

An Efficient and Reliable Fibonacci Tree Based Overlay Multicast Protocol^{*}

Jing LI^{*}, Mi WEN, Lin ZHOU, Mei XUE, Yong WANG

Dept. of Computer Science and Technology, Shanghai University of Electric Power, Shanghai 200090, China

Abstract

Efficiency and reliability are both important for overlay multicast to give the best service to end-users. In this paper, an efficient and reliable Fibonacci tree based overlay multicast protocol- RFOM is proposed on top of mesh overlays. In RFOM, it built a two-layer architecture based on a location-aware CAN mesh network. The novel cluster formation algorithm and cluster core selection algorithm is proposed by fully utilizing the properties of mesh networks. In addition, a novel Fibonacci tree approach and a forward flooding approach are proposed that address the multicast routing scheme. Moreover, the reliability of end host is considered when constructing the hierarchical architecture to improve the efficiency and reliability of RFOM. At last, the evaluations show that RFOM is efficient, reliable and scalable.

Keywords: Overlay Multicast; Mesh; Hierarchical Architecture; Fibonacci Tree; Reliable

1 Introduction

Because of its inherent drawbacks, IP Multicast lacks ubiquitous multicast support among all ISP. Overlay multicast offers a viable solution to overcome the limitation of IP Multicast availability. It builds a multicast architecture by having the end hosts self-organize into logical overlay networks. According to overlay topology design, current proposed overlay multicast protocols can be classified into three flavors: the mesh-first, tree-first and implicit overlay multicast protocols.

Take NARADA [1] as example of mesh-first protocol. NARADA firstly organizes the multicast group members into a mesh topology and then constructs a spanning tree whose root is the multicast source. To guarantee robust, every multicast group member maintains a state list about all members. But this condition compromises the scalability of NARADA.

YOID [2] is a case of tree-first overlay multicast protocol. It builds a shared data delivery tree among members. The tree structure has logarithmic scaling behavior. Its drawback is a direct control over every aspect of tree structure and this will result in high costs.

^{*}Project supported by the Innovation Program of Shanghai Municipal Education Commission Grant(No. 09YZ346)

^{*}Corresponding author.

Email address: lijing@shiep.edu.cn (Jing LI).

NICE [3] and CAN-based multicast [4] is two representatives of implicit overlay multicast protocol. CAN-based multicast is based on a special infrastructure-CAN. CAN is an overlay network whose constituent nodes form a virtual d-dimensional Cartesian coordinate space. The flooding scheme is used to multicast packets. NICE is hierarchical infrastructure. It involves several layers and each layer has a set of clusters. Each cluster has a cluster leader. So NICE is a scalable protocol. However, it does not consider underlying topology when running hierarchy and clusters. This compromises the delay performance.

Although there are many protocols proposed [5], few of them considered the reliability of end host. If an end host who serves an important role fails frequently, the protocol has bad efficiency and reliability performance. Therefore, designing a reliable and efficient application-layer multicast protocol is still an open problem. In this paper, a reliable and efficient overlay multicast protocol is proposed. It organized the group members into two-layer architecture. Group members all lie in member layer and they are partitioned into several clusters. Each cluster has a cluster core and all cores lie in core layer. In RFOM, it routes the packets in the two layers in parallel by novel Fibonacci tree approach and forwarding flooding approach.

The rest of the paper is organized as follows: Section 2 gives a description of RFOM design. The performance analysis is presented in Section 3. The evaluation is described in Section 4. Finally, in Section 5 we present the conclusions.

2 RFOM Design

In RFOM, group members are mapped into a topology-aware mesh. The whole mesh space is divided into several clusters and each cluster includes the mesh zones that are occupied by the members with the closeness relationship in terms of overlay hops. Each cluster has a cluster core. In the two-layer architecture, a degree-constrained flooding scheme and Fibonacci tree scheme are adopted as the multicast routing scheme to deliver packets. In addition, a new conception is introduced: local area. Local area is composed of the group members that attach to the same router directly or through several local network components (e.g. the hubs or switches).

2.1 Mesh architecture

In RFOM, we assume that group members have been mapped into a topology-aware CAN mesh. Without loss of generality, we suppose that an end host owing a zone is mapped onto the central point of the zone. We use the number of overlay hops to measure the length of the shortest path between two zones.

In addition, we assume the existence of an agent set. The function of each agent is to register the multicast group for the end hosts that send requests to it. A member list of the end hosts that have registered is maintained by the agent. After the members join in the group, end hosts are organized into a two-layer architecture. Fig.1 shows an example of such two-layer architecture. Fig. 1(a) gives the original network. Through some hashing function, end hosts can be mapped into the topology-aware mesh shown in Fig. 1(b). This overlay mesh is the lower layer of two-layer architecture which called member layer. Some selected end hosts will construct the upper layer named core layer in Fig. 1(c).

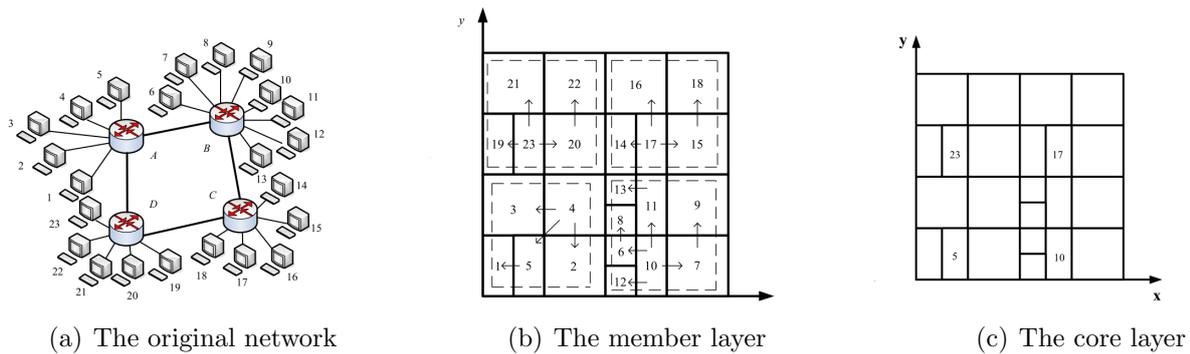


Fig. 1: An example of two-layer architecture.

2.2 Cluster formation

After the group members are mapped into a location-aware CAN mesh, they are assigned into several different clusters by using a cluster formation scheme. Firstly, some new conception is introduced: member selection distance unit D is the maximum number of overlay hops between the constructor (or sub-constructor); the cluster size bound S is the maximum number of end hosts that a cluster may contain.

One agent is initially selected through negotiating among all agents in agent set. The selected agent then select one unassigned end host at random from its member list. This selected end host becomes constructor of the cluster and achieves the zone information of group members. The constructor then starts the cluster construction and it selects the unassigned end hosts with the number of overlay hops not greater than D as its cluster members. If the constructor finishes its cluster member selection procedure due to it finds a member with the overlay hop greater than D , a sub-constructor will be selected to continue the current cluster construction. The current cluster formation will not terminate until all current cluster members cannot find any unassigned end host with the number of overlay hops not greater than D or the number of selected cluster members equals to the cluster size bound S . Then, it initiates the formation of another cluster among those unassigned end hosts. When there is no unassigned end host in group, the cluster formation completes.

The cluster core is responsible for forwarding the packets from the outside group members to the inside cluster members and distributing the inside cluster packets to other clusters. Hence, the selection of cluster core decides the efficiency and reliability of RFOM. Therefore we select the cluster member who has the minimum sum of delay time to all other cluster members as the cluster core. In the meantime, this chosen cluster core should have considerable reliability performance. If this core has low reliability, the end host with sub-minimum delay performance would be selected as the cluster core.

2.3 Multicast routing

In RFOM, packets are transmitted in parallel by different routing schemes. A degree-constrained flooding scheme is adopted in the member layer and a Fibonacci tree based scheme is adopted in core layer for the short delay performance.

Firstly we introduce degree-constrained flooding scheme. Denote n_v as the number of cluster

neighbors (i.e., the cluster members in the zones with 1 overlay hop to v) of any cluster member v . In member layer, if the cluster core c satisfies $n_c < D_c$ (D_c denote the maximum number of direct receivers of c), it forwards the outside packets to all its cluster neighbors; otherwise, it only forwards the packets to the closest $D_c - 1$ neighbors for reserving the capacity to forward packets to one direct receivers in the core layer. Those $n_c - D_c + 1$ neighbors may receive the packets from their other cluster neighbors. The neighbors received packets from c continue forwarding the packets to their neighbors who haven't received the packets before by the same way under considering the capacity constraint of end hosts.

All cluster cores in member layer consists the core layer. All cores are organized into a Fibonacci tree and packets are delivered along this Fibonacci tree. The Fibonacci tree is achieved by adopting the Fibonacci series based multicast algorithm (Algorithm 1). Before the Algorithm 1 is performed, the cluster cores should be first organized into a member sequence. The source serves as the root of the Fibonacci tree.

Because the capacity of end host is not comparable with router, the packet processing delay of end host is considerable and this factor is introduced in RFOM. In addition, the reliability of end host is also taken into account to achieve better reliability and delay performance. The cores are organized into the sequence by using the following regulations. The core named m_i should maintain three parameters: 1) d_i : the delay distance of logic link from itself to the root; 2) deg_i : the degree constraint of core (the value is assigned initially); 3) l_i : the packet processing delay of member m_i ; and 4) r_i : the reliability of core m_i (the value is also assigned initially). The parameters $d_{max}, deg_{max}, l_{max}$ and r_{max} denote the maximum of d_i, deg_i, l_i and r_i respectively. A new parameter w_i is constructed by using the four parameters above. $w_i = \alpha \frac{d_i}{d_{max}} + \beta (1 - \frac{deg_i}{deg_{max}}) + \gamma \frac{l_i}{l_{max}} + (1 - \alpha - \beta - \gamma) (1 - \frac{r_i}{r_{max}})$, where α, β and γ are balance factors. In our experiment, we would like to give greater weight to delay distance and reliability. The value of w_i is used as the weight of these members. Then these members are sorted into a sequence (i.e. the input of the algorithm) with an ascending order of the w_i value. This guarantees that the least time-cost and the most reliable end host is treated first while the outgoing links of nodes are utilized adequately. It makes the better efficiency and reliability performance.

In decision of members weight, besides the delay distance and the degree constraint together with packet processing delay of end host, it also considers the reliability of end host. This makes the protocol reliable and efficient.

Algorithm 1 adopts the idea of Fibonacci series to partition the members sequence into parts with different sizes. The Fibonacci series $\{f_i\}$ satisfies the following condition: $f_0 = 0, f_1 = 1; f_n = f_{n-1} + f_{n-2}, \text{if } n > 1$.

Algorithm 1 (Fibonacci series based multicast):

Input: member sequence $\phi = (d_1, d_2, \dots, d_K), d_s$ is the core which serves as the source node in ϕ . The number of members in ϕ is $K. f_n \leq K < f_{n+1}$

Output: a multicast tree constructed for all members in ϕ

1 If $K = 2$, d_s sends packets to the only destination;

2 If $K > 2$, ϕ is partitioned into two subsequences ϕ_1 and ϕ_2 where d_s is in the larger subsequence and the smaller one includes f_{n-2} members;

2.1 If $s > f_{n-2}$, then $\phi_1 = (d_1, d_2, \dots, d_{f_{n-2}}); \phi_2 = (d_{f_{n-2}+1}, d_{f_{n-2}+2}, \dots, d_K);$

Else $\phi_1 = (d_1, d_2, \dots, d_{K-f_{n-2}}); \phi_2 = (d_{K-f_{n-2}+1}, d_{K-f_{n-2}+2}, \dots, d_K);$

2.2 If $s > f_{n-2}$, d_s firstly sends packets to d_1 , then d_1 is in charge of multicasting in ϕ_1 and d_s is in charge of multicasting in ϕ_2 ;

Else d_s firstly sends packets to $d_{K-f_{n-2}+1}$, then d_s is in charge of multicasting in ϕ_1 and $d_{K-f_{n-2}+1}$ is in charge of multicasting in ϕ_2 ;

3 Multicast packets from d_1 to all members in ϕ_1 and from d_s to all members in ϕ_2 (or multicast packets from d_s to all members in ϕ_1 and from $d_{K-f_{n-2}+1}$ to all members in ϕ_2) by recursive calls Algorithm 1.

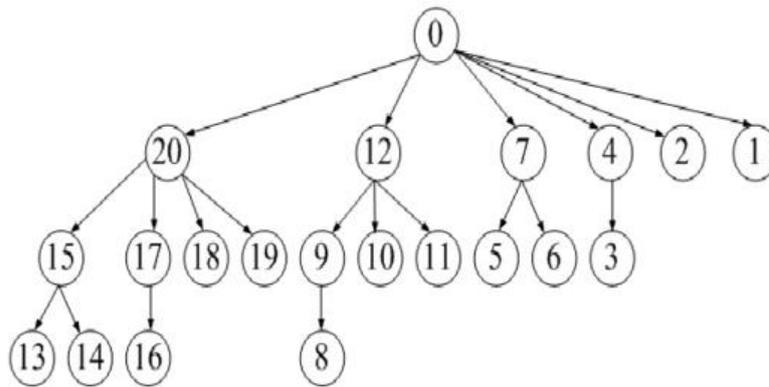


Fig. 2: The multicast tree achieved by Algorithm 1 on sequence (20,19,18,...,1,0)

Fig.2 shows an example of Fibonacci multicast tree achieved by this algorithm based on member sequence (20,19,18,...,1,0).

3 Architecture Analysis

In this section, we illustrated the Fibonacci tree by using the Fibonacci series base multicast protocol is efficient. In the process of Fibonacci multicast tree building, each member in sequence is processed once. Obviously, the time complexity of tree building is $O(K)$, where K is the number of members in the initial input sequence.

The following is the analysis of time complexity in the process of multicast packets to K members by using the Fibonacci series based multicast algorithm. L denotes the propagation time on an overlay link and t is the packet processing time on each end host. For simplicity, L is supposed to be same for each overlay link. $T(K)$ denotes the time used to multicast a packet from one source to K destinations. Obviously, $T(K)$ is a monotonous increasing function of K . When packet multicast begins, the source firstly sends the packet to the partition node. It takes $L + t$ seconds to deal with the packet. As a result, the initial member sequence is divided into two sub-sequences.

So it takes $O(\log K)$ time to multicast a packet to a member group by using Algorithm 1, where K is the number of members. It illustrated that this algorithm is efficient in term of delay performance. Moreover, the consideration of reliability of end host also improved the efficiency and reliability of RFOM.

$$\begin{aligned}
 T(K) &= \max\{T(f_{n-2}) + L + t, T(K - f_{n-2})\}, T(1) = 0, T(2) = L + t \\
 \text{If } f_n \leq K < f_{n+1}, T(K) &= \max\{T(f_{n-2}) + L + t, T(K - f_{n-2})\} \\
 &\leq \max\{T(f_{n-2}) + L + 2t, T(f_{n+1} - 2f_{n-2})\} \\
 f_{n+1} - 2f_{n-2} &= f_{n-1} + f_{n-3}, \text{ we obtain } f_{n-1} \leq f_{n+1} - 2f_{n-2} < f_n \\
 T(K) &\leq \max\{T(f_{n-2}) + L + 2t, \max\{T(f_{n-3}) + L + 2t, T(f_{n-1})\}\} \\
 &= \max\{T(f_{n-2}) + L + 2t, T(f_{n-1})\} \\
 \text{In addition } T(f_{n+1}) &\leq \max\{T(f_{n-2}) + L + 2t, T(f_{n-1})\} \\
 T(f_{n-1}) &\leq \max\{T(f_{n-4}) + L + 2t, T(f_{n-3})\} \\
 T(f_{n-2}) &\leq \max\{T(f_{n-5}) + L + 2t, T(f_{n-4})\} \\
 \text{If } f_n \leq K < f_{n+1}, T(K) &\leq \max\{T(f_{n-2}) + L + 2t, T(f_{n-1})\} \\
 &\leq \max\{T(f_{n-5}) + 2L + 4t, T(f_{n-3})\} \\
 &\dots \\
 &\leq \left\lfloor \frac{n}{2} \right\rfloor * (L + 2t) \\
 \text{Moreover } f_n &= \left\lfloor \frac{\psi^n}{\sqrt{5}} + \frac{1}{2} \right\rfloor, \psi = \frac{1 + \sqrt{5}}{2} \\
 \text{So } n &= 1.44(\log f_n - 1.66), T(K) \leq \lfloor 0.72 \log K \rfloor * (L + 2t)
 \end{aligned}$$

4 Simulation Evaluation

4.1 Model design

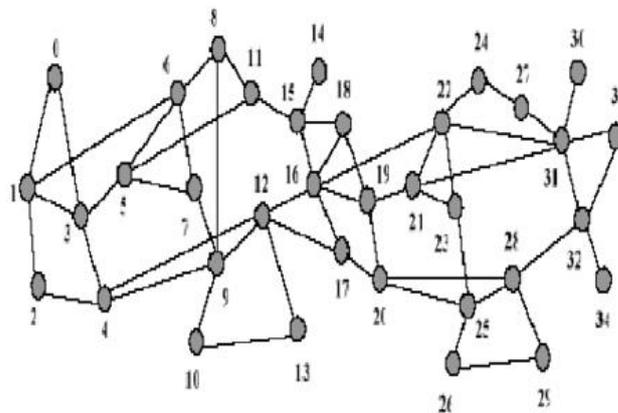


Fig. 3: The MCI ISP backbone network used in the simulation

We use the tool NS-2 to accomplish our simulation experiments. The backbone network in the simulation is shown in Fig. 3. It is the well-known MCI ISP backbone. The end hosts in multicast group are connected to routers directly or indirectly. The bandwidth of links in the backbone network is 1000Mbps and that in the local area is 100Mbps. The cost of each link of backbone network is a random integer between 20 and 40 and that in the local areas is a random integer

between 1 and 4. Each node gets a reliability assigned from the range of 0 1. The simulation traffic is the 1.5Mbps MPEG-1 video streams.

In the simulation, performance comparisons are given among RFOM, HFTM [6] , NICE and CAN-based multicast. The following metrics are used: 1) Average Link Stress (ALS): the ratio of the sum of times that identical packet copies traverse over the underlying links to the number of links in the group; 2) Average End-to-end Delay (AED): the ratio of the sum of end-to-end delay from a multicast source to each group member to the number of group members; 3) Number of Links Used (NLU): it refers to the total number of links that are used during the multicast communication.

4.2 Comparison results

Two simulations are done to compare different protocols along the ALS, AED and NLU metric under single source and multiple sources occasion. In the first simulation, the number of group members varies from 70 to 1015. The number of sending sources is 1.

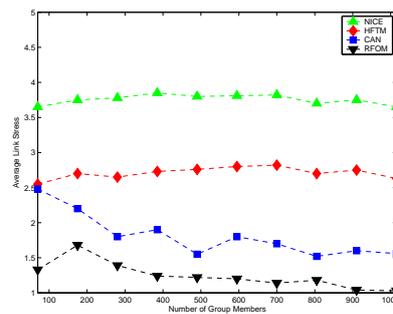


Fig. 4: The ALS performance of four protocols

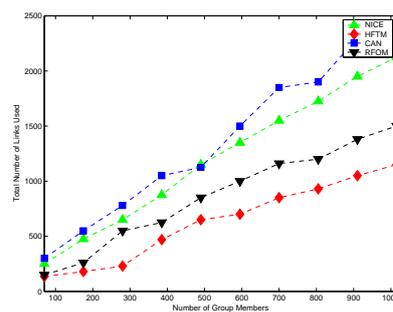


Fig. 5: The NLU performance of four protocols

Fig. 4 illustrates the comparison of ALS performances. The curves show that NICE incurs the worst ALS among the four protocols. This is expected due to the fact that cluster leaders in NICE forward the packets to all the cluster members in the same cluster, which causes that the links near the leaders carry the identical packet many times. CAN-based multicast has better ALS performance than NICE because the flooding routing scheme of CAN-based multicast can evenly distribute the link stress among all overlay hops. Moreover, RFOM has better ALS performance than CAN-based multicast because the members are divided into several clusters which enable some members to have less cluster neighbors.

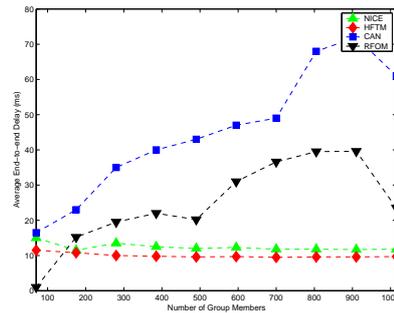


Fig. 6: The AED performance of four protocols, single source

Fig. 5 illustrates the comparison of NLU performance. We can see that CAN-based multicast consumes the most physical links during the multicast communication. The cluster and the Fibonacci tree routing scheme of RFOM incurs its better NLU performance. In summary, Fig. 4 and Fig. 5 shows that RFOM has better ALS and NLU performance because it has the novel cluster and two-layer hierarchical architecture. In addition, the consideration of end host reliability enables RFOM to have better performances and it can work well in large scale multicast group.

Fig. 6 illustrates the AED performance comparison among four protocols. From the diagram, we can see that RFOM has better AED performance than CAN-based multicast and such trend becomes more obvious when the group size becomes large. It is mainly because the two-layer architecture and selection of cluster core decrease delay time while multicasting in each cluster. The construction of Fibonacci multicast tree and degree-constraint of end host also enable RFOM better AED performance.

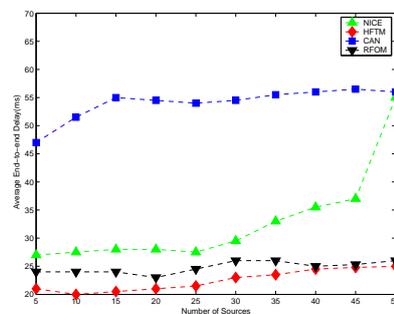


Fig. 7: The AED performance of four protocols, multiple sources

In the second simulation, the AED performances comparison among the four protocols are observed under the situation that the number of the sending source varies and the number of group members is always 490. The additional sources are determined randomly.

Fig. 7 illustrates the AED performances of the four protocols when the number of sending source varies from 5 to 50. The curves show that CAN-based multicast has worst AED performance because of its flooding routing scheme. In multiple sources occasion, RFOM has better AED performance than NICE and this advantage becomes obvious especially when there have more multicast sources. The properties of CAN mesh and the construction of two-layer architecture in RFOM make that the data packets take less delay to reach group members than in NICE.

The second simulation illustrates that RFOM have good AED performance with the increasing number of sources in the multi-sources cases and so they are scalable and efficient.

The simulations illustrate that RFOM are reliable and scalable in term of ALS metric and NLU metric. Moreover, it is efficient in term of AED metric. Therefore, RFOM is scalable, efficient and reliable.

5 Conclusions

In this paper, an efficient, reliable and scalable overlay multicast protocol-RFOM is proposed based on mesh topology. In RFOM, there exists a two-layer architecture. All group members are mapped onto a location-aware CAN mesh named member layer. They are partitioned into different clusters and each cluster has a cluster core. All cluster cores construct the core layer. An efficient degree-constrained flooding scheme is utilized to delivery multicast packets on member layer. On core layer, packets are delivered along a novel Fibonacci tree structure. Moreover, the consideration of degree-constraint and reliability of end host enable RFOM to have better ALS and AED performance. The evaluations show that RFOM is a reliable, efficient and scalable protocol and it works well, especially in large scale network.

Acknowledgement

The research described here is supported by Innovation Program of Shanghai Municipal Education Commission Grant No.09YZ346, the National Natural Science Foundation of China under Grant No.60903188.

References

- [1] Y. H. Chu, S. Rao, S. Seshan, and H. Zhang. A Case for End System Multicast. Proc. of ACM SIGMETRICS 2000, Santa Clara, California, USA, June 2000: 1-12.
- [2] P. Francis. Yoid: Extending the multicast internet architecture. White paper <http://www.aciri.org/yoid/>, April 2000.
- [3] S. Banerjee, B. Bhattacharjee, and C. Kommareddy. Scalable Application Layer Multicast. ACM SIGCOMM'02, Pittsburgh, August 2002: 205-217.
- [4] S. Ratnasamy, M. Handley, R. Karp, and S. Shenker. Application-Level Multicast Using Content-Addressable Networks. Proc. of the 3rd International Workshop on Network Group Communication, London, UK, November 2001: 14-29.
- [5] M. Brogle, L. Bettosini, and T. Braun. Quality of service for multicasting in content addressable networks. In 12th IFIP/IEEE International Conference on Management of Multimedia and Mobile Networks and Services (MMNS 09). Springer LNCS, October 2009.
- [6] Jing Li, Mi Wen, Yong Wang, Wei Zhang. Quality of Service Enabled Multicasting Protocol for Overlay Networks. Journal of Computational Information Systems, Vol 6, No. 2: 575-583, 2010.