Content-based Routing For ebXML Messages in E-Commerce

Felor Mohammadi
Department of Computer Engineering
South Tehran Branch,
Islamic Azad University
Tehran, Iran

Mir Ali Seyyedi
Department of Computer Engineering
South Tehran Branch,
Islamic Azad University
Tehran, Iran

Afshin Salajegheh
Department of Computer Engineering
South Tehran Branch,
Islamic Azad University
Tehran, Iran

ABSTRACT
The main component of electronic commerce (E-commerce) is the exchange of documents and messages. In the field of commercial messages standardization, the ebXML framework offers the valid standard for message modeling.

Due to the importance of standardization ebXML framework in E-commerce, creating more facilities, in the ebXML messaging is so important. Today, the exchange of messages within the ebXML framework uses explicit addressing, in which, the message ultimate recipient is determined based on message producer partner and the information existed in collaboration agreement. In case of explicit addressing Use, among the messages that are reached by the partners in the business process, only a few of them may be appropriate to the needs of the recipient and as a result the recipient is forced to discard some of them. Due to releasing a large number of relatively high-volume messaging in system, not only the bandwidth consumption will rise, but it also slows down the message precision regarding to the message received by the partner that is appropriate to its needs or not. The purpose of this paper is providing a method in order to convert an explicit addressing to the implicit addressing for ebXML messages. The proposed method is based on an algorithm called "Rendezvous", that's one of the algorithms which have been widely used in networks based on contents and specifically in publish/subscribe systems. At the end of work though, bandwidth parameters, speed and precision in both traditional ebXML messaging method and the proposed messaging method are compared.

General Terms
Content-based routing for ebXML message

Keywords
ebXML message, Partner, Content-based routing, Publish/subscribe, Rendezvous

1. INTRODUCTION
Ebxml framework provides credible standards for messaging in the business to business collaboration space. Today, the exchange of messages in ebXML framework uses explicit addressing, in which the final destination or the final message recipient is determined by message producer partner and based on the information existed in agreement that's exchanged between both partners , and is set in each message.

The purpose of this paper is using the implicit addressing rather than explicit addressing and trying to enter explicit addressing into ebXML framework by using concepts and algorithms related to content-based routing that's considered in publish/subscribe systems. Content-based routing refers to the routing form in which messages are routed based on their content or profits and consumer's interests [1]. An example of an application for content-based routing is publish/subscribe event notification service that provides facilities either implicitly or asynchronously for transferring data from message producer to all partners that expressed their interests for that message[2,3].

Using content-based routing has many advantages among which the following items can be mentioned:

1. Since messages are sent only to the actual consumers, network traffic is reduced [4].

2. It is appropriate for distributed systems with a large number of consumers with diverse interests.

3. Consumers with few resources can limit their input network traffic by registering their interests for receiving a special type of massage [1].

4. Determining route of messages based on the interests of consumers, causes messages which the recipient's interests for receiving them are not recorded in the system not to be published and this will result in system efficiency [5].

In this paper we're attempting to add some facilities such as reducing consumption of bandwidth in terms of the total number of exchanged messages in the system and also increasing the accuracy, to the traditional messaging model in ebXML framework. Obviously, if the numbers of messages that are exchanged in the system are reduced, consumable bandwidth will be also reduced. Also in terms of accuracy, two parameters are considered, the first parameter is checking accuracy of the number of partners who are looking forward to receive messages and so the expected message will be received, the second parameter, is checking the accuracy so that partners will not receive inappropriate message.

2. 2. ebXML framework
In early 1999, members of the center UN/CEFACT[1] give theories for producing a global XML framework for electronic business, they were associated with a group in the OASIS[2]

[1] United Nations Centre for Trade Facilitation and Electronic Business
[2] Organization for the Advancement of Structured Information

center for forming this standard, and after 18 months and by attending more than 120 companies and standards, the first generation of ebXML was passed in may 2001, and its first infrastructure delivery has begun [6].

Perspective of ebXML, is creating a global electronic unit market by which companies of any size and in any location could meet each other and do e-commerce together through exchanging messages based on XML [7]. The main purpose of ebXML messaging service (ebMS) is to facilitate electronic exchanging of business messages in XML framework that uses common standard of internet. It determines ebMS, message packet, protocol and messaging functions which are considered for business on SOAP [9]. The ebXML message service is conceptually divided into three sections as below:

- Message service interface
- Given functions by message service handler (MSH)
- Mapped to the underlying transport services[10]

EbMS messages are packaged as a SOAP 1.1 message or SOAP 1.2 and are independent of communication protocols. Figure 1displays the general structure of an ebMS Message [9].

3. CONTENT-BASED NETWORK

Content-based network is a communication network which shows a new communication model called content-based communication. By using a content-based communication service, the recipients notify their interests through selective predicates and the message are delivered to the recipients that the election statement by the recipient matches with the message content statement [1,11].

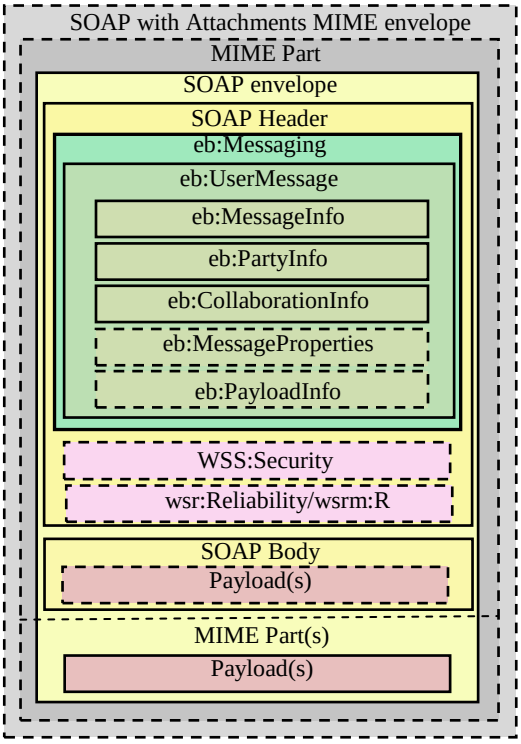

Figure 1: ebXML message structure

Content-based communication paradigm is appropriate for a variety of application domains including: event notification publish/subscribe, system monitoring, service discovery, data sharing, distributed electronic auction and so on [5]. According to the definition which is in references [1], [2], [12] and [11], content-based routing refers to the form in which, messages have not any explicit destination addresses, and messages destination is determined based on their content.

3.1 Publish/Subscribe event notification

In a publish /subscribe communication system, each of the participants can have a role of publisher or subscriber. Publishers produce information as events, these events are consumed by subscribers who indicate their interests for special events by using subscriptions that export. Figure 2 shows a high-level model of interaction of a publish/subscribe system with its clients [12]. Publish/subscribe systems are divided into two categories: Topic-based and Content-based.

In topic-based publish / subscribe systems, events are classified based on their subjects, while in content-based systems, subscribers express their interests to receive events by using the filters on the content of the events.In general, in these systems, event routing algorithms are divided into two categories, filter-based and rendezvous-based[12].

Standards

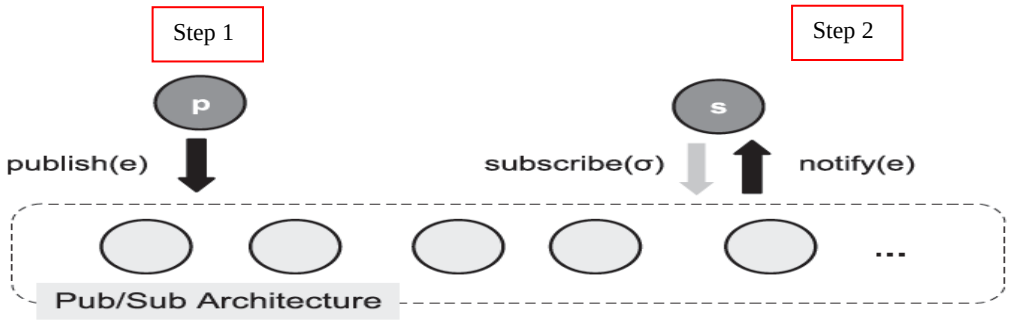

Figure 2: High-level interaction model of a publish/subscribe system with its clients (p and s indicate a generic publisher and a generic subscriber respectively)

4. PROPOSED METHOD OF CONTENT-BASED ROUTING FOR EBXML MESSAGES

In traditional method messaging within ebXML framework, to conduct business collaboration with other partners, the destination of the ebXML messages are determined explicitly and based on the information in collaboration agreement between business partners. In other words, in this method, the receiver of a message which is generated and sent, is specified by its sender. In the proposed method that's provided for content-based routing in ebXML messages, the concept called rendezvous nodes is used. Also in this method, in addition to the original message, announcement, subscription and match messages are also exchanged. Figure 3 displays the general structure of proposed method.

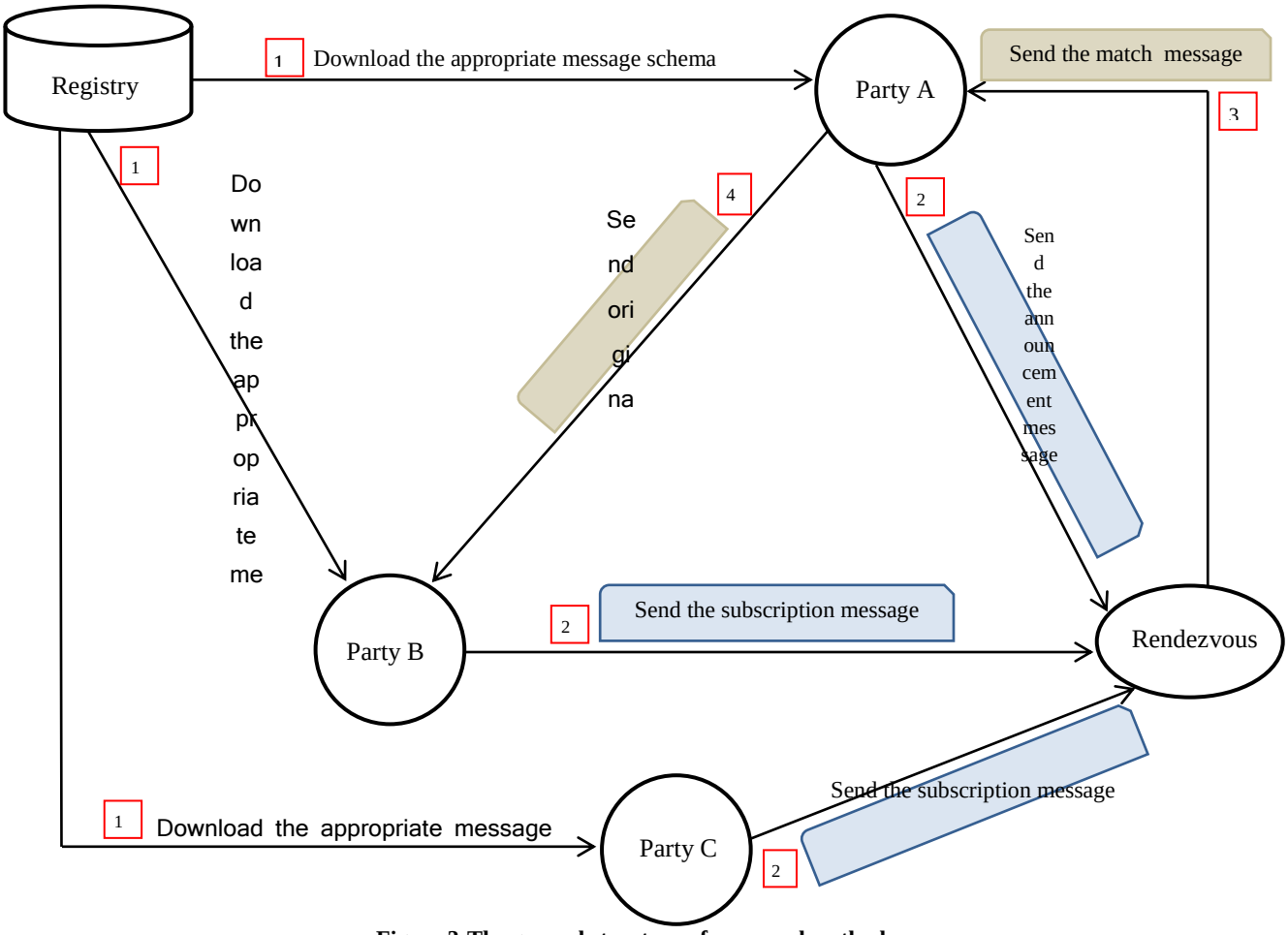

Figure 3:The general structure of proposed method

According to figure 3, in this method at first, for each type of messages that is supposed to be exchanged in the system, a message schema is defined and a rendezvous node can also be assigned to it. Information relating to each of nodes that meet a specific message type, and also defined message schemas are as well in a registry. Each partner that generate an ebXML messages, at first before sending it, based on appropriate pattern of the message that is in the registry, generates a message called announcement message and then sends this message which refers to a part of the original message content to rendezvous node related to its supposed message.

Other partners that are also interested in receiving a message, generate messages as subscription message and send it to the related rendezvous node based on the patterns of an appropriate message that are in registry. In fact, subscription messages, shows the interest of each partner to receive messages with a specific content. Then in rendezvous node, comparison operation between announcement messages and subscription messages will be done and a list of partners who are interested to receive messages with announced content by the original message producers, will be sent to the original message producer in the form of match message. The producer of the original message extracts the ID of recipient's partners from the match message, put it in the original message and send it. Thus the destination of generated message is determined based on its content and the announced interests by each partner.

5. ASSESSMENT PROPOSED METHOD

By using three parameters: bandwidth consumption, speed and accuracy, the proposed method will be evaluated.

5.1 Bandwidth consumption

Bandwidth consumption is studied of the total number of messages that are published [13]. In the traditional method, only the original message which is the same as ebXML message included business document, is exchanged. Whereas, in the proposed method, according to the presented solution, in addition to the original message, announcement, subscription and match messages are also exchanged.

One of the reason of releasing additional original messages in the system is that, in a business collaboration the agreement which is exchanged between the partners and is based on their profiles, is only agreed upon some certain features such as business process name, role of each partner that performs in the business process and so on, and there's no reference to the content of the document which is supposed to be exchanged. In this condition, it may occur that the message which is received by one of the partners and after it's processing won't be appropriate to its actual needs and the partner will discard the message. In this case it can be said that, ebXML message is additionally released in the system and it has used the bandwidth ineffectively.

According to the feature of proposed method that the destination of a message is determined based on its content and by performing match operation between announcement messages and subscription messages, therefore in compared with the traditional method, as far as possible the original messages are published to their actual users in the system.

5.2 Accuracy

In general, the accuracy is defined as proportion of the number of partners who received the appropriate message to their needs and interests, to the total number of partners who have received the message. Since in the proposed method, by recording subscription messages the partners define filters on message content and the producer partner also by rendezvous node is notified that which of the partners are interested in receiving a message with produced content by him, it's expected that in proposed method in compare with the traditional method, accuracy of algorithm in which the number of partners who receive the message are real consumer and won't throw it away, will increase.

5.3 Speed

Speed is defined as the time distance from the moment a partner generates a message, to the moment the message is sent. In the traditional method, when a partner generates a message, immediately sends it after determining the target. Whereas in the proposed method, before sending the message, first the announcement message should be sent to the related rendezvous node, after doing match operations between announcement message and subscription messages, and a list of interested partners to the produced message is provided and sent for exporting partner of announcement message in match message format. After receiving this message by the original message producer, the receivers are defined based on existed IDs in match message and then the specified message will be sent to the intended receivers. Obviously, according to given definition, the purpose of speed is messages delivery rate.

It is concluded that in the traditional method, the time is only spent to produce the original message, in order that the last recipient receive the generated message, while in the proposed method, since a partner generates a message to send it, the spent time is as much as the total time taken on generating announcement message, time taken for matching in rendezvous node, sending match message to message producer and also required time to send the original message so that the generated message will be reached to the recipient partner.

6. SIMULATION

In order to measure each of these metrics, a simulation environment is used. Simulation environment is designed and performed by using programming language C # and in the environment Microsoft Visual Studio 2012. Through this simulation environment, both the traditional and the proposed method have been evaluated in a quite similar environment. In order to compare the bandwidth consumption, speed and accuracy between the traditional and the proposed method, the simulation is performed through 5 round and with different numbers of participants in each round. Some of the key parameters which are considered constantly in each program running are as follow:

The number of process and the rendezvous nodes = 3

Total number of attributes = 10

The number of producing partners of announcement message = 30% of total partners

The number of producing partners of subscription message= 70% of total partners

6.1 Evaluation of bandwidth

As mentioned in section 5.1, bandwidth consumption is evaluated based on the total number of exchanged messages. Table 1 displays comparatively the total numbers of exchanged messages in both methods. It should be noted that in the proposed method the total number of messages means the original, announcement, subscription and match messages.

Table 1. Total number of exchanged messages between the traditional and the proposed method (Tm:Traditional Method, Pm: Proposed Method)

Iteratio / Partner	1	2	3	4	5	Average
100 (Tm)	797	1171	969	977	584	899.6
100 (Pm)	182	176	165	171	170	172.8
200 (Tm)	2901	3016	2892	2451	2365	2725
200 (Pm)	476	593	464	490	516	507.8
300 (Tm)	5401	5766	5308	5274	6048	5559.4
300 (Pm)	729	856	820	937	858	840
400 (Tm)	10019	11131	11617	10401	9028	10439.2
400 (Pm)	1403	1462	1289	1311	1534	1399.8

As the results indicated in Table 1, although the proposed method, in addition to the original message, four types of other messages are also exchanged, but the average of exchanging messages in the proposed method is lower than the traditional one. This is because, since in the proposed method partners recipient are determined based on their contents, therefore the number of exchanged original messages makes a huge difference rather than the traditional method. So, can conclude that bandwidth consumption in terms of the total number of exchanged messages, in the traditional method is more than the proposed method. Figure 4 displays the results shown in table 1 comparatively.

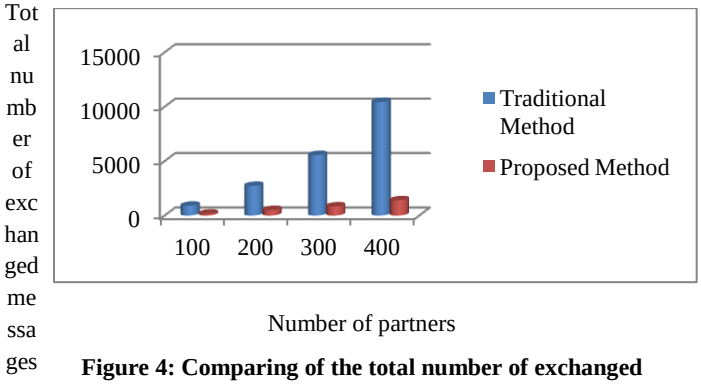

Figure 4: Comparing of the total number of exchanged messages between the traditional method and the proposed method

As it can be seen in the chart above, the blue bars show total numbers of the exchanged messages in the traditional approach and the red bars show total numbers of the exchanged messages in the proposed approach. Regarding to the results shown in table 1 and provided analysis for it, there's a noticeable difference between the exchanged messages in both approaches which is due to performing the action of matching in rendezvous node, that prevents extra sending of the main message of ebXML in the proposed approach and it causes to reduce the number of exchanged messages in this approach. So according to the given definition for bandwidth consumption, it can be concluded that in the proposed approach, the bandwidth consumption is reduced.

6.2 Evaluation of accuracy

To evaluate the accuracy of the proposed method and the traditional method, two parameters are considered. The first parameter, is likely the partners who are waiting for a message, receive the message. The second parameter is likely that a partner does not receive an appropriate message to its needs or in other words, receives inappropriate message to its needs.

6.2.1 Evaluation of accuracy regard to the first parameter

In the traditional method because the produced message, is sent to a wider range of partners, we can conclude that in this method, the probability that a produced message won't be received by a partner who is waiting for, is almost zero. While in the proposed method, because this possibility depends on match function used in rendezvous node, and since the accuracy of match function's not one hundred percent, therefore, some partners are possibly waiting to receive a message and that message won't be get into their hands. In summary it can be said that in the proposed method, the probability that a message won't be received by a partner who's waiting for it, is a positive number and greater than zero. So in terms of this parameter, the traditional method will be more accurate than the proposed method.

6.2.2 Evaluation of accuracy regard to the second parameter

To evaluate the accuracy regard to the second parameter, the following assumptions have been considered at first:

- ✓ M = total number of messages of a specified type. (E.g. Tender Ad.)
- ✓ N = number of messages that belong to a particular group. (e.g. belong to Department of Information Technology)
- ✓ P (A) = possibility that a partner receives an inappropriate message to its needs.
- ✓ P (B) = possibility that a partner receives an appropriate message to its needs.

Also the assumption is that a partner is waiting for a message from the IT Department.

In the traditional method, the probability that a partner receives an inappropriate message to its needs is:

P (A) = 1 - P (B) = 1 - (N / M)

Where, (N / M) << 1, and therefore P (A) tends to a large number.

In the proposed method, the probability that a partner receives an appropriate message to its needs is: P (A) = 1 - P (B)

Because the match function may not work one hundred percent accurately, it will be P (B) ~1, but since this number tends to one, P (A) will be a small number. According to these evaluations, it'll be concluded that if it's important that all the partners receive a message which were waiting for in a message exchanging, consequently, the traditional method would perform better and more accurately in this case. While if it is important that partners do not receive inappropriate message, the proposed method will be more accurately than the traditional method.

6.3 Evaluation of speed

According to the definition given in section 5.3 for speed and in order to find the required time for generating each of the original, announcement and match messages, the averaged method is used. To calculate the time related to the original, announcement and match messages, generate a specified number of each message by using a Loop in the program, and production time of each of them is calculated and finally by using averaged operation an average time is obtained for each messages. In order to calculate the time taken for the match operation and sending match message, with 20 round of program running, it was observed that in each round, 34 match messages are generated in average, and again we observed 2 partners' ID in these messages, average spent time was recorded as 0.0653 seconds. (It is recommended to reduce percent of the time due to simulation nature and subscription operation on collections and performing match operation so that it does not exist in the real world. However, base of calculations is considered approximately, and as the mentioned form) Thus we can conclude that the time taken to perform the match operation and send 680 kinds of these messages is equal to 0.0653 seconds.

Table 2 displays the average time taken with 1,000,000 times round for the original and announcement messages and 680 rounds for match message and match operation related to it, and also obtained time with each message.

Table 2. The average times obtained for different types of messages

Type of Message	Number of Message	Time (Second)	Time for a message(μs)
Original message	1,000,000	0.115	0.115
Announcement message	1,000,000	2.685	2.685
Match message and matching operations	680	0.0653	96

Thus in evaluating speed parameter it can be concluded that in the traditional method, since a partner generates a message up to the moment that send it, the time taken is 0.115 μs, while in the proposed method with regard total time spent for announcement message, matching operations, and match message and original message, the time taken is 98.8 μs. The outcome of evaluation above, is shown in Figure 5.

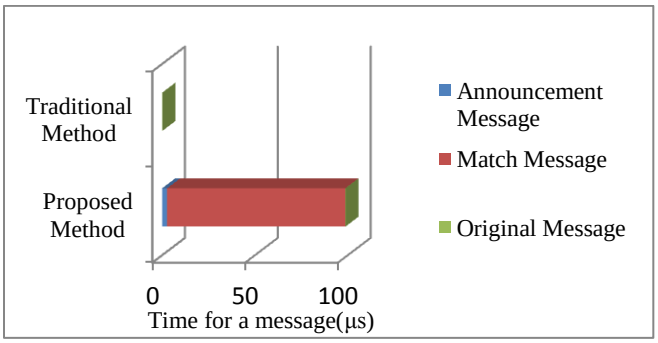

Figure 5: comparison the time used for the original message

According to the above description the important point is that nevertheless in proposed method sending time of messages is increased but due to message recipient partners, receive less messages, thus processing time taken in these nodes will be reduced and partners' speed in processing the messages will be increased. It should be noted that in all the above calculations, bandwidth is considered unlimited.

7. CONCLUSION AND FUTURE WORK

Regarding the importance of standardization of business messages and documents in e-commerce and by considering the ebXML framework as a valid standard for message modeling and because of the role and importance of this standard in e-commerce, creating greater convenience is so important in the traditional messaging method in this standard.

Provided solution in order to create content-based routing in ebXML messaging, is based on an algorithm called the "rendezvous" which is a known content-based routing algorithm in event notification systems or publish / subscribe systems. In the proposed method, in addition to enter the logic usage of a partner (node) as rendezvous, also four new messages, as schema, announcement, subscription and match messages are added to the ebXML messaging framework. As the evaluating of proposed method and simulation results in the previous sessions show, however in the proposed method, some extra messages, like announcement, match and subscription are also exchanged, but because the destination of messages is determined by their contents, numbers of the original published messages are noticeably reduced the strategy and this causes the reduction of consumption bandwidth in the proposed method. The accuracy in the proposed method in terms of the partners do not receive inappropriate messages has been increased, and it has been decreased in terms of all the nodes that are waiting to receive a message, receive it. About speed parameter, we observe the reduction of speed messages in the proposed method rather than the traditional method, it is because in the proposed method in comparison with the traditional method, from the moment that a partner prepare a message to send to the moment it is sent, more levels and therefore more time will be spent.

One of the proposals to reduce the volume of operations in rendezvous node can be creating covering between subscription and announcement messages recorded in rendezvous node. It means that among one or some of announcement and

subscription messages that cover each other, only one of them will be kept in rendezvous node. Since in the proposed method the fault tolerance category in rendezvous node is not issued, providing a solution to support fault tolerance in this node, is one of the future work in this case. Also in terms of announcement and subscription messages and match function definition, anthology concepts can be used.

8. REFERENCES

[1] R. Chand, P.A. Felber:" A Scalable Protocol for Content-Based Routing in Overlay Networks, "Network Computing and Applications, NCA 2003. Second IEEE International Symposium.

[2] Antonio Carzaniga, David S. Rosenblum, Alexander L. Wolf: "Content-Based Addressing and Routing: A General Model and its Application," University of Colorado/ Department of Computer Science/ January 2000.

[3] Yuanyuan Zhao, Sumeer Bhola, Daniel Sturman: "A General Algorithmic Model for Subscription Propagation and Content-based Routing with Delivery Guarantees," Technical report, RC23669, IBM Research, 2005.

[4] Antonino Virgillito, Roberto Beraldi, Roberto Baldoni: "On Event Routing in Content-based Publish/Subscribe through Dynamic Networks, " Distributed Computing Systems, 2003.The Ninth IEEE Workshop on Future Trends of.

[5] J. Legatheaux Martins, Senior Member: "Routing Algorithms for Content-based Publish/Subscribe Systems," Communications Surveys & Tutorials, IEEE (Volume: 12, Issue: 1), 2010.

[6] Eric Chiu: "ebXML Simplified-A Guide to the New Standard for Global E-Commerce,"2002.

[7] http://www.oasis-open.org, "ebXML Messaging Version 3.0 Part 1, Part 2 and AS4, "2011.

[8] Christian Huemer: "ebXML: An Emerging B2B Framework, " University of Vienna, Institute of Computer Science ,"2002.

[9] OASIS ebXML Messaging Services, Version 3.0: Part 1, Core Features, OASIS Standard, October 2007.

[10] Message Service Specification, Version 2.0, OASIS ebXML Messaging Services Technical Committee, April 2002.

[11] Antonio Carzaniga, Matthew J. Rutherford, Alexander L. Wolf: "A Routing Scheme for Content-Based Networking" Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies (Volume: 2), INFOCOM 2004.

[12] Roberto Baldoni, Leonardo Querzoni, Antonino Virgillito: "Distributed Event Routing in Publish/Subscribe Communication Systems: a Survey," Dipartimento di Informatica e Sistemistica Universit`a di Roma La Sapienza, 2005.

[13] Peter R. Pietzuch, Jean M. Bacon: "Hermes: A Distributed Event-Based Middleware Architecture," Distributed Computing Systems Workshops, 2002. Proceedings. 22nd International Conference.