On Internet Multicast Architectures: Fully Distributed and Hierarchical Vs Service-Centric

Omar Said, Ahmed Ghiduk, Sultan Aljahdali
Faculty of Computers and Information Technology, University of Taif, KSA

Abstract: *There are two approaches for the multicast routing architecture. The first approach is a traditional multicast architecture that constructs and updates the multicast tree in distributed manner. The second and most recent approach is called service-centric, in which there are two types of routers. Efficient router, which is called m-router, handles many to many multicast functions. The other routers are called i-routers that handle only minimum multicast functions. This approach has drawbacks, originating from the centralization idea. This paper proposes two approaches that enhance the performance of the service-centric architecture; hierarchical architecture and fully distributed architecture. In our proposed architectures, the service-centric m-router is divided into three sub m-routers. The functions of each sub router are determined. How these routers communicate with each other to build the multicast tree is demonstrated. Management of the multicast tree in the new architectures is showed. How the new architectures recover the drawbacks of the current approaches is clarified.*

Keywords: *Routing Protocols; Internet Multicast Trees; Internet Protocols; TCP/IP*

1. Introduction

There are three fundamental methods for transmitting data over a network: unicast, broadcast, and multicast. Unicast can be defined as a traffic sent to a single specific destination such as a host computer, web server, or a particular end user. Broadcast can be defined as a traffic forwarded to all users of a network. Multicast traffic can be defined as a traffic delivered to a specific subset of the network's users. The implementation of both unicast and broadcast traffic is easy for networks. This is because the data packets will either be delivered to a single unique destination, or they will be propagated throughout the network for all end users. The implementation of multicast traffic is considerably more complex because users must be identified, and traffic must be sent to their specific locations. The network should also refrain from sending traffic to unnecessary destinations to maintain security and to avoid wasting valuable bandwidth. ISPs are concerned about the effects of multicast traffic on their networks. However, multicast traffic is increasing over the Internet [1, 2, 3, 4, 5]. Applications such as data casting, video and audio transmissions, and training seminars all depend on multicast technology. These applications are designed to deliver identical packets to a large number of receivers and the packets must be replicated at an exponential rate. The resulting bandwidth requirements and routing overhead associated with these applications can be quite daunting [6].

The paper proceeds as follows; Section 2 introduces the related work. In Section 3, the new hierarchical architecture is demonstrated. In Section 4, the new fully distributed architecture is demonstrated. In Section 5, general aspects for the two proposed architectures are introduced. A comparison between the current and the proposed architectures, is showed in Section 6. The conclusion and the future work are demonstrated in Sections 7 and 8.

2. Related Work

In recent years, there have been more research in the area of multicast routing. Traditional multicast routing protocols are classified into two classes: Shortest Path Tree (SPT) based multicast routing protocols and Shared Tree (ST) based multicast routing protocols. SPT based protocols build a separate multicast tree for each (source, group) pair rooted at the source. DVMRP (Distance-Vector Multicast Routing Protocol) [8] and MOSPF (Multicast Extensions to Open Shortest Path First Protocol) [9, 10] are SPT-based protocols. SPT-based protocols minimize end-to-end delay. However, there are three problems in SPT-based multicast routing protocols. These problems can be stated as follows, scalability problem for a large network, adopting DVMRP or MOSPF wastes a large portion of the network bandwidth due to flooding, and multicast trees generated in DVMRP or MOSPF are shortest path trees, which may not be the lowest cost multicast trees [7].

The scalability and bandwidth wasting problems are handled by proposing the ST-based multicast routing protocols. Core-Based Tree (CBT) [11], Protocol-Independent Multicast Sparse Mode (PIM-

SM) [12] and Simple Multicast (SM) [13], [14] are ST-based protocols [15]. However, the ST-based multicast routing protocols introduce new problems. These problems are; the less efficient multicast communication mechanism from a source to a multicast groups specially in terms of multicast tree cost and communication delay, the elected core has the same architecture as any other routers in the network, thus has limited computing and packet forwarding capability, and the ST-based approach may cause traffic jam around the core. In addition, the traffic concentration will cause the problems of packet loss and longer communication delay. Finally, the multicast communication between a source and a multicast group cannot tolerate the core failure [7].

In a more recent approach called service-centric multicast architecture proposed by yang [7], a powerful router, called m-router, collects multicast-related information and processes multicast requests based on the information collected. The m-router handles most of multicast related tasks, while other routers only need to perform minimum functions for routing. The m-router is designed to be able to handle simultaneous many-to-many communications efficiently. The Service-Centric Multicast Protocol (SCMP), builds a dynamic shared multicast tree rooted at the m-router for each group. The tree construction is performed by a special type of self-routing packets [7]. However, the centralization idea raises some problems.

The study and analysis of the service-centric approach extracted the following drawbacks; hardware complexity is high, the m-router needs lot of bandwidth to handle the multicast functions and this requirement may not be available, the fault tolerance idea is straightforward and is not clarified in details, the multicast architecture in service-centric approach is costly, the tree message is too complex to transfer from one level to another in the multicast routing tree, what will be done if the tree message is lost?, and the changing between the architecture messages (Branch and Tree) within the multicast tree management is ambiguous.

3. Proposed Hierarchical Architecture

The proposed hierarchical architecture has N management routers. In this paper, N equals three. The architecture management routers are called basic router, m1-router, and m2-router. The domain of ISP is divided into two parts. The first part is for m1-router and the second one is for m2-router. Other system routers are distributed depending on three factors; the link delay, link cost, and load balance. When a new router or host needs to login the ISP domain, it should communicate with m1-router or m2-router. The routers should send a unicast message to the up level m-routers. Consequently, the m1-router sends an upgrading aggregation message to the basic router. Regarding the aggregation message infrastructure, see Section 5.

At the m-routers level, a sub-multicast tree for each domain part is constructed. Each router on the sub-domain considers its m-router as a root of its sub-tree. Each m-router sends its sub-tree to the high level router that is called basic router. The basic router merges the two sub-trees in one multicast tree. Hence, the root of the multicast tree will be considered the basic router. The merging process is accomplished using the technique found at [16].

In the following subsections, the architecture components, load balance and fault tolerance aspects, multicast tree construction, relation between basic router and domain routers, and a case study to describe the architecture operations are discussed.

- *Architecture components*

Our proposed architecture consists of three different management levels. The first level contains a basic router. The basic router is the main manager in the multicast tree construction system. It has two inputs and two outputs leading to low hardware complexity. The basic router functions are session management, merging the multicast sub-trees in one tree, and recovering the m-routers in case of failure occurrence.

The second level of our proposed architecture contains two management routers called m1-router and m2-router. These two routers share the basic router in the management processes. Each router is responsible for the domain part. The main functions of this level are to construct the multicast sub-trees, collect hosts or routers information in its sub-domain, transfer the messages between connected domain routers and the basic router, and in case of the basic router failure, it can be replaced by one m-router. To solve this tip, at the system start-up, one of the m-routers is elected as an alternative of the basic router to complete the tree management functions. The election process depends on the load on each m-router (load balance).

- *Load balance aspect*

Two points of view as regards the load balance in our proposed hierarchy architecture should be demonstrated. The first one considers how other system routers will be distributed on the m1 and m2 management routers. The second is how the management functions will be distributed on the architecture main components. To clarify the first aspect, every m-router in the domain should have a general variable (counter) in its configuration file. This general variable will be incremented when the new router is connected to either m1 router or m2 router. Also, the link and delay costs are taken in consideration. The last values of the general variable will be stored at the basic router to determine with which m-router the new host will communicate. The following algorithm describes the process:

Algorithm

1- Assume that router1 is connected to m1 router
2- The m1 router will increment its counter and send the new value to the basic router.
3- If ((m1.counter <= m2.counter) && (link delay and link cost are accepted)),
   3.1 The basic router will send a confirmation message to m1 router informing with a correct connection for the new router
4- Else,
   4.1 The basic router sends a fail message to the m1 router informing it that the new router should communicate with m2 router.
   4.2 The m1 router sends a failure message of basic router to the new router for changing its connection to m2 router.
5. End

Regarding the second load balance aspect, every m-router contains the application software to run its main function. Also, the m-routers contain applications to run a session and a group membership functions, but these applications are inactive and will be excited at a recovery state only (i.e., at the basic router failure). In addition, the basic router contains applications for managing the multicast session and group membership. On the same idea, the basic router contains the inactive m routers applications. Inactive applications don't represent an overload on the system complexity.

- *Fault tolerance aspect*

Firstly, one of the m routers is elected from the basic router to recover it in case of failure. The election process depends on the load balance factor (i.e., each m-router). Assume that the m1 router is elected for this target (recovery of basic router). The m1 will communicate with m2 router using its IP address. The m1 fires its inactive applications to manage the multicast session till the basic router is repaired. The m2 router sends its multicast sub-tree to the m1 router. The tree message, which is stated in the service-centric approach, is used for transferring the m2 multicast sub-tree to the m1 router. Hence, the m1 router merges the two sub-trees (m1 sub-tree and m2 sub-tree) in one multicast tree and the root of that tree will be the m1 router. Also, the m2 router sends a multicast message to its downstream routers informing them with the new state.

- *Multicast tree construction in hierarchal architecture*

In the service-centric technique, the multicast tree is constructed virtually at the m-router. Consequently, the m-router sends a tree message to the downstream routers to start the physical construction of multicast tree. The main disadvantage of this technique is the complexity of the tree message. This complexity may lead to the loss of the tree message that causes a distortion in a multicast tree. So, our proposed idea is focused on how the multicast tree is constructed from downstream routers to upstream routers. Each router constructs a simple multicast tree and sends it to the upstream router that merges its downstream trees in one tree. This operation continues till the m routers level. Each m router constructs its multicast tree. The basic router will receive the two multicast trees from m1 and m2 routers and merge them. When a new host needs to join a group, it sends a group join message to the upstream router and this message continues transferring until it finds the router that is responsible for the target group. This router will be called the target router. The target router upgrades its routing entry and sends a message to the m-router (m1 or m2). Consequently, the m-router (the root of the new host) sends a message to the basic router informing it with the change that should be done in the multicast tree and routing entries. The construction of the simple trees and group join operations will be done simultaneously.

- *Basic router vs. domain routers*

The contact of system routers occurs only in relation to the m-routers (m1 and m2). When a communication between a system router and its m-router fails, it takes a permission to contact the basic router to recover the error by changing the m-router for that router or recover the failed connection (i.e. domain routers are not authorized to contact the basic router except in case of a failure occurrence). So, each router should have the IP address of the basic router with a conditional communication. The communication permission and restriction are built in the configuration file that will be installed at each system router. These permissions are adapted to be inactive in case of stable state, but they are fired in failed communication state.

- *Case study for hierarchal architecture*

To investigate our idea, a case study for constructing the multicast tree is demonstrated. A simple internet topology that may be found at any ISP is assumed. This can describe how the three management routers accomplish their functions in the multicast tree. The Delay Constrained Dynamic Multicast (DCDM) algorithm is used to construct a multicast tree with link delay and link cost. It is known that the DCDM algorithm can resolve the loops resulting in the supposed tree topology [7]. The result of DCDM algorithm execution is two multicast sub-trees; one for each m-router. Group joining and leaving is also discussed. The proposed topology contains 9 nodes, numbered from 4 to 13, with 3 management routers that are numbered as follows: 0 for the basic router, 1 for the m1 router, and 2 for m2 router. The topology contains three groups called G1 {4, 7, 12}, G2 {5, 8, 9}, and G3 {6, 10, 11}. We denote the values of link delay and link cost on each link, as shown in Fig. 1.a. To apply our idea in this case study, two steps should be clarified: 1- Extraction of multicast tree using the DCDM algorithm and solving the resulting loops, see Fig. 1.b. 2- Transformation of the multicast tree into two sub-trees one for each m-router, see Fig. 1.c. It's notable that the loops in the supposed topology, which are extracted after applying the DCDM, are (m1, m2), (m1, 5), (4, 5), (8, 9), (9, 10), (10, 11), (11, 12), and (7, 12) (denoted with dotted line). After the loops are deleted, the multicast tree is extracted. Consequently, the resulting tree is transformed into two sub-trees taking into consideration three factors: the load balance, the link delay, and the link cost. So, the first sub-tree is for m1 and contains the nodes 4, 7, 8, and 12. The second sub-tree, which resulted by division too, is for m2 and contains the nodes 5, 6, 9, 10, and 11. Assume that the router number 13 needs to join g2. It should contact the router number 8 for load balance factor and without neglecting the other two important factors (the link delay and the link cost). The general variables, which stored at the basic router (B0), determine the number of routers (or hosts) at each m-router. If a new router tries to contact router 5 or router 9, it will be redirected to router 8 by m2 router, see Fig.1.d. Suppose that the router 11 needs to leave the group g3 it should send a prune message to the router 6 that is considered the group manager (or tracer). Hence; the router 6 sends an update message to m2 router that sends the same type of message to the basic router, see Fig 1.e. By using this technique each sub-tree will be upgraded to the new state.

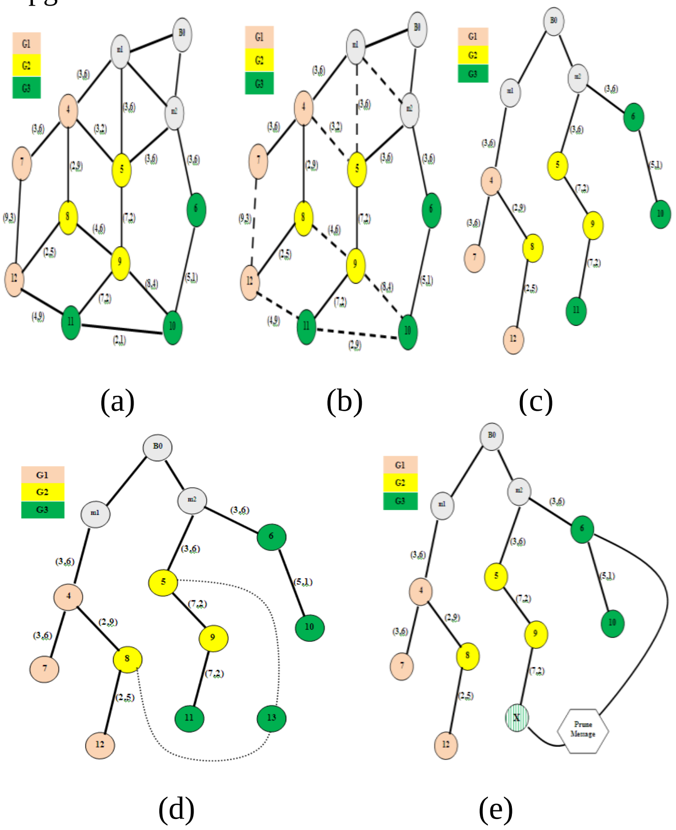


Figure 1. Hierarchical architecture case study

4. The Fully Distributed Architecture

It's notable that the service-centric architecture is fully centralized and the hierarchical architecture is considered a semi-centralized due to the basic management router idea. So, architecture with no centralization idea should be proposed and compared to the service-centric and the hierarchical architectures. The third architecture is called fully distributed.

- *Architecture components*

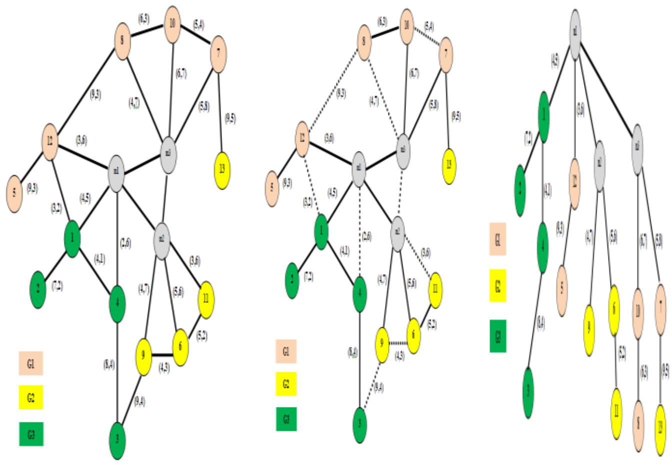
Simply, the proposed fully distributed architecture has three m routers. The domain of ISP is divided to three parts. The first part is for the first management router (m1); the second part is for the second management router (m2); the third part is for the third management router (m3). The privilege and restrictions are equal in the three management routers. Also, the factors that should be taken in consideration while other system routers distribution, are the link cost, the link delay, and the load balance. When a new router needs to connect the system, it should send a multicast message to the up level routers to change its multicast sub-tree. This operation is done persistently till one of the management routers sense the last updates in the routing tree. Each management router constructs its multicast sub-tree (i.e. it is considered a root of the constructed routing sub-tree). A master and its alternative management routers are elected from the three management routers simultaneously. The election process for fault tolerance aspect mainly depends on the load balance factor. After the election process, a multicast message contains master and spare management routers address is sent to other routers in the system. The load balance aspect in the fully distributed architecture is the same as in the hierarchical architecture.

- *Multicast tree construction in fully distributed architecture*

Each router in the system constructs a simple multicast tree (sub-tree) and sends it to the upstream router that merges its downstream trees in one tree. This operation continues till the m routers level. Each m router constructs its sub-multicast tree. The elected master m router receives the two sub-multicast trees from other management routers (considered m2 and m3 if m1 is the master management router) and merges them. When a new host needs to join a group it sends a group join message to the upstream router and this message continues transferring until it finds the router that is responsible for the target group. This router will be called the target router. The target router upgrades its routing entry and sends a message to the m-routers.

- *Case study for fully distributed architecture*

The proposed topology contains 13 nodes, numbered from 1 to 13, with 3 management routers that are called m1, m2, and m3. We denote the values of link delay and link cost on each link, as shown in Fig. 2.a. Firstly, the routers are normally distributed on the three management routers. As shown in Fig. 2.a, m1 manages the routers 5, 12, 2, 1, and 4, m2 manages routers 11, 6, 9, and 3, and m3 manages routers 13, 7, 10, and 8. The topology contains three groups called G1, G2, and G3. G1 {7, 8, 0, 12, 5}, G2 {9, 6, 11, 13}, and G3 {1, 2, 3, 4}. Fig. 2.b shows the resulting tree after the DCDM algorithm is applied. Assuming that the m1 is elected as a master router, the resulting multicast tree constructed by m1 is shown in Fig. 2.c. The new routers connection and deletion processes are the same as in the hierarchal architecture case study.



(a) (b) (c)

Figure 2. Fully distributed architecture case study.

5. General Aspects for the Proposed Architectures

In this section, the shared aspects for the two proposed architectures are demonstrated. These aspects are; the additional architectures messages, the routers configuration upgrades, and the architectures advantages.

- *Architectures messages*

Our proposed architectures contain three new types of messages; the multicast tree construction message, the tip message, and the fail message. These messages should be simple to guarantee that our architecture doesn't cause an overload when compared with the service-centric approach.

A. Multicast tree construction message

This message is sent from the downstream router(s) to the upstream router(s). This message contains the IP addresses of the downstream routers. This type of message is sent in unicast mode.

B. Tip message

This message is sent to the router in two states. The first state: when one from m-routers is failed, the basic router or the elected master router sends this message to the domain routers to inform them with the new state. The second state: when the basic router or the elected master router is failed, the alternative router that recovers the basic router sends this message to the domain routers informing them with the new management state. This type of message is sent in multicast mode.

C. Fail message

The basic (or master) router sends this message to the m1 router informing it that the new router should communicate with m2-router. This message contains the IP address of the new router and identity field. This type of message is sent in unicast mode.

- *Router configuration upgrades*

There are some upgrades that should be done in the configuration file of system routers. The IP addresses that should be added to the configuration file of domain routers are; the basic router or the elected master router IP address that will be used only at the recovery state, and m1 or m2-routers that will be used to connect the sub-domain. Also, the IP addresses that should be added to the configuration file of the management routers. Each router in the three management routers should have the IP addresses of other two routers. This is to complete the management functions, and recovery process. There is a general variable that is inserted only at the basic router or the elected master router, and will be incremented with one when a new router is connected to the domain.

- *The advantages of the proposed architectures*

A. Fault tolerance guarantee

Our architectures have an alternative for each management router (m1, m2, and basic). Hence, if the basic router or the master router is failed, m1-router or m2 router can take place. Also, if one m-router is failed the basic router can manage its sub-domain till repairing process is accomplished.

B. Scalable architectures

It's notable that our architectures idea is built on division of ISP domain into two or three parts one for each m-router. So, they can receive duplicated number of routers comparable to the service-centric idea.

C. Low hardware complexity

The basic or the elected master router needs only two ports to communicate with the downstream routers (m-routers). The m-routers needs n/2 (n/3 for fully distributed architecture) ports to manage its sub-domain routers, where n is the number of routers in the ISP domain.

D. Low required bandwidth

It's notable in the two proposed architectures that the load on each m-router is decreased. Consequently, the number of management messages will be decreased, hence; the required bandwidth for this link also will be decreased. In addition, the tree message that is used in the service-centric approach is too large. Hence; it requires more bandwidth than our system messages.

E. No centralization drawbacks

The main drawback of centralized architecture is that all the system functions are accomplished by one node (i.e., the scalability, reliability, and performance analysis of the system rely on the m-router). In our proposed architectures the management functions are distributed in balance with m1, m2, and m3 routers.

6. Summarization Table

Table 1. Comparison between the service-centric, the hierarchical, and the fully distributed architectures

Parameter	Service-Centric	Hierarchical	Fully Distributed
Centralization	Yes	No	No
Fault Tolerance	No	Yes	Yes
Load Balance	No	Yes	Yes
Simplicity	Yes	Yes	Yes
Tree Construction	Yes	Yes	Yes
Hardware Complexity	High	Less than Service-Centric	Less than Hierarchical[Moderate Delay]

7. Conclusion

It's difficult to build efficient multicast tree using traditional multicast architectures due to lack of complete information about the network infrastructure. Service-centric approach treats some drawbacks of the traditional approach, but suffers from the drawbacks of the centralization operation. This paper proposed two new modifications for the service-centric multicast architecture; namely: hierarchical and fully distributed architectures. The proposed modification replaces the service-centric m-router by three sub m-routers with specific functions, management, and interconnection strategies. Our proposed architectures theoretically outperform the efficiency of the service-centric and traditional architectures.

8. Future work

Our two proposed architectures will be simulated and compared with the service-centric architecture efficiency. The simulation environment will be constructed using ns2 [17]. Also, the recommended architecture issue (either hierarchical or fully distributed) will be decided after the simulation results extraction.

Acknowledgements

I would like to thank Taif University for providing support and material related to our research. Also, I am grateful to College of Computers and Information Technology for the early experience in teaching this subject.

References

[1] Ramjee P., Albena M., 2009. New Horizons in Mobile and Wireless Communications: v. 4: Ad Hoc Networks and PANs. Artech House Publishers, USA.

[2] Jian Xu Xiang-Cai Zhu, 2009. Aware Routing Protocol for MANET Accessing Internet. Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, Chengdu, China, Pages: 571 – 574.

[3] Shu-Quan Li Yue Wu, 2010. A Survey on Key Management for Multicast. The IEEE International Conference on Information Technology and Computer Science (ITCS), Ukraine, pages: 309

[4] Schoeneich, R.O. Golanski, M. 2007. Mesh Cluster Based Routing Protocol: Enhancing Multi-hop Internet Access using Cluster paradigm. The International Conference on Computer as a Tool, Warsaw, Canada, Pages: 962 - 965.

[5] Longxin Lin, 2010. Multicast routing algorithm for DSN network. The IEEE International Conference on Computer Application and System Modeling (ICCASM), Chiana, Pages:516-520.

[6] Bill K. 2005. Multicast Communications Supporting Shared Applications. Product Manager, Spirent Communications. URL: http://www.usipv6.com/6sense/2005/apr/05.htm.

[7] Yuanyuan Yang, et. al., 2008. A Service-Centric Multicast Architecture and Routing Protocol. IEEE Transactions on Parallel and Distributed Systems, Volume: 19, page(s): 35 – 51.

[8] D.Waitzman and C. Partridge, 1998. Distance Vector Multicast Routing Protocol. RFC 1075.

[9] J. Moy, 1998. Multicast Extension to OSPF. Internet draft.

[10] J. Moy, 1998. OSPF version 2. RFC 2328.

[11] A. Ballardie, B. Cain and Z. Zhang, 1998. Core Based Trees (CBT version 3) Multicast Routing Internet draft.

[12] S. Deering et al., 1998. Protocol Independent Multicast-Sparse Mode (PIM-SM): Motivation and Architecture. Internet draft.

[13] R. Perlman et al., 1999. Simple Multicast: A Design for Simple, Low-Overhead Multicast. Internet draft.

[14] A. Adams, J. Nicholas and W. Siadak, 2004. Protocol independent multicast - dense mode (PIM-DM): protocol specification. Internet draft.

[15] B. Waxman, 1988. Routing of multipoint connections. JSAC, Dec. , pp. 1617-1622.

[16] LI Xiao-yong , GUI Xiao-lin, 2006. Merging Source and Shared Trees Multicast in MPLS Networks. Seventh International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT'06), Taiwan, Pages: 23-28.

[17] The Network Simulator - ns-2, 2008, http://www.isi.edu/nsnam/ns/

Omar Said received his BSC degree in Computer Science from Menoufia University, Egypt in 1997; MS.C from Menoufia University, Egypt in 2002; Ph.D. degree from Menoufia University, Egypt in 2005. He is currently an assistant professor in Computer Science department, College of Computers and Information Technology, at Taif University, KSA. His research interests include Networking, Multimedia Communication, and Internetwork Protocols.

Ahmed Ghiduk is an assistant professor at Beni-Suef University, Egypt. He received the BSc degree from Cairo University, Egypt, in 1994, the MSc degree from Minia University, Egypt, in 2001, and a Ph.D. from Beni-Suef University, Egypt in joint with College of Computing, Georgia Institute of Technology, USA, in 2007. His research interests include software engineering especially search-based software testing, genetic algorithms, and ant colony, and image prossing. Currently, Ahmed S. Ghiduk is an assistant professor at College of Computers and Information Systems, Taif University, Saudi Arabia.

Dr. Sultan Aljahdali is the Dean of the College of Computers and Information Technology at Taif University. He received the B.S from Winona State University, M.S. with honor from Minnesota State University Mankato, and Ph.D., in Information Technology from George Mason University in Fairfax, Virginia. He is the recipient of the prestigious higher education scholarship from the government of Saudi Arabia for pursuing his B.S., M.S., and Ph.D., in a row. He is a member of professional societies like ACM, IEEE, Arab Computer Society (ACS) and International Society for Computers and Their Applications (ISCA).