Receiver-Oriented 'Pull' Model for RosettaNet Trade Documents Interchange

T. T. Ting, K. T. Khoo
School of Computer Sciences,
Universiti Sains Malaysia
11800 Minden, Penang.

Abstract — A major challenge in integrating trading partners' processes is effective document interchange. Traditional business-to-business (B2B) process integration is based on a 'Push' model through which documents are pushed from the senders to the receivers, as in emailing and electronic data interchange. However, as the volume of document increases, the 'Push' model suffers from diminishing data quality and overall service quality. Pushing large volume of pre-defined standard documents among the traders usually results in data redundancy and/or inaccuracy. The larger companies interchanging high volumes of daily transactions among their trading partners are most affected by these setbacks. In this paper, however, the 'Pull' model is explored. This model facilitates users to pull only the latest necessary and sufficient information to avoid data redundancy and inaccuracy, as in the case of RSS 2.0 (Really Simple Syndication). In a B2B document interchange environment, it is believed that smaller blocks of structured data pulled by the receivers are more easily transmittable compared to the large standard documents 'pushed' to the receivers. In the receiver-oriented 'Pull' model, only the required information are identified, transmitted and received in a XML-based format. The streamlined XML-based information can thus be more easily captured and scheduled for automated updates at the backend corporate databases. After exploring the weaknesses in the current B2B document interchange, this paper outlines a possible data-block structure for B2B information retrieval through the RosettaNet Partner Interface Processes.

INTRODUCTION

The 'Push' or 'Pull' (P/P) models have been used in various domains including business supply chain, e-commerce [8,17], database caching in Web applications [10], data broadcasting/dissemination [7,9], network management [18], and computer system performance [11]. Although the 'P/P' models are commonly used in different domains, the 'Push' model is basically used for the 'publish/subscribe/distribute' functions or mass production in the 'selling' paradigm, while the 'Pull' model is used for the 'request/response' functions or in the one-to-one paradigm [18]. In the domains of business supply chain, e-commerce and data broadcasting/dissemination, 'P/P' model is a business strategy used to serve the customer needs either on a group basis ('Push' model) or on a personalized/one-to-one basis ('Pull' model). However, in network management and computer system performance, 'P/P' model is a technical conceptual idea used to improve the existing technology. In this paper, the 'P/P' models are adopted to further improve the data flow process in B2B integration.

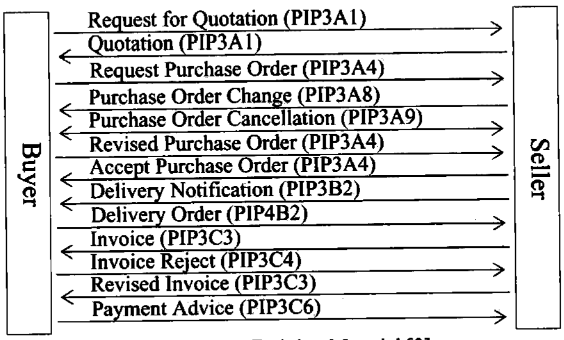
B2B messages are transformed data in certain format packed in data envelopes and transmitted through the transport protocol. Therefore, the flow of messages determines the state of business processes in B2B integration. These messages are business documents such as purchase orders and invoices. In the traditional B2B 'Push' model, all business documents are pushed to the trading partners based on the terms and conditions specified in the Trading Partner Agreements. For example, in a procurement process, business documents such as quotations, purchase orders and invoices are pushed to each other. The problems associated with the traditional 'Push' model is discussed in the next section.

RosettaNet is a non-profit organization established in 1998 defining standards in global supply chain. RosettaNet Partner Interface Processes (PIPs) standardize business processes by defining business document format and content (in XML Schema) in seven sectors: demand creation, design, forecast, order, payment, logistics, and manufacture. Each document is complemented with a business document specification (human-readable and machine-readable), business process specification, PIP process specification and message structure to ease the implementation of the standards. Through standardized business process and documents' content and format, the flow of data can be observed, analyzed and controlled to improve the efficiency of the integration. RosettaNet standards are widely endorsed by some 500 companies worldwide. More details about RosettaNet PIPs standards can be found in [12].

A case study has been carried out in [4] to realize the RosettaNet PIPs compositions as Web service orchestrations. It proposes a framework to execute PIPs for Web services. However, it is still a 'Push' model. Many researchers have attempted to enhance the 'Push' model, but few have shown interest in the 'Pull' model. A broad literature review on the approaches to enhancing B2B integration in RosettaNet environment has been summarized in [2]. This paper focuses on the control of high-level data flows in the B2B processes via RosettaNet PIPs.

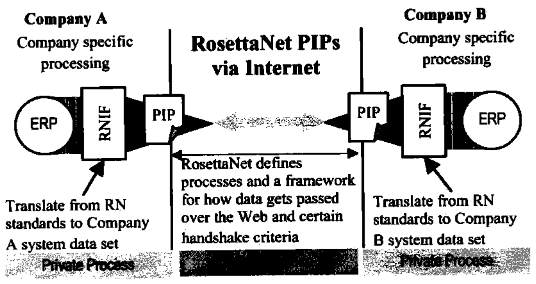
B2B Procurement Process Using PIPs

This section examines the 'Push' model by observing the business documents exchanged in a typical B2B procurement process. Based on the latest update from the RosettaNet Web sites when this paper was written, there was a total of 117 PIPs to-date [12]. Fig. 1 and Fig. 2 show how PIPs and RosettaNet Implementation Framework (RNIF) function in a B2B integration process. It is noted that business documents are exchanged by being pushed through the Internet using the 'Push' model.



Adapted from RosettaNet Training Material [3]

Fig. 1. B2B transactions being pushed through PIPs.



Adapted from RosettaNet Training Material [3]

Fig. 2. PIPs and RNIF in B2B integration.

Emailing is a good example of the use of a 'Push' model. A sender can send an unlimited number of emails to a receiver without the need for any response. As a result, junk messages tend to overload mailboxes. The size of a PIP is usually between 10 megabytes and 100 megabytes. A complete PIP business transaction normally consumes 100's of gigabytes. As a consequence, the need for larger disk storage and higher network speed will increase exponentially as a business grows.

Data Quality and Quality of Service Problems in PIPs

In the following sections, the multi-national companies (MNCs) represent the senders who use PIPs and RNIF to deal with their trading partners. The MNCs exchange thousands of PIPs with their trading partners daily. The sequence of processes is fixed in the Trading Partner Agreement.

Data Quality

Relevance and completeness are two of the main attributes for data quality. Enterprises invest heavily to provide customized documents in specific format and content to meet the needs of their partner's request on data relevance and completeness. Although PIPs take into account the entire content of trade documents which comply with completeness, [6] asserts that PIPs have a problem in "large messages with redundant content". In this study, an analysis is carried out to compare the contents of two PIPs, they are Request Purchase Order and Query Order Status. Among the first 100 data fields in Query Order Status, 90 data fields are redundant in the Request Purchase Order. In other words, 90% of the information exchanged in these two processes are redundant. The entire Request Purchase Order and Query Order Status has been calculated to have 56.21% of redundant fields. Therefore, a large amount of redundant data is actually being interchanged between the partners. It is pointed out in [13] that large amount of emailing reduces network efficiency and speed and thereby affecting a company's ability to meet tight deadlines. Likewise, redundant data could also clog the network and delay the business processes. Nevertheless, customizing or personalizing documents to meet different partners' request is very costly and time consuming.

PIPs standardize the documents' format and content. As the business grows, the content of a predefined document needs to be updated. In PIPs, it is costly and time consuming to alter the content. Therefore, partners are reluctant to change the predefined content. In short, although PIPs standardize the document content, they sometimes do not comply with the 'relevance' criterion.

In the business processes, many business documents need to be produced manually or system generated. For example, in PIP3A, Quote and Order Entry, at least 10 business documents with thousands of data fields need to be created and sent to the trading partners. PIP Request Purchase Order and PIP Query Order Status are two of the business documents that must be created and sent to the business partners. However, duplication of the content in both documents amounts to 56.21%, as stated earlier.

Quality of Service

The 'Push' model is considered not resilient to failure according to the research conducted in [9]. The main reason is that in the 'Push' model, the states of a process are kept at the sender's server. Whenever a sender's server is recovered from a failure, the server needs time to retrieve the states of processes before it can continue to push data to the receiver.

In addition, real time accessibility of data depends on the sender. If the sender does not push the data, the receiver will not get the data. Accessibility refers to the chance whereby a service is capable to meet a request. The 'Push' model has lower accessibility in terms of data retrieval [9].

In view of the numerous problems encountered in the 'Push' model, this paper attempts to propos a receiver-oriented 'Pull' model.

RECEIVER-ORIENTED 'PULL' MODEL

'Pull' Model

The 'Pull' model is more suitable for one-to-one and personalized requests and responses. In this study, the model facilitates the receivers to pull the necessary data through the data services provided. Identification number and authentication information are the main data pushed between the trading partners. Instead of pushing a business document (PIP) to the receiver, data is stored in a database and only the ID is sent (with authentication information) to the receiver. The receiver accesses the data-blocks indirectly from the sender's server and captures data by using the ID provided. The same technique is used for the rest of the business processes. Therefore, ID is pushed between trading partners instead of PIPs. The approach benefits both the senders and receivers where senders do not have to push large volume of documents and receivers request for only the required data.

The main components in a 'Pull' model are ID and authentication information, business state control engine, data capturing and scheduling agent, data service and data-blocks (PIPs in chunks). The data-blocks are structured based on PIPs. They are always updated and published for sharing among the partners. Only partners who have ID and authentication information are authorized to retrieve the data-blocks via the data services. The data capturing and scheduling agents will use the authentication information to capture data in data-block and update the required database. The process flow is controlled by the business state control engine via the confirmation message sent by the data capturing and scheduling agent. Data servicing is an important role to ensure that data-block is extracted from database and compliant to PIPs and ready to be pulled as a service. For example, partner A might request: "I want to get partner's shipping address in invoice." The data service agent will prepare the data-block "Ship To" ready to be pulled by the requestor, who does not need to know what has happened behind the process of getting the information. They get what they want.

Really Simple Syndication (RSS) is an example of the 'Pull' model. RSS is in XML format to syndicate volatile Web site contents. A user can subscribe to RSS Web feed and get updated content through the feed. If a user is interested with the content, s/he can retrieve (pull) the full content through the link provided in the feed. Some examples of RSS web feeds include CNN News Update, XML.com, etc. The main ideas in RSS are adopted in the 'Pull' model.

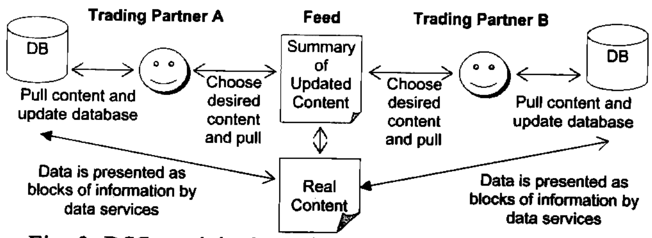


Fig. 3. RSS model adopted in receiver-oriented 'Pull' model.

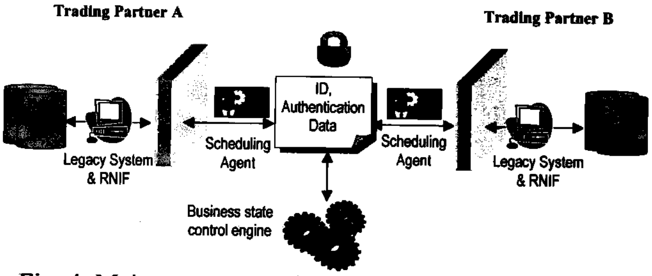


Fig. 4. Main components in the 'Pull' model with PIPs data-blocks and RNIF.

Receiver-Oriented 'Pull' Model in Procurement Transaction

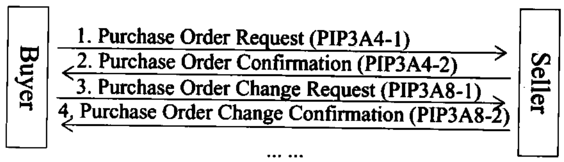


Fig. 5. Four business processes in procurement transaction.

In this section, four processes are examined, namely Request Purchase Order, Purchase Order Confirmation, Purchase Order Change, and Purchase Order Change Confirmation. These four processes are chosen because both the PIP Request Purchase Order (3A4-1 3A4-2) and PIP Request Purchase Order Change (3A8-1, 3A8-2) are modular PIPs which will ease the comparison between the PIPs. Purchase Order Request document will be sent to the Seller. Once the Seller confirms the Purchase Order, the Seller will send the Purchase Order Confirmation to the Buyer. Should there be any changes to the Purchase Order, the Buyer will send the changes to the Seller who will then send the changed confirmation in return. The Buyer is a receiver for processes 2 and 4, while the Seller is a receiver for processes 1 and 3. Processes 1 and 3 will be discussed in the receiver-oriented 'Pull' model. In the following discussion, the Receiver refers to the Seller and Sender refers to the Buyer. In processes 1 and 3, two documents are received: Purchase Order Request and Purchase Order Change Request. The following table shows some possible data-blocks in both Purchase Order Request and Purchase Order Change Request documents. The possible data-blocks are compared between two documents and shaded fields in TABLE 1 are similar data-blocks with the same content.

TABLE 1 POSSIBLE DATA-BLOCKS IN PURCHASE ORDER REQUEST AND PURCHASE ORDER CHANGE REQUEST

Purchase Order Request	Purchase Order Change Request
AccountDescription	AccountDescription
BillTo	GeneralServicesAdministrationNumber
BusinessDocumentReference	BusinessDocumentReference
ContractInformation	ContractInformation
contractOwner	OriginalScheduledEvent
FinancingTerms	FinancingTerms
GovernmentContractIdentifier	GovernmentContractIdentifier
InstallAt	InstallAt
ListPrice	PurchaseOrderIdentifier
OrderShippingInformation	OrderShippingInformation
RequestedEvent	RequestedEvent
RespondTo	ReceivedQuantityInformation
SecondaryBuyer	SecondaryBuyer
ShipTo	ShipTo
SpecialDiscount	RequestedShipFrom
TaxExemptStatus	TaxExemptStatus
TotalAmount	TotalAmount
-	ScheduledEvent

After the Sender has successfully sent the Purchase Order Request in process 1 (Purchase Order Request), the Receiver would have all the data-blocks in the Purchase Order Request. In process 3 (Purchase Order Change Request), the Receiver's server may choose to pull only certain data-blocks and eliminate similar data-blocks, provided that there are no changes to the previous data-blocks sent. In other words, the Receiver only pulls the updated or new data-blocks from the Sender.

The data capturing and scheduling agents play the role of deciding when and what to pull. Once the Receiver's agent receives an updated summary from the Sender, it will choose the appropriate data-blocks to be pulled from the Sender's server.

Resolving problems in the 'Push' model

In the 'Pull' model, the partners can choose to pull the necessary information into their databases. Therefore, the Sender does not have to customize documents satisfying different partners' requests. The Sender only needs to supply data-blocks ready as service to serve the partners. Since the partners pull the data by themselves, the data will eventually be relevant and complete. As such, data redundancy is reduced.

The scheduling agent provides the schedule to capture the data-blocks. Once the data-blocks are captured, the database files could be updated. Data-blocks could flow more easily through the network due to its reduced size. Therefore, quality of service in reliability is increased. 'Pull' model meets the real time requirement in term of data flow and database update.

In addition, the 'Pull' model can also handle ad-hoc requests. Since the data-block is provided via data services, even after one business transaction is fully committed, summary of caption of data-blocks is still available in archive and can be retrieved anytime. Once the requestor has an ID, authentication information and caption, they can pull the data-block at the most convenient time.

In the 'Pull' model, data flow depends on the receivers rather than the senders. At the most appropriate time, the receiver's agent can be scheduled to retrieve the data-blocks. This is crucial especially for international enterprises where partners have different time zones; services have to be highly accessible at anytime for effective business processes.

Lastly, it is argued that the 'Pull' model is resilient to failures [9]. In the 'Push' model, the state information about the receiver's need is maintained by the Sender. Once the sender's machine fails, the receiver has to wait for the sender to recover from the state information.

One challenge in realizing the 'Pull' model is how to compose data-blocks from thousands of lines in a PIP? In other words, how to transform a PIP document into smaller pieces of data to be pulled across the network? The following section proposes a possible data-block structure.

Possible Data-Block Structure

PIPs use an object-oriented architecture to construct business documents. Elements in a business document could belong to a Business Related Reusable Structure Domain, a Non-Business Related Reusable Structure Domain or a domain proprietary to a particular PIP. Both the Business Related Reusable Structure Domain and the Non-Business Related Reusable Structure Domain contain reusable classes. In the PIPs specification, a data structure is the description of elements in the business documents which collects classes, class attributes with associated data type, constraints, patterns, restrictions and relationships with other classes [12].

Each PIP has certain portions of redundant data compared to other PIP as discussed earlier. It is observed that most of these redundant data fields are reusable data structure from the Business Related Reusable Structure Domain or Non-Business Related Reusable Structure Domain. For example, both the PIP Request Purchase Order and PIP Query Order Status contain data structure ContractInformation from Procurement (Business Related Reusable Structure Domain). It is liken to physical business documents whereby both *purchase order* and *delivery order* have trading partners' contact information, products information, credit card information, etc.

Possible data-blocks can be formed from the original data structure. For example, in Procurement Business Related Reusable Structure Domain, some data-blocks can be formed, e.g. AccountDescription, ContractInformation, Financing Terms, etc. To further illustrate the possible data-blocks, PurchaseOrder section in PIP Request Purchase Order is studied. TABLE 1 shows the possible data-blocks formed.

The following diagram shows how business processes can be enhanced by using data-blocks.

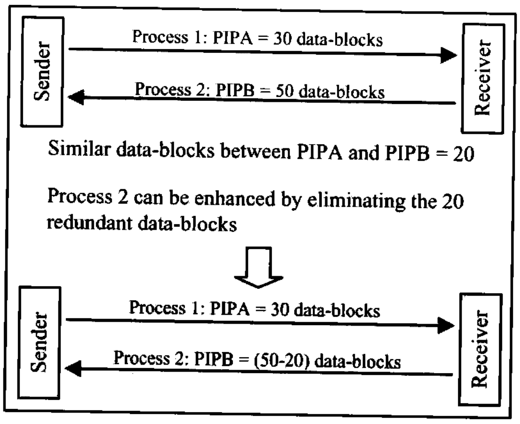


Fig. 6. Enhanced business process using data-blocks.

CONCLUSION

This paper has examined the existing B2B 'Push' model and its problems associated with the RosettaNet PIPs. In an ordinary business document interchange process, documents are formatted and pushed to the trading partners. After analyzing the problems of diminishing data quality and quality of service as the volume of transaction increases, a receiver-oriented 'Pull' model was proposed. In the proposed 'Pull' model, data-blocks are proposed as the basic units of data interchange. With the identification and authentication information, business partners can retrieve and update the necessary data into their systems. Further study is required to examine in detail various possible data-block structures for the 'Pull' model to function more effectively.

REFERENCES

[1] W. Shen, Q. Hao, S. Wang, Y. Li, and H. Ghenniwa, "An Agent-Based Service-Oriented Integration Architecture for Collaborative Intelligent Manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 23, pp. 315-325, June 2007.

[2] T. T. Ting, K. T. Khoo and Y. -N. Cheah, "Approaches to Enhancing RosettaNet Standards Component," presented at the Computer Science Postgraduate Colloquium, Gelugor, Penang, June 25-26, 2007.

[3] RosettaNet Engineering, "Transitioning to Gen. 2 PIP Specifications", presented at the RosettaNet Training, Georgetown, Penang, April 9-24, 2007.

[4] A. Schönberger, "Modelling and Validating Business Collaborations: A Case Study on RosettaNet," *Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik*, vol. 65, March 2006.

[5] D. Bollier, "When Push Comes To Pull - The New Economy and Culture of Networking Technology," report of the 14th Annual Aspen Institute Roundtable on Information Technology, Washington, DC, 2006.

[6] S. Damodaran, "B2B Integration over the Internet with XML: RosettaNet Successes and Challenges," in *Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers & Posters*, 2004, pp. 188-195.

[7] D. Aksoy and M. S. -F. Leung, "Pull vs Push: A Quantitative Comparison for Data Broadcast," in *Proceedings of the Global Telecommunications Conference*, 2004, pp. 1464-1468.

[8] R. Patricio, "Internet Based Technologies (IBTS) Integration in B2B Models - The Impact on Competitive Advantage of Firms," *Frontiers of E-Business Research*, pp. 369-390, 2002.

[9] M. Bhide, P. Deolasee, A. Katkar, A. Panchbudhe, K. Ramamritham and P. Shenoy, "Adaptive Push-Pull: Disseminating Dynamic Web Data," *IEEE Transactions on Computers*, vol. 51, pp. 652-668, June 2002.

[10] C. Mohan, "Caching Technologies for Web Applications," in *Proceedings of the 27th International Conference on Very Large Data Bases*, 2001, pp. 726.

[11] C. -L. Yang and A. R. Lebeck, "Push vs Pull: Data Movement for Linked Data Structures," in *Proceedings of the 14th International Conference on Supercomputing*, 2000, pp. 176-186.

[12] RosettaNet, "RosettaNet Standards," http://www.rosettanet.org, 2007.

[13] Surfcontrol, "Excessive Spam E-Mails were Beginning to Slow Down the Network, Jeopardizing Fast Turnarounds on Customer Projects". http://www.surfcontrol.com/general/assets/case_studies/imagemaster.pdf, 2007.

[14] B2B Commerce (M) Sdn. Bhd., "B2B Commerce", http://www.b2b.com.my/estorefront/pages/main/index.jsp, 2007.

[15] Wikipedia, "RSS", http://en.wikipedia.org/wiki/rss_(file_format), 2007

[16] Saskatchewan Efuturecentre, "Business-to-Business E-Commerce Basics". http://www.e-future.ca/sask/pdf/b2b_basics.pdf, 2007.

[17] Alberta Efuturecentre, "Supply Chain Basics", http://www.e-future.ca/alberta/pdf/efc_supply_chain_basics.pdf, 2007.

[18] J. P. Martin-Flatin, "Push vs Pull in Web-Based Network Management", *in Proceedings of the 6th IFIP/IEEE International Symposium on Integrated Network Management*, 1999, pp. 3-18.