{2n!-1}$

$\Rightarrow E = 2(n-1)+n!\,(p-(2n!)^k) = f(p)$

Again, $g(p) = \frac{f(p)}{p} = \frac{2n(n-1)}{p} + \text{n!}\left[1-\frac{(2n!)^k}{p}\right]$

Where g(p) is the ratio of number of links to the number of processors and ρ is the ratio of link cost to processor cost.

Hence, CEF(p) = $\frac{1}{1+\rho\times g(p)} = \frac{1}{1+\rho\times\left[\frac{2n(n-1)}{p}+n!\left(1-\frac{(2n!)^k}{p}\right)\right]}$
(proved)

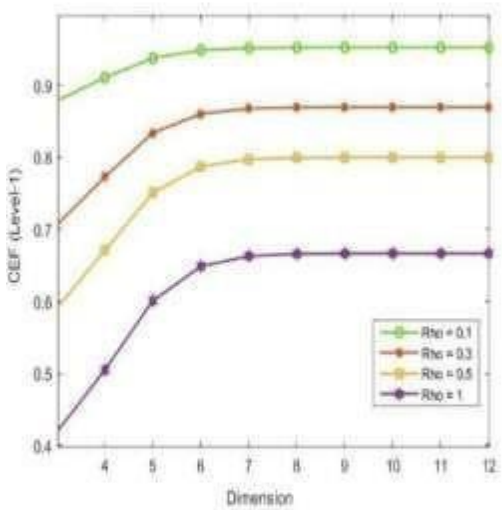


Figure 5(a): Cost Effectiveness Factor (Level-1)

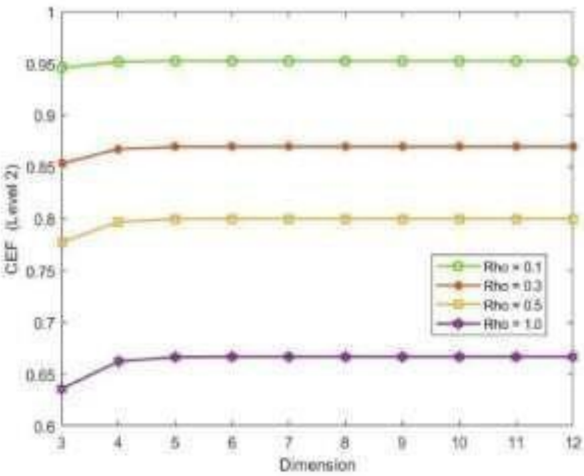


Figure 5(b): Cost Effectiveness Factor (Level-2)

Figure 5(a) shows the variation of CEF at level-1 with respect to the dimensions. The nature of the curves is monotonically increasing to dimension-6 and then shows constant nature from dimension-6 onwards. So, it is optimal for higher dimensions.

Figure 5(b) shows the variation of CEF at level-2 with respect to the dimensions. The nature of the curves is increasing to dimension-4 and then after dimension-4 onwards it shows constant nature. So, it is optimal for higher dimensions.

D. Comparison of Node

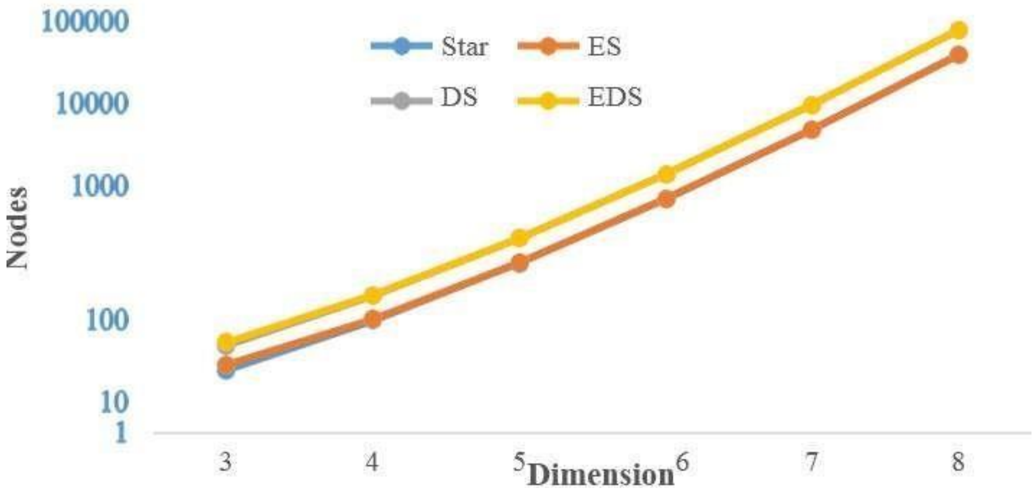


Figure 6: Comparison of Nodes

The comparison of nodes for EDS network is shown in Fig.6. The EDS network contains the highest number of nodes among all and its just the double size of the ES network. Star graph being the base network is the smallest and others are derived from it.

E. Comparison of Degree

The comparison of node degree against dimension is shown in Fig.7. Addition of NCs though eases communication, but it increases the overall degree of the graph. But with inner ring structure in EDS, the nodes serve as back up nodes and help incase of node failure. The ES and EDS have almost equal node degree being hierarchical in structure.

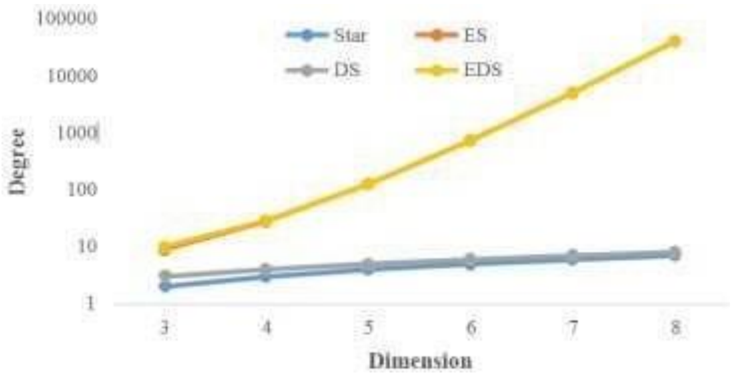


Figure 7: Comparison of Degree

F. Comparison of Diameter

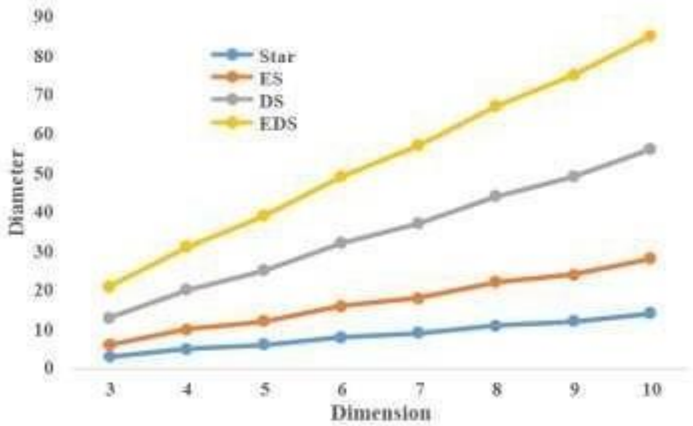


Figure 8: Comparison of Diameter

The most attractive property of the EDS network it its diameter. As the double star is a regular graph with degree and dimension both n similar to hypercube, the derived network EDS has the best values for diameter while packing the highest number of nodes among all networks as shown in Fig. 8.

G. Comparison of Cost

The comparison of cost for EDS network with star, ES and DS networks is shown in Fig. 9.Though the size of EDS is just double that of the ES network, but the cost of travelling in EDS is close to that of the ES network. This comparative study reveals that EDS is more cost effective.

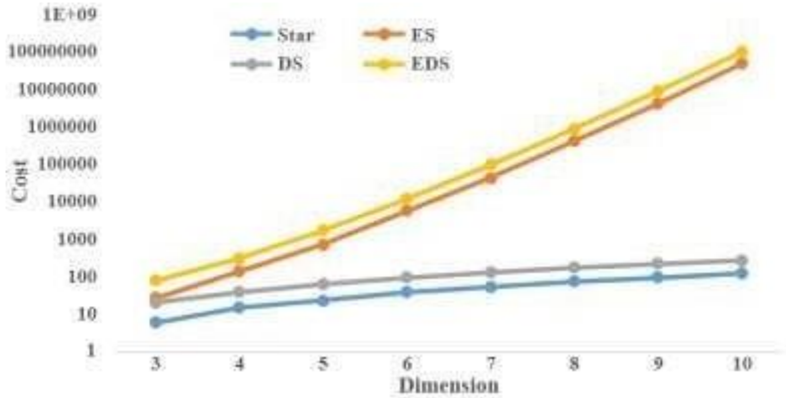


Figure 9: Comparison of Cost

5. ROUTING AND BROADCASTING

In multi computer systems, communication is an important issue. For message communication in multi computer system, processors exchange message effectively and reliably. This issue can be addressed by two important concepts, routing and broadcasting. An optimal routing algorithm always finds the shortest path between any two communicating nodes.

A. Routing

This section discusses the routing issues in EDS graph. The use of NC categories the inter - processor communication into two sub classes namely the local communication and the global communication. Communication among the PEs belonging to the same cluster is classified as local communication. Communication among the PEs of different basic modules via the network controller is called global communication. Nodes in the EDS have two parts namely ring component and star component. Here we assume that ‘S’ and ‘D’ be the source and destination nodes respectively.

B. Routing for local communication

This section discusses the routing issues in EDS graph for local communication when PEs belong to the same cluster. Depending upon the position of S and D in the EDS there can be following two possible cases.
Case I: - Both S and D belong to same ring.Case II: - S and D belongs to different rings.
Inside both inner and outer ring, the normal star routing is used. Thus, the star routing is a subfunction of the EDS routing due to the hybrid structure.

C. One to One Routing Algorithm:

Step I: If (S and D) belong to inner ring then follow star routing process within the ring.
Step II: If (S and D) belong to the outer ring then follow star routing in that ring.
Step III: If S is in inner ring and D is in outer ring then take leaf edge. Follow star routing in outer ring.
Step IV: If S is in outer ring and D is in inner ring then take leaf edge. Follow star routing in inner ring.

D. All to All Routing

Every node in the network sends a message to all other node in the network. Hence the one toone algorithm can be executed for each node for complete routing in the network.

E. Routing for Global Communication

This section discusses the routing issues in EDS graph for global communication when PEs belongs to the different clusters. In global communication the NCs are involved. The topmost NC transmits the message form source to destination PEs via the ring of NCs. The message passing operations in the global communication involves: a) the source PE, b) Up to *2(l-1)* NCs and c) the destination PE. The transfer of message between two nodes at different levels of hierarchy is referred to as the vertical shift.The routing in EDS involves two vertical shifts for level-to-level communication and a star shift for movement in the double star. The algorithms first check whether it is a local communication or a global one.

The message routing for different source and destination PE has been given for illustration in Table 3.

Table 3: Message Routing Sequence

Distance	Destination	Routing sequence from (00,060) source
1	(00,061)	(00,060) – (00,061)
2	(00,090)	(00,060) – (01,060) - (0,00) – (0,01) – (0,02) – (0,03) – (01,090) – (00,090)
3	(10,030)	(00,060) – (01,060) - (0,00) – (1,00) – (00) – (1,03) – (11,030) – (10,030)
4	(11,041)	(00,060) – (01,060) - (0,00) – (1,00) – (0,0) – (1,04) – (11,040) – (11,041)

The following procedure describes the routing process.

Let u be the source node with address $D_L{}^s D_L{}^s{}_{-1} D_L{}^s{}_{-2} \dots \dots D_0{}^s$, and v be the destination node with address $D_L^d D_{L-1}^d D_{L-2}^d \dots \dots D_0^d$
An algorithm is proposed below for message routing.

Algorithm:

MsgEDS (u, v)

1. *Begin*
2. *If* $[D_L{}^s \dots . D_0{}^s] = [D_L{}^d \dots . . D_0{}^d]$
3. *Then destination is source; terminate.*
4. *Else set* $j = 0$
5. *While* $j = 0$

6. Do for $i = 1$ to L
7. Begin
If $D^s_{L-1} = D^d_{L-1}$ then j=I;
End;
8. Begin
Vertical shift from $D_L{}^s D_L{}^s{}_{-1} D_L{}^s{}_{-2} \ldots \ldots D_0{}^s$ to $D^d_L D^d_{L-1} D^d_{L-2} \ldots \ldots D^d_j$;
Star shift from $D_L{}^s D_L{}^s{}_{-1} D_L{}^s{}_{-2} \ldots \ldots D_j{}^s$ to $D_L{}^d D_L{}^d{}_{-1} D_L{}^d{}_{-2} \ldots \ldots D_j{}^d$;
Vertical shift from $D^s_L D^s_{L-1} D^s_{L-2} \ldots \ldots D^s_j$ to $D^d_L D^d_{L-1} D^d_{L-2} \ldots \ldots D^d_0$;
9. End
10. End

Illustration
For example, The routing paths between source node (00,060) of Fig.4(b) for inner ringcluster (06) to destination node (00,090) for cluster (09) of same inner ring are as follows.
(00,060) – (01,060) – (0,00) – (0,01) – (0,02) – (0,03) – (01,090) – (00,090)
The routing paths between source node (00,060) for inner ring cluster (06) to destinationnode (10,030) for cluster (03) of inner ring are as follows:
(00,060) – (01,060) – (0,00) – (1,00) – (0,0) – (1,03) – (11,030) – (10,030)

F. Broadcasting:
There can be two broadcasting options in EDS networks namely up and down broadcasting.

Assumptions
'M' The message to be transmitted in EDS(k,l)
'S' The source node in EDS(k,l) one of the basic module.
($D^s_l D^s_{l-1} D^s_{l-2} \ldots\ldots\ldots D_0{}^s$) The source address
$D_l = 0$, $D^i_l = \text{i}$ where $0 \leq \text{i} \leq k! - 1$

Procedure: EDS_broadcast(s, k,l)
Begin
For i = 0 to l
Up broadcast from ($D^s_l D^s_{l-1} D^s_{l-2} \ldots\ldots\ldots D^s_l$) to ($D_l$) via next level NC
If I = l-2 Para do
All NC's perform Down broadcast to all PEs ($D_l D_l{}^i{}_{-i}$)
End

The broadcasting procedure is to make copy of a single message of the source node at all others nodes. In the above procedure during the Up broadcast the top most NC is reached in *(l-1)* number of steps. Next the NC's will do down broadcast in one step in parallel. Thus the total number of steps to do complete broadcast is (*l1+Pl*) for P number of processing elements. Thus, the broadcast algorithm can be executed in *O(l)* parallel steps.

6. Complexity Comparison for EDS Network

This Section is devoted to discuss the complexity of Extended Double Star Graph(EDS) and compare it with other contemporary parallel interconnection networks like star graph, ECC, EVHC and the ES topology. The following complexity evaluation for EDS graph are considered for comparison.
a) Linked Complexity growth rate,
b) One-to-one routing complexity,
c) Broadcast time complexity.

For any PIN the link complexity is the growth in number of links. For the EDS graph the link complexity growth rate is calculated as follows.
Number of links aNumber of nodes.

$$\Rightarrow E \text{ a } v$$
$$\Rightarrow E = \mathbb{C}V$$

$$\Longrightarrow \mathbb{C} = \frac{E}{V}$$

Here V = Total number of nodes, E is total number of edges and $\mathbb{C}$ = Constant of Proportionality.

In the graphs shown below, among all cube derivatives only EVHC is shown for comparison whereas EH and ECC are not shown as all the three networks contain equal number of nodes.

In this figure 10(a) Link complexity growth rate for EDS, ES, and EVHC are compared. Here X-axis represents network dimensions and Y-axis represent the link complexity growth rate. In this comparison EDS link complexity growth rate is less in comparison to ES and EVHC network as morenumber of nodes exists in EDS. So it is a better network in comparison with ES and EVHC for massive parallel communications with respect to practical implementation. The slower rate in link complexity growth is due to connection of network controllers in the outer ring only.

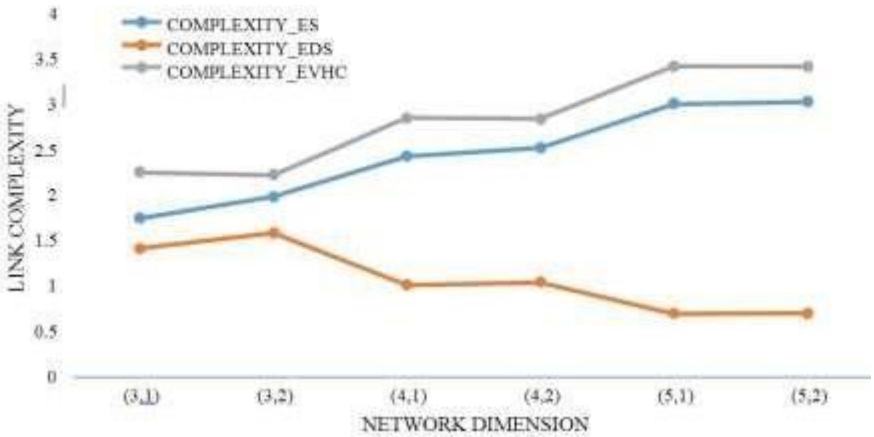


Figure 10(a): Link Complexity Growth Rate Comparison

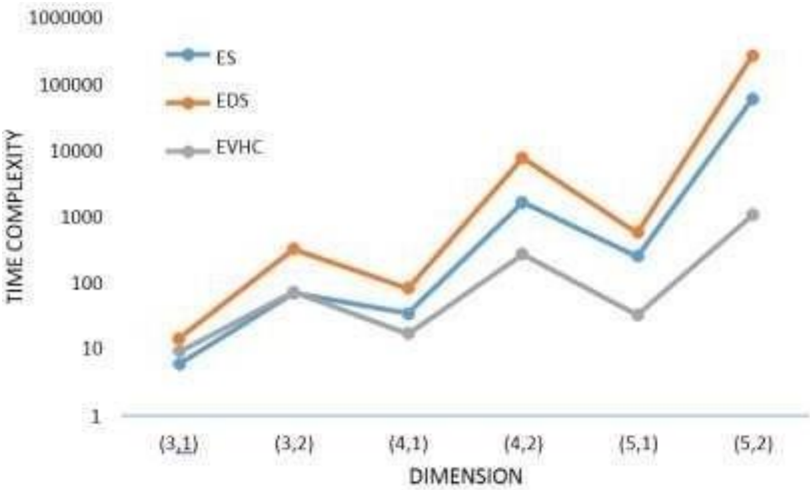


Figure 10(b): One to one Routing Time Complexity Comparison

For EDS graph the one-to-one routing algorithm finds a shortest path between source and destination node in O(n log n) time. In this case for both inner and outer rings, the normal star routing is used. The time complexity of star routing is O(n log n) [5]. In this EDS routing different clusters with NCs are involved. The top most NC transmits the message from source to destination PEs via rings of NCs connected in star fashion. So the total time complexity in this case is (n log n) or (n log n +1) $\cong$ n logn for large n. So the time complexity of EDS graph is O(n log n). The comparison is shown in Fig. 10(b) Here X-axis signifies the dimension of the network and Y-axis represents the time complexity. The figure depicts the EDS network takes more time as comparison to ES and EVHC. But in EDS the number of nodes is just the double as compared to the ES and EVHC networks. In comparison to the packing density the routing time is not increased in the same rate. Hence in comparison to other extended networks namely ES, EVHC, EH and ECC, EDS provides better and faster routing. In the graph, only EVHC is shown whereas EH and ECC are not shown as all the three networks contain equal number of nodes.

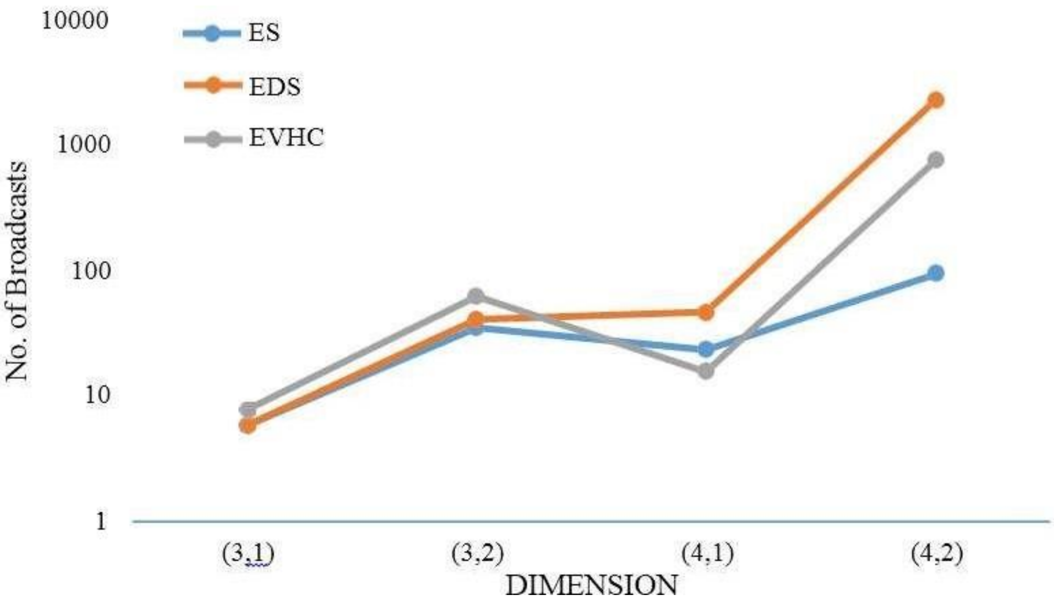


Figure 10(c): Comparison of Broadcasts against Level of Network

In the broadcasting procedure only the NCs are involved in the extended type of graphs. The broadcasting algorithm can be executed in O(*l*) parallel steps. Hence comparison is made to show the number of nodes covered in one step with respect to the level of NCs and shown in the Fig. 10(c). Here X-axis denotes the dimensions of the network with two parameters (n and k) and Y-axis represents the quantity of broadcasts. At level 0, all the extended networks do nearly same number of broadcasts. But in higher levels, the EDS network more number of nodes are covered both in inner and outer rings during routing/message passing process as compared to ES and EVHC networks. So EDS is a superior network in connection with broadcasting issues at higher dimensions. The ES network runs with a midrange value as it packs more nodes than the cube based extended networks.

7. CONCLUSION

The current paper has attempted to propose a new hierarchical parallel interconnection network derived from star graph called as Extended Double Star graph EDS(n,k). The proposed network is hierarchical in nature, possesses highest number of nodes at low node degree. The cost of communication is less. The addition of the network controllers makes message passing faster and thus low cost message communication will be beneficial for the Big data system. The fault tolerance is better in EDS in comparison to star and extended star networks. The nodes in inner and outer rings can serve as back up nodes which may be more helpful in design of map reduce framework of big data systems.

The shortest path issue is a classic problem and studied by many researchers. The difficulties are more in massive data environment. As the EDS network a different flavour of hybrid network has been designed for big data systems, the current study paid attention to optimal routing issues and suggested one to one routing, one to all routing and all to all broadcasting algorithms. The Link complexity growth rate is evaluated and found to be less in EDS than other extended networks. Maximum number of nodes exits in EDS topology and is just the double the count of the ES network. In less time more nodes are covered in routing/message passing process at different levels of the network that is *O(nlogn)+1.*

8. REFERENCES

1. Tsai C F, Lin W C, Ke S W (2016) Big Data Mining with Parallel Computing: A comparison of Distributed and Map Reduce Methodology. Journal of Systems and Software, Volume 122, Pages 83-92
2. Zhang Y (2016) Parallel Processing Systems for Big Data: A Survey. In *Proceedings of the IEEE*, vol. 104, no. 11, pp. 2114-2136
3. Feng T (1981) A survey of Interconnection Networks. IEEE Computers, 1(4) , pp.(12-27)

4. Saad Y, M. H, Schultz (1988) Topological Properties of Hypercube. IEEE Transactions on Computers, vol. 37, no. 7, pp. 867-872.
5. Akers S B, Krishnamurthy B (1987) The Fault-tolerance of Star Graphs. Proceedings of International Conference on Supercomputing, pp. 270.
6. Pilas M A, Rajsekharan S, Davis SL Wei (1989) General Routing Algorithms for Star Graphs. Scholarly Repository, University of Pennsylvania.
7. Arai D, Li Y (2015) Generalized-Star Cube: A New Class of Interconnection Topology for Massively Parallel Systems. Third International Symposium on Computing and Networking (CANDAR), pp. 68- 74.
8. Kumar J M, Pattanaik L M (1992) Extended Hypercube: A Hierarchical Interconnection Network of Hypercubes. IEEE Transactions on Parallel and Distributed Systems. Volume 3, Issue 1, pp 45–57.
9. Tripathy C R (2004) Star-cube: A New Fault Tolerant Interconnection Topology for Massively Parallel Systems. IE(I) Journal, ETE Div., vol.84, no 2, pp.83- 92.
10. Tripathy C R, Dash R K (2008) A New Fault-tolerant Interconnection Topology for Parallel Systems. IE(I) Journal CP.
11. Adhikari N, Tripathy C R (2014) Star crossed cube: an alternative to star graph. Turkish Journal of Electrical Engineering and Computer Sciences, Vol.22, pp. 719- 734.
12. Adhikari N (2017) Extended Crossed Cube: A New Fault Tolerant Interconnection Network. Int. Journal of Engineering Science Invention, Vol.6, No.9, pp. 60-70.
13. Adhikari N, Nag B (2011) On Topological Properties of a Star based Large Scale Parallel System. Proceedings of ETNCC2011, International Conference on Emerging Trends in Networks and Computer Communications, IEI Udaipur Section.
14. Adhikari N (2018) On the Extension of a Star Based Parallel Interconnection NetworkTopology. Proc. of The Eighth Intl. Conf. On Advances in Computing, Control and Networking - ACCN 2018, pp.1-5.
15. Barik H, Swain N K, Rout L, Adhikari N (2019) Double Star: A high performance network for big data. SSRN eLibrary CTFC.
16. Barik H, Swain N K, Adhikari N (2020) Load Balanced Double Star: A High Performance Architecture for Data Intensive Applications. *International Conference on Computer Science, Engineering and Applications (ICCSEA)*, pp. 1-6.
17. Bushong, Michael (2016) Six considerations for big networks. Seachsdn.techtarget.