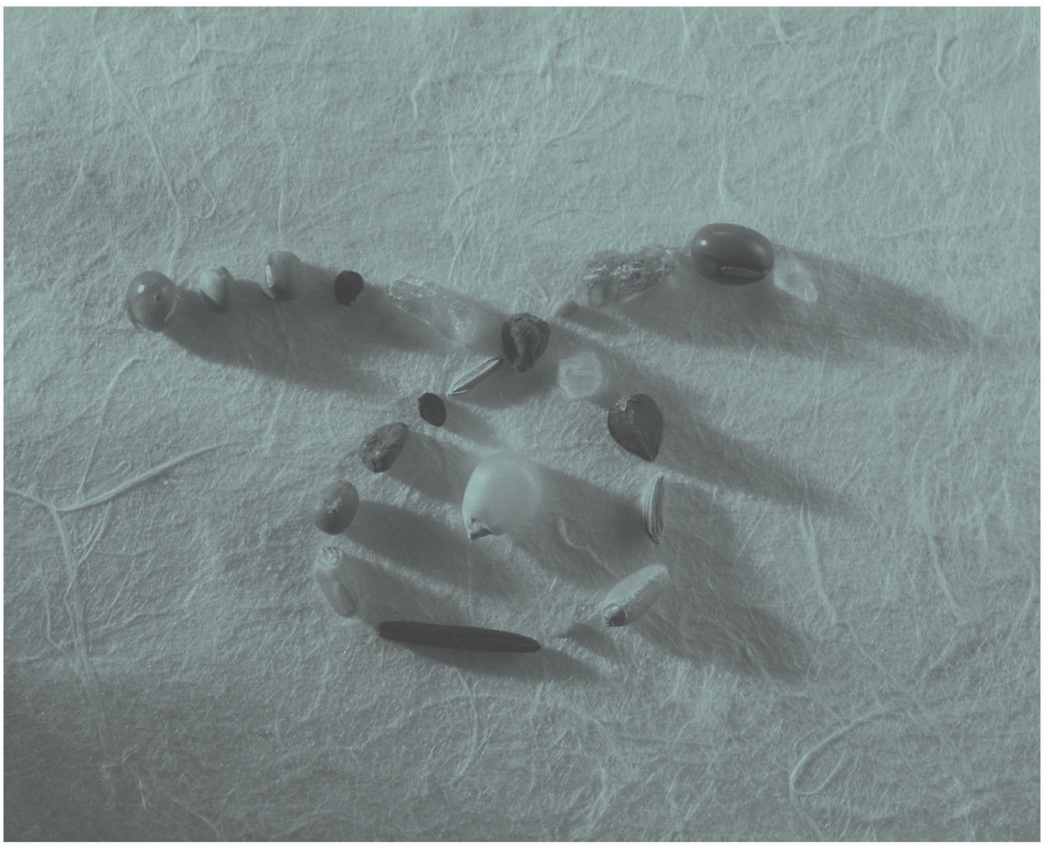
Vol. 1, No. 1
January - March 2005

**An official publication of
the Information Resources
Management Association**

INTERNATIONAL JOURNAL OF INTELLIGENT INFORMATION TECHNOLOGIES



IDEA GROUP PUBLISHING
Publisher of IT books, journals and cases since 1988
www.idea-group.com

The *International Journal of Intelligent Information Technologies* (ISSN 1548-3657; eISSN 1548-3665) is published quarterly by Idea Group Publishing, 701 E. Chocolate Avenue, Suite 200, Hershey, PA 17033-1240, USA. Subscription fee to this journal is US$85 individual and US$195 institutional. Sample copies may be requested (based on availability). **Advertising information:** Idea Group Publishing, 701 E. Chocolate Avenue, Suite 200, Hershey, PA 17033-1240, USA, 717/533-8845, <www.idea-group.com>. **Postmaster:** Send all address changes to above address.

Idea Group Publishing
Publisher in information science, technology and management

701 E. Chocolate Avenue, Suite 200, Hershey, PA 17033-1240
Tel: (717) 533-8845 • Fax: (717) 533-8661 • E-mail: cust@idea-group.com
www.idea-group.com

Special Issue Proposal

IJIIT publishes special issues on the recent trends in Intelligent Information Technologies and related issues. A short proposal containing the title of proposed special issue, tentative list of invited authors, suggested date of publication and a call-for-papers should be sent to *sugumara@oakland.edu* for review.

Idea Group Publishing
http://www.idea-group.com

Hershey • London • Melbourne • Singapore
Tel: 717/533-8845 • Fax: 717/533-8661 • E-mail: cust@idea-group.com

International Journal of
Intelligent Information Technologies

Jan-March 2005, Vol 1, No. 1

Table of Contents

IJIIT is listed in the following indexes: GetCited; MediaFinder; Standard Periodical Directory; Ulrich's International Periodicals Directory

International Journal of Intelligent Information Technologies

Guidelines for Manuscript Submissions

Mission: The advent of the World Wide Web has sparked renewed interest in the area of intelligent information technologies. There is a growing interest in developing intelligent technologies that enable users to accomplish complex tasks in web-centric environments with relative ease, utilizing such technologies as intelligent agents, distributed computing in heterogeneous environments, and computer supported collaborative work. The mission of the **IJIIT** is to bring together researchers in related fields such as information systems, distributed AI, intelligent agents, and collaborative work, to explore and discuss various aspects of design and development of intelligent technologies. For more information, visit *www.idea-group.com/ijiit*.

Coverage: The **IJIIT** encourages quality research dealing with (but not limited to) the following topics: intelligent information systems development using *Design Science* principles; models, architectures and behavior models for agent-oriented information systems; intelligent agent and multi-agent systems in various domains; intelligent information retrieval and business intelligence; Semantic Web; Web services and ontologies; reasoning, learning and adaptive systems; intelligent decision support systems; agent-based auction, contracting, negotiation, and ecommerce; privacy, security, and trust issues; and other topics related to Intelligent Information Technologies.

Originality: Prospective authors should note that only original and previously unpublished manuscripts will be considered. Furthermore, simultaneous submissions are not acceptable. Submission of a manuscript is interpreted as a statement of certification that no part of the manuscript is copyrighted by any other publication nor is under review by any other formal publication. It is the primary responsibility of the author to obtain proper permission for the use of any copyrighted materials in the manuscript, prior to the submission of the manuscript.

Style: Submitted manuscripts must be written in the APA (American Psychological Association) editorial style. References should relate only to material cited within the manuscript and be listed in alphabetical order, including the author's name, complete title of the cited work, title of the source, volume, issue, year of publication, and pages cited. Please do not include any abbreviations. See the following examples:

Example 1: Single author periodical publication.

Smith, A.J. (2002). Information and organizations. *Management Ideology Review*, 16(2), 1-15.

Example 2: Multiple authors periodical publication.

Smith, A.J., & Brown, C.J. (1988). Organizations and information processing. *Management Source*, 10(4), 77-88.

Example 3: Books.

Smith, A.J. (2002). *Information booklet*. New York: J.J. Press.

State author's name and year of publication where you use the source in the text. See the following:

Example 1: In most organizations, information resources are considered a major resource (Brown, 1988; Smith, 2002).

Example 2: Brown (2002) states that the value of information is recognized by most organizations.

Direct quotations of another author's work should be followed by the author's name, date of publication, and the page(s) on which the quotation appears in the original text.

Example 1: Brown (1989) states that "the value of information is realized by most organizations" (p. 45).

Example 2: In most organizations, "information resources are considered to be a major organization asset" (Smith, 2002, pp. 35-36) and must be carefully monitored by the senior management.

For more information please consult the APA manual.

Review process: To ensure the high quality of published material, **IJIIT** utilizes a group of experts to review submitted manuscripts. Upon receipt of the manuscript, two reviewers are selected from the Editorial Review Board of the journal. The selection is based upon the particular area of expertise of the reviewers, matched to the subject matter of the submission. An additional ad-hoc reviewer is also selected to review the manuscript. Therefore, each submission is accordingly blind reviewed by at least three reviewers. Revised manuscripts will be reviewed again by the original review panel with the addition of one new reviewer. Return of a manuscript to the author(s) for revision does not guarantee acceptance of the manuscript for publication. The final decision will be based upon the comments of the reviewers, upon their second review of the revised manuscript.

Copyright: Authors are asked to sign a warranty and copyright agreement upon acceptance of their manuscript, before the manuscript can be published. All copyrights, including translation of the published material into other languages are reserved by the publisher, Idea Group Inc. Upon transfer of the copyright to the publisher, no part of the manuscript may be reproduced in any form without written permission of the publisher.

Submission: Authors are asked to submit their manuscripts for possible publication by e-mail as a file attachment in Microsoft Word (Rich Text Format), or PDF to *sugumara@oakland.edu*. The main body of the e-mail message should contain the title of the paper and the names and addresses of all authors. Manuscripts must be in English. The author's name should not be included anywhere in the manuscript, except on the cover page. Two versions are needed for PDF submissions: one version with author names/affiliations and contact information and another version without this information (for review purposes). Manuscripts must also be accompanied by an abstract of 100-150 words, precisely summarizing the mission and object of the manuscript.

Length: The length of the submitted manuscript must not exceed 25 double-spaced, typed pages using at least 11-pt fonts with 1 inch margins, including all illustrations, tables and references. Discussion and analysis should be complete but not unnecessarily long or repetitive. The papers should not exceed 6,000 words.

Correspondence: The acknowledgment letter regarding the receipt of the manuscript will be promptly sent. The review process will take approximately 8-16 weeks, and the author will be notified concerning the possibility of publication of the manuscript as soon as the review process is completed. All correspondence will be directed to the first author of multi-authored manuscripts. It is the responsibility of the first author to communicate with the other author(s). Authors of accepted manuscript will be asked to provide a final copy of their manuscript in either Word or RTF text format stored on a 3 1/2" disk, zip disk or CD-ROM, accompanied by a hardcopy of the manuscript and the original signed copy of the Warranty and Copyright Agreement. The accepted manuscript will be edited by the Journal copyeditor for format and style.

Book Reviews: IJIIT invites prospective book reviewers to submit their review of either textbooks or professional books for possible inclusion in the journal. Reviewers should focus on the following guidelines when developing the book review:

- Book reviews must not exceed 1,500 words; • Reviews should summarize the book and indicate the highlights, strengths, and weaknesses of the book; • Reviews should evaluate the organizational and managerial applications of the material discussed in the book relevant to information resources and technology management; Reviews should critique and constructively evaluate the author's work and not merely list the chapters' contents; The writing style, accuracy, relevance, and the need for such a work in the discipline should be analyzed; The review must include the title of the book, author, publishing company, publication date, number of pages, cost (if listed), and ISBN number; and each submission must be accompanied by a short biography of the reviewer.

Case Studies: IJIIT also encourages submission of case studies based on actual cases related to different issues and aspects of intelligent technologies and systems. Case studies must provide adequate information regarding the organization upon which the case is based, discussion of the issues involved, coverage of any experiments or trials of techniques or managerial approaches, and finally, discussion of any lessons learned or conclusions drawn from this study. Case studies must not exceed 6000 words.

ALL SUBMISSIONS AND QUESTIONS SHOULD BE FORWARDED TO:

Vijayan Sugumaran, Editor-in-Chief; sugumara@oakland.edu

PREFACE

The Inaugural Issue of the *International Journal of Intelligent Information Technologies*

Vijayan Sugumaran, Editor-in-Chief

INTRODUCTION

Fierce competition, globalization, and the digital economy have forced organizations to search for new ways to improve competitive advantage. Consequently, there is great need for improving communication and information flow, and providing decision-making capabilities for organizations so they can respond quickly to market changes. To increase organizational productivity, emerging information technologies (such as intelligent agents) are being applied to create a cooperative and group-based work environment.

The advent of the Web has sparked renewed research interest in the area of intelligent information technologies. There is growing interest in developing intelligent information systems that enable users to accomplish complex tasks in Web-centric environments with relative ease. Although artificial intelligence (AI) technologies such as expert systems and neural networks have been successfully used in aerospace, communication, medical, and financial applications, they have not made a significant impact on improving overall productivity due to their narrow scope. In contrast, the new breed of "intelligent information technologies" hold greater potential in that they can be applied to a large number of domains and a diverse set of problems. A generic intelligent-agent based application, for instance, can be customized for different domains and a variety of problem scenarios.

Intelligent systems are generally characterized as systems that help users in carrying out difficult tasks by minimizing complexity and, hence, the user's cognitive burden. These systems have a learning component and gain "experience" over time. They are designed to respond to changes in the environment and new situations with minimal or no human intervention. They are context sensitive and capable of making sense out of ambiguous or contradictory information. They also maintain user profiles including user preferences and previous actions, and serve as a tutor, critic, consultant or advisor by providing suggestions and/or courses of action to take. These systems exhibit "intelligent" behavior by dealing with complex situations and applying their knowledge to manipulate the environment by recognizing the relative importance of different

elements within a problem scenario. Several enabling technologies and techniques such as intelligence-based modeling, grid computing, data mining and knowledge discovery; data warehousing and business intelligence; fuzzy computing, neural networks, machine learning and evolutionary algorithms, business components; and Web technologies are being utilized in creating intelligent systems.

Intelligent agent technology is emerging as one of the most important and rapidly advancing areas. Researchers are developing a number of agent-based applications and multi-agent systems in a variety of fields such as electronic commerce, supply chain management, resource allocation, intelligent manufacturing, mass customization, industrial control, information retrieval and filtering; collaborative work, mobile commerce, decision support, and computer games. While research on various aspects of intelligent agent technology and its applications is progressing at a very fast pace, many issues remain to be investigated and explored in areas such as agent design, implementation, integration, deployment, evaluation and business value.

For example, salient characteristics of agents that are specific to different domains, formal approaches for agent-oriented modeling, reusable agent patterns, platforms and tools for designing and implementing agent-oriented information systems, collaboration and coordination in multi-agent systems, integration, testing and configuration management of agent-based systems; scalability and robustness of agents; security issues related to mobile agents and multi-agent systems; and understanding the organizational impact of agent-based systems are some of the areas requiring further research. Some of the future directions for agent research include integrating component-based development and agent technology; transferring research themes and prototypes from the laboratories into industrial strength business applications; use of agents in grid and pervasive computing, performance and accountability of agents in mission critical applications, and effective utilization of agents for knowledge and content management in the digital economy.

SCOPE & STRUCTURE OF THE JOURNAL

Research on intelligent information systems covers a broad range of disciplines and technologies. The *International Journal of Intelligent Information Technologies* (*IJIIT*) provides a forum for academics and practitioners to explore research issues related to not only the design, implementation and deployment of intelligent systems and technologies, but also economic issues and organizational impacts. Papers related to all aspects of Intelligent systems including theoretical work on agent and multi-agent systems as well as case studies offering insights into agent-based problem solving with empirical or simulation based evidence are welcome. The journal is interested in publishing both basic and applied research.

Some of the focal points of *IJIIT* include: intelligent agent and multi-agent systems; ontologies and Web services; business intelligence through data warehousing and data mining; adaptive and reasoning systems; intelligent decision support, and individual, organizational and social impacts of intelligent information systems. While the thrust of the journal is on agent-based systems and their application in various domains,

the journal is also receptive to manuscripts dealing with business applications of other intelligent technologies such as case based reasoning; fuzzy computing, neural networks, evolutionary algorithms, machine learning, and so on. Furthermore, *IJIIT* is not only interested in publishing technical articles but also articles containing empirical investigations addressing the business value and use of intelligent information systems.

Due to the diversity and the rapidly changing nature of intelligent information technologies, *IJIIT* will publish different types of articles to facilitate timely dissemination of research results and applications. *IJIIT* intends to publish the following types of articles:

- *Research papers*: These are traditional research papers describing novel research, design and implementation of intelligent information systems, application of intelligent technologies in a variety of domains, large-scale experiments, and exciting systems of relevance to the journal's academic audience as well as practitioners. Each research paper must be an original submission prepared according to generally accepted standards of rigorous scientific analysis and should not exceed 6,000 words.
- *Research notes*: These are short papers that focus on state of the art thinking about a specific problem or a key innovation or new developments whose timely publication is important. Research notes are just as original and rigorous as research papers but concise and do not exceed 3,000 words.
- *Prototype demonstrations*: Prototype demonstration papers are short papers that showcase new research ideas and emerging technological solutions. These papers explain what the prototype does, why it is useful and provide a link to the working prototype so that any reader can use it (either onsite or after download and local installation). These papers do not exceed 2,000 words and are reviewed and accepted for publication only if the prototype works and is of high quality.
- *Case studies*: These studies are based on actual cases related to different issues and aspects of intelligent technologies and systems. Case studies (about 6,000 words) must provide adequate information regarding the organization upon which the case is based, discussion of the issues involved, coverage of any experiments or trials of techniques or managerial approaches, and finally, a rich discussion of any lessons learned and or conclusions drawn from this study for research and practice.
- *Book reviews*: These are reviews of either textbooks or professional books. They provide a summary and carefully examine the strengths, and weaknesses of the book. Most importantly, reviews discuss the contribution in the context of the extant literature. They evaluate the organizational and managerial applications of the material discussed in the book relevant to information resources and technology management. Book reviews do not exceed 1,500 words.

To ensure the high quality of published material, *IJIIT* utilizes a group of experts to review submitted manuscripts. Upon receipt of the manuscript, two reviewers from the Editorial Review Board and an additional ad-hoc reviewer are selected to review the manuscript. Therefore, each submission is blind reviewed by at least three reviewers.

The Associate Editors assess and make recommendations pertaining to the manuscript's disposition: accept, revise or reject. They make these decisions based on their own review as well as the feedback from the evaluation forms previously completed by the three reviewers assigned to the paper. Associate Editors occasionally provide an editorial preface that discusses some issues, trends, problems or solutions pertaining to the theme of the journal.

ABOUT THIS ISSUE

This inaugural issue of the *International Journal of Intelligent Information Technologies* contains four research papers. Two of these articles discuss agent-based applications, another article discusses evolutionary agents and the last article relates to maintaining privacy during data mining (association rule mining). In their article, "An Agent Based Approach for Sourcing Business Rules in Supply Chain Management," Ram and Liu discuss the importance of business rule management in supply chains to support real-time decision making and provide an agent-enabled mechanism that dynamically maps business rules to the enterprise data model. They describe a semantics-based data sourcing service that supports the execution of business rules. This requires data to be retrieved from various data sources spread across the enterprise, including the enterprise data warehouse. Their contribution is a prototype system to illustrate the sourcing service and demonstrate the feasibility of their approach.

In the next article, "Memes and Mutation: Societal Implications of Evolutionary Agents in Push Technologies," Kendall and Kendall describe and define autonomous and evolutionary agents in the context of push technologies. These agents are utilized for seeking, organizing and creating information via the Web and other pervasive and innovative information technologies. They explore how memes will influence evolutionary agents. Memes are dominant messages, strong enough to be copied over and over again from one agent to another. In particular, they discuss the social implications of meritorious and malevolent memes exchanged by evolutionary agents. The Kendall's significant contribution is to raise awareness of memes and help focus research on how messages evolve as well as how agents evolve. This will prove useful as agents proliferate.

Heine, Herrler and Kirn discuss the use of agents in the medical domain in the third article, "ADAPT@AGENT. HOSPITAL: Agent-Based Optimization and Management of Clinical Processes." They present a system that can substantially increase the efficiency of hospital process management through agent-based simulation. They describe the architecture of this system and its functionalities and discuss an example scenario for "clinical trials" to illustrate how the system supports distributed clinical process management. Their contribution is an outstanding example of using agents in the medical field.

The last article discusses a data mining algorithm (association rule mining) that preserves privacy. In their article, "PPDAM: Privacy Preserving Distributed Association Rule Mining Algorithm," Ashrafi, Taniar and Smith argue that data mining using Web data may disclose some sensitive information. As a result privacy is one of the prime concerns in data mining research, and the authors present a methodology that generates asso-

ciation rules without revealing confidential inputs such as statistical properties of individual sites, but still retains a high level of accuracy in the resultant rules. They make a measurable contribution by reducing the communication cost significantly compared to other well-known distributed association rule mining algorithms.

CONCLUSION

Intelligent technologies will attain a permanent place in the modern economy and will be deployed in many roles to increase productivity in business either as assistants to human operators or as autonomous decision-making components of complex systems. One can easily envision a world filled with millions of knowledge agents and tools where the boundary between humans and machine agents is blurred. Intelligent information systems will be an important enabling component in the transformation of business and work place. Human agents can delegate a range of tasks to personalized software agents that can not only make decisions based on the criteria provided by their human counter parts, but also model the reasoning, action, communication, and collaboration skills involved in performing human job functions. Capturing organizational knowledge in a reusable form, and having intelligent systems that access and utilize this corporate knowledge seamlessly is going to further revolutionize the work environment.

We are at a critical juncture in the development of intelligent information systems. Several new and exciting technologies have emerged that could make it possible to design and develop intelligent systems in order to solve a number of business problems that have remained too difficult and complex for other approaches. *IJIIT* is committed to become an integral part of scientific discovery in this area of intelligent technology, supporting a research community that has not always received its due attention. *IJIIT* will report on current and future developments; create synergy among a variety of research groups; and promote interdisciplinary research. As we forge ahead, I look forward to receiving your papers. I encourage involvement and support in making *IJIIT* a premier publication and a valuable resource. I would like to take this opportunity to thank all the members of the Editorial Board, Associate Editors and Idea Group for making this journal possible. I would also like to thank all the authors who have chosen to publish their research in *IJIIT* and the paper reviewers for their valuable efforts in reviewing the submissions. My special thanks to the Association for Information Systems' Special Interest Group on Agent Based Information Systems (AIS SIGABIS) for endorsing and promoting this journal to its members. I hope that you will find the current and future issues of *IJIIT* to be informative, useful and an important resource for the state of the art in intelligent information systems. Welcome to the inaugural issue of the International Journal of Intelligent Information Technologies!

Vijayan Sugumaran
Editor-In-Chief
IJIIT
Dept. of Decision & Information Sciences
School of Business Administration
Oakland University
Rochester, MI 48309
sugumara@oakland.edu

Vijayan Sugumaran is an associate professor of Management Information Systems in the Department of Decision and Information Sciences at Oakland University, Rochester, MI. Dr. Sugumaran received his BTech in mining engineering from Indian School of Mines, MS in mining engineering from the University of Alaska – Fairbanks, and PhD in information technology from George Mason University. His research interests are in the areas of intelligent agent and multi-agent systems, component based software development and reuse, ontologies, Semantic Web and Web services, and Internet technologies.

An Agent-Based Approach for Sourcing Business Rules in Supply Chain Management

Sudha Ram, University of Arizona, USA
Jun Liu, University of Arizona, USA

ABSTRACT

In recent years, business rule management has become an important component of enterprise information systems. Business rules represent guidelines about how an enterprise should conduct its business and provide better service for customers. Business rules are being widely deployed in supply chains to support real-time decision making. The research reported in this paper is aimed at designing a dynamically adaptable data-sourcing service for deploying business rules effectively in supply chain management. Such a data-sourcing service is important since execution of business rules requires data to be retrieved from various data sources spread across the enterprise, including the enterprise data warehouse. We propose a semantics-based approach to implement the data-sourcing service for business rules. Our approach captures semantics of business rules and provides an agent-enabled mechanism that dynamically maps business rules to the enterprise data model. A prototype system is implemented to illustrate our sourcing service and demonstrate the feasibility of our approach.

Keywords: business rules; e-business; enterprise data management; intelligent agents; supply chain management

INTRODUCTION

In recent years, business rule management has become an important component of enterprise information systems. Business rules are defined as "statements that define or constrain some business aspect, intended to assert business or to control or influence business behavior" (Business Rule Group, 2000). In other words, business rules represent policies and guidelines about how a business should conduct its work to better service customers (Leite & Leonardi, 1998; Rosca, Greenspan, & Wild, 2002). Nowadays, the need to incorporate business rules into information systems is becoming imperative due to the rapid development of e-

business. E-business implies that customers are just one mouse click away. To get a competitive edge in e-business, enterprises must be capable of making fast business decisions based on the market condition without the involvement of humans because any delayed response could easily lead to lost business. Therefore, deploying business rules and accommodating frequent changes to the rules have become major issues that affect the competitiveness of an enterprise.

Because of the importance of business rules, organizations have been paying increasing attention to the management of business rules as a strategic corporate asset (Shao & Pound, 1999). This new focus has led to the development of technologies around managing and executing business rules, such as rule engines providing inference capabilities over business rules (Ceri, Gennaro, et al., 2003; Rosca et al., 2002) and repositories for storing and managing rule sets (Herbst, 1996, 1997; Von Halle, 2002). These technologies follow the principle of "externalizing business rules" (Date, 2000) and manage business rules separately via a centralized facility that resides outside of core business systems. The centralized management of business rules provides enterprises with greater competitive advantage as it enables them to be more responsive to changes within the company (internal policy changes) and the marketplace (products and pricing changes in response to volatile customer demand). However, as a part of the information system of an enterprise, a complete business rule service still requires other software components that interface with core business systems to be effective. For example, a software component that sources data for business rules is important because the logic of business rules must operate on data stored in the databases. Nevertheless, integration of business rules services into information systems has not been adequately addressed in literature.

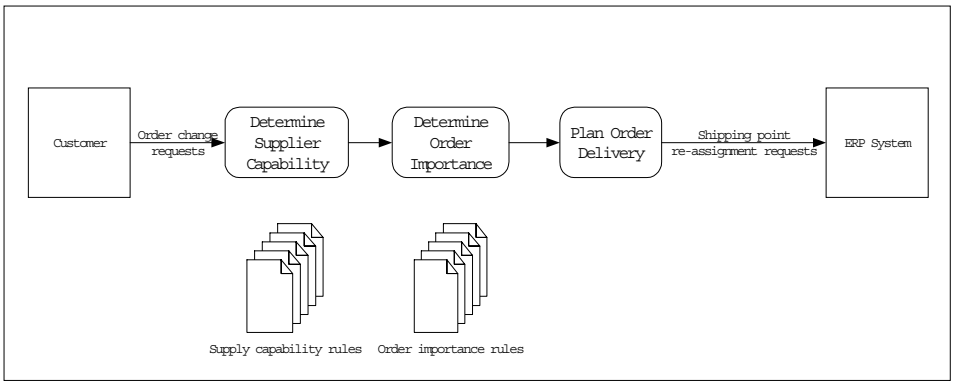
In this work, we propose a semantics-based approach for sourcing business rules for efficient supply chain management. Our approach employs an agent-based design and thereby emphasizes the flexible interaction of the business rule service with business information systems. A software agent can be defined as an encapsulated computer program situated in some environment and capable of flexible, autonomous action in that environment in order to meet its design objectives (Wooldridge, 1997). Work in the area of software agents has been ongoing for some time, and agent-based systems have proved useful in a broad array of applications (e.g., see Boudriga & Obaidat, 2004; He, Jennings, & Leung, 2003; Jennings, Sycara, & Wooldridge, 1998). In this paper, we extend the agent-based design paradigm to the domain of business rules. According to Jennings (2001), agent-based systems can potentially help find solutions to a complex task in a widely distributed, heterogeneous, uncertain information environment. Because business rules are often managed independently and the business rule service interacts with distributed business applications across the enterprise, it is a natural choice to create an agent-based approach for managing and sourcing business rules.

The rest of the paper is organized as follows. We first motivate the need for sourcing business rules using a scenario from a manufacturing organization and identify some challenging issues. Next, we review related research on modeling business rules. We then present our semantics-based approach for sourcing business rules. Next, we describe a prototype system developed using our approach and then summarize our contributions. Finally, we conclude with some directions for future research.

IMPORTANCE OF BUSINESS RULE SOURCING

Consider the following scenario for a large semiconductor manufacturing company in the management of its supply chain network. The biggest challenge facing the company is that late order changes happen very often, and order changes account for approximately 20% of its total order request. To ensure that it has on hand the specific products that the customers are favoring at a particular point, in spite of the order changes, the company deploys a delivery-planning module in its supply chain that accepts order-change requests and then reassigns shipping points for orders that were in jeopardy of not being fulfilled. A decision support agent embedded in the delivery-planning module makes order reallocation decisions (Figure 1). After receiving order-change requests from an upstream supply chain node, the decision support agent determines the importance of the order, evaluates the supplier's capability to fulfill the order, and then creates an order delivery plan and submits a request to the ERP system to reassign the shipping point for the order. As shown in Figure 1, two sets of business rules, that is, "order importance rules" and "supplier capability rules," are deployed to support decision making. The decision support agent utilizes order importance rules to evaluate the importance of a specific order, and supplier capability rules to determine the supplier's capacity to fulfill the order. An example of an order importance rule is, "If order-criticality is

Figure 1: Decision Process of the Delivery-Planning Module



very-low and order-size is small, then order-importance is very-low," and an example of a supplier capability rule can be, "If supplier-inventory availability is available and supplier responsiveness is good, then supplier-delivery capability is good." The data required to execute the rules and make decisions is obtained (sourced) from a number of distributed information sources in the enterprise. Once a set of rules is executed and a decision is made, the data generated from the decision (e.g., a new shipping point) is fed back into one or more information sources, such as the ERP system.

Our main objective in this project is to design a data-sourcing service for deploying business rules. The sourcing service is crucial because execution of the business rules requires data to be retrieved from various data sources spread across the enterprise, including the enterprise data warehouse and other databases. We propose that such a data-sourcing service should map business rules to an enterprise data model. In designing our approach to accomplish this, we address the following issues.

1. Business rules often represent the strategic or tactical decisions of a business (Rosca et al., 2002). Business experts usually specify business rules in natural language (Sandifer & Von Halle, 1993; Von Halle, 2002) so that they can provide complete and appropriate business requirements without being restricted by the models of the existing databases. This results in a "semantic gap" between business rules and the data models. Deploying the business rules requires data to be retrieved from the enterprise databases. This requires the terms in the business rules to be mapped to the enterprise data. Our approach attempts to perform this mapping in an easy-to-use and flexible manner using a conceptual model of the rules and the enterprise data.
2. Business rules model factors affecting the competitiveness of a business, and they are thus subject to frequent changes (Shao & Pound, 1999). Our approach for sourcing business rules needs to be flexible enough to accommodate this dynamic feature of business rules. Thus, it must allow users to easily update the mapping between business rules and the enterprise data model whenever a change occurs in the business rules or the underlying database conceptual structure.

Having elucidated the challenging issues we are confronted with, we discuss how our approach addresses these issues in the following sections.

RELATED WORK

In recent years, business rules have received a lot of attention in literature as offering solutions to many information technology problems (Rosca & D'Attilio, 2001). Advocates of the business rule approach recommend that an explicit manipulation of business rules be used for deploying rules in enterprise information systems. Consequently, research has been carried out to explicitly model business rules at the conceptual design stage, and

the modeled rules are the basis for subsequent development of a business rules repository (Ram & Khatri, in press).

In much of the current literature (Sandifer & Von Halle, 1993; Von Halle, 2002), the expressiveness of business rules is limited to whatever can be expressed in the ER-model. Some other studies including Halpin (2001), Ram and Khatri (in press), Rosca et al. (2002), and Ross (1997) try to capture business rules using a more general, conceptual modeling methodology.

Ram and Khatri (in press) present a taxonomy for set-based business rules and describe an overarching framework for modeling set-based business rules. This framework formalizes the constraints that cannot be captured with any existing conceptual model and integrates with an extant conceptual design methodology in a seamless but flexible way. Rosca et al. (2002) propose a high-level architecture for the business rules environment, in which business rules are attached to an enterprise model that adopts the enterprise modeling paradigm of LiveModel (Martin & Odell, 1997) to represent both objects and processes of an enterprise.

A somewhat different approach by Ross (1997) formally defines the semantics of business rules and then associates the rules with a conceptual model. The Ross method extends the syntax of ER diagrams by providing an elaborate diagrammatic representation for different types of business rules. Another similar example is the object role model (ORM; Halpin, 2001), which was originally intended for conceptual modeling of databases and was then extended to include an extensive graphical notation for representing constraints.

In this paper, we propose a semantics-based approach to managing an integrated view of business rules by mapping them to an enterprise data model. Our approach is similar to prior research such as Ram and Khatri (in press) and Rosca et al. (2002) in the sense that it focuses on integrating business rules with the enterprise data model. However, our framework is significantly different from methods such as the Ross method and the ORM in that it does not model multiple business rules in a single data model.

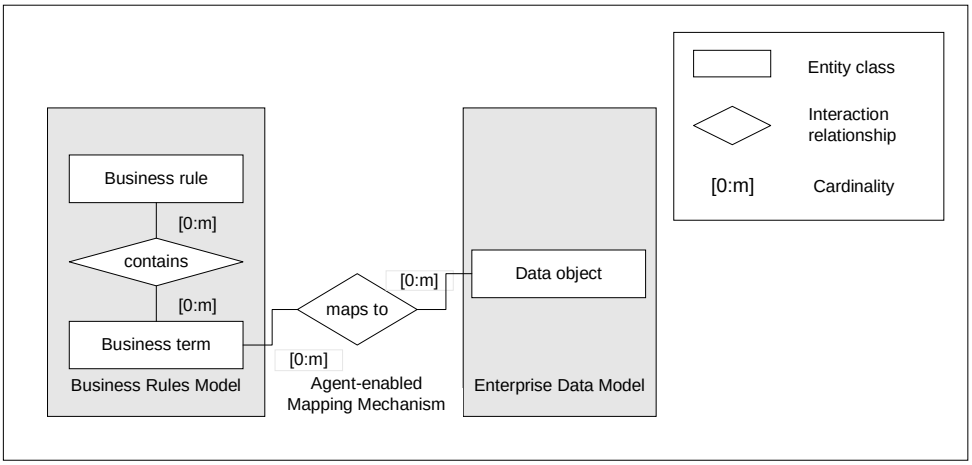
OUR SEMANTICS-BASED APPROACH

An overview of our approach is presented in Figure 2. It consists of three parts: a business rules model, an enterprise data model, and an agent-enabled mapping mechanism linking these two models. The enterprise data model defines the data objects about which business rules are expressed. We assume that a predefined enterprise data model is available. The business rules model captures semantics of the business rules. The agent-enabled mapping mechanism maps business terms (used in expressing business rules) to data objects, usually attributes, in the enterprise data model.

Business Rules Model & the Enterprise Data Model

The business rules model captures the semantics of business rules. As we noted earlier, ideally, business rules should

Figure 2: Rule-Sourcing Approach



be expressed in a language that is natural to business experts (Whitten, Bentley et al., 1994). Natural language, however, may include rambling and imprecise expressions and logic, which makes the representation of business rules less amenable to the application of formal methods or tools that would permit some form of automated reasoning (Rosca et al., 2002). Therefore, the need to capture semantics of business rules and formalize business rule expressions is essential to the execution and understanding of business rules. To formalize business rules statements, our framework adopts the Business Rules Group's idea that business rules are formed of standardized business terms, which are words or phrases that have specific meaning for a business in some designated context (Business Rule Group, 2000). This business rules model formalizes business rules using terms retrievable from a common ontology. Here we use the term common ontology as a vocabulary of business terms (concepts) with agreed-upon definitions.

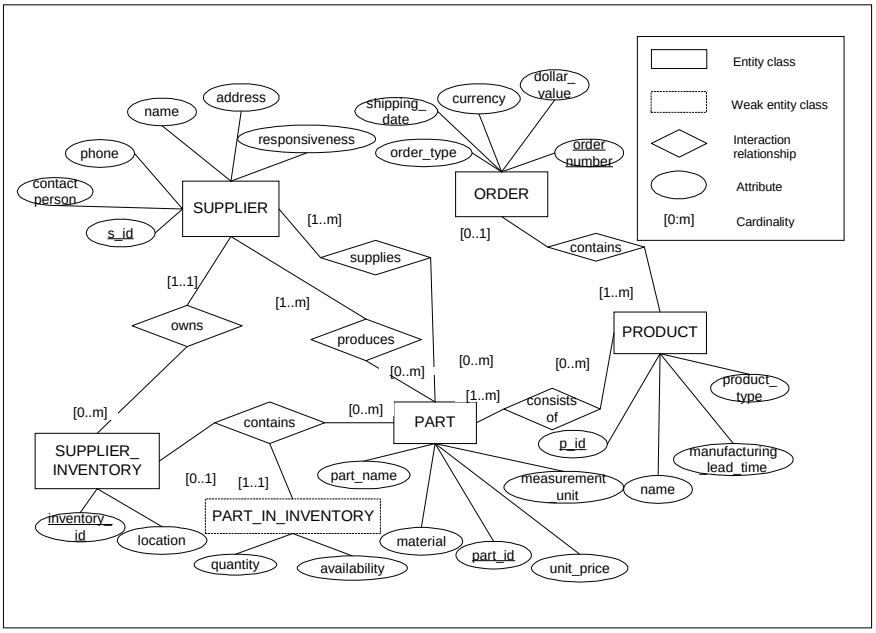
An example of a business rule formulated in our model, with business terms highlighted, is shown in Figure 3. This business rule consists of three business terms retrievable from the ontology: order-criticality, order-size, and order-importance.

The enterprise data model is a semantic model developed to provide a precise and unambiguous representation of an organization's information content (Peckham & Maryanski, 1988). In our work, the enterprise data model captures

Figure 3: Example of Business Rule

If **order-criticality** is very low and **order-size** is small
then **order-importance** is very low

Figure 4: Subset of the Enterprise Data Model



the information content that is spread across multiple data sources in the enterprise and represents the universe of discourse to which the business rules apply. This enterprise data model is constructed in the form of an entity-relationship model (Chen, 1976). Figure 4 shows a part of the enterprise data model for a manufacturing organization. It contains entity classes such as ORDER, PRODUCT, PART, SUPPLIER, and SUPPLIER_INVENTORY. Each entity class has a set of attributes whose values are retrieved and evaluated in the execution of business rules.

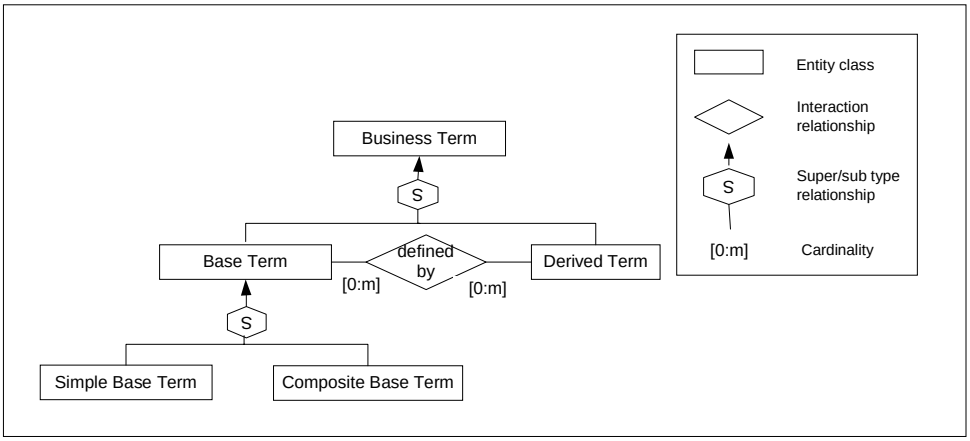
An Agent-Enabled Semantic-Mapping Mechanism

This semantic-mapping mechanism driven by a semantic-mapping agent establishes a link between the business rules and the enterprise data model. To connect business rules with the enterprise data model, the semantic-mapping agent performs the following steps.

- It identifies business terms in the business rule.
- It maps the identified business terms to data objects in the enterprise data model and thereby creates a data structure required by the business rules.
- It issues change notifications whenever a change to an established mapping should be made.

Among these steps, the most essential one is to establish a mapping between terms in the business rules and data objects in the enterprise data model. Such a mapping is necessary because the busi-

Figure 5: Business Terms Taxonomy

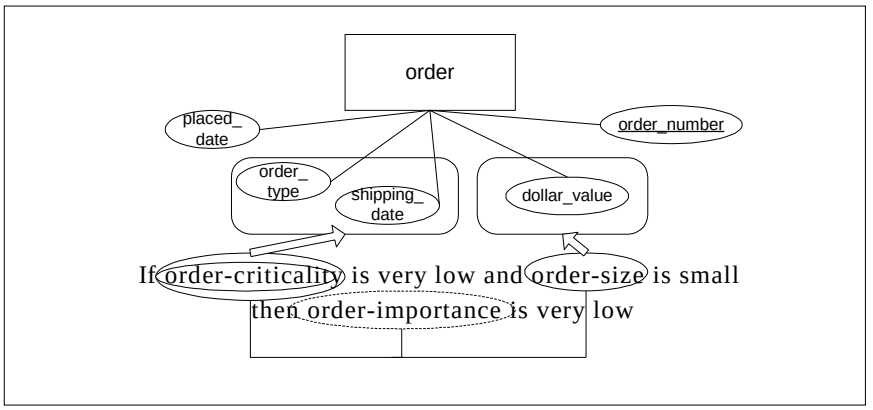


ness rules are specified independent of the form and availability of the actual data, whereas the enterprise data model captures the definition of the various enterprise information sources and must be updated if new sources are added or old sources are deleted. Mapping between business terms and data objects is not always one to one. Sometimes, a business term is associated with multiple data types, and some business terms are not directly connected with any data object. A classification of business terms based on different ways of being mapped to data objects is shown in Figure 5.

In this classification, base terms are those that are directly mapped to one or more data objects in the enterprise data model. Base terms can be further classified into simple and composite base terms. A simple base term is mapped to a single data object while a composite base term is mapped to multiple data objects. A derived term is defined by multiple terms, indicating that the value for this term is determined by values of a set of base terms. An example of mappings of different types of terms is shown in Figure 6. The ellipses signify business terms. An ellipse with a solid line denotes a simple base term; an ellipse with double lines denotes a composite base term and an ellipse with a dotted line is a derived term. The term order-size is a simple base term mapped to the attribute dollar_value in the enterprise data model. The composite term order-criticality is mapped to order_type and shipping_date, indicating that the value of order criticality is determined by the values of order_type and shipping_date obtained from one or more databases. The term order-importance is derived from the base terms order-size and order-criticality, and it cannot be mapped to any attribute in the data model, meaning that the value of order-importance is determined by the values of order-criticality and order-size.

This taxonomy of business terms is useful because the semantic-mapping agent employs different mapping schemes for different types of terms. The mapping

Figure 6: Example of Business Rule Mapping

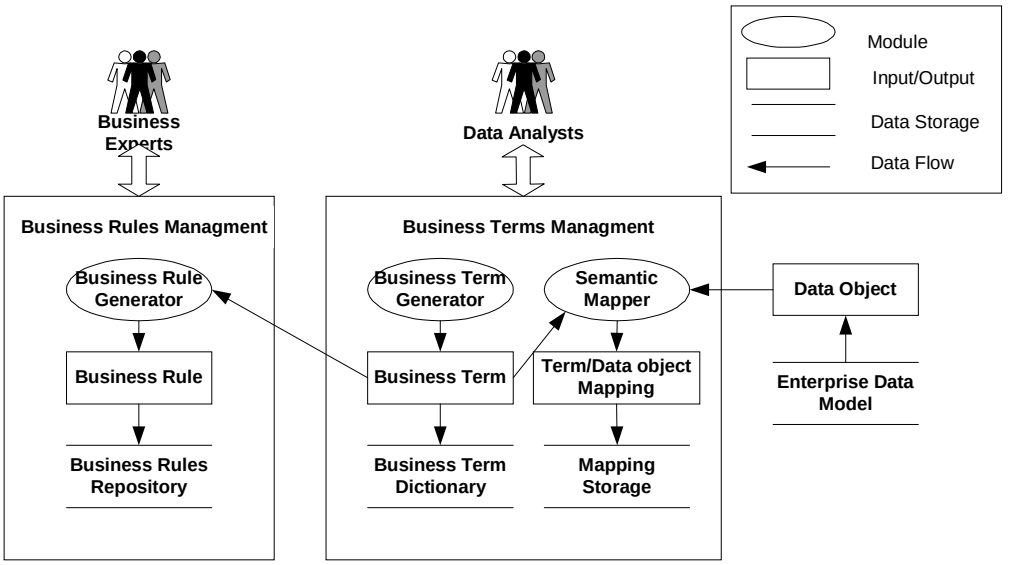


for a base term is normally predefined and stored in the agent's knowledge base as it requires specialized knowledge of both the business terms and the data model, whereas the mapping for a derived term is created at execution time. While creating mappings for a derived term, the semantic-mapping agent collects mapping information for base terms associated with the derived term.

The mapping agent also maintains the established mappings between business terms and data objects in the enterprise data model. As actual data of an enterprise reside in distributed information sources, changes made in local systems often require corresponding modifications to the enterprise data model, which in turn may invalidate the established mappings. To maintain the mapping information, the mapping agent should be able to handle event notifications (e.g., a new data object is inserted or a data object is deleted) sent by downstream agents and inform the data analyst of the changes that could affect the mappings. For example, when the mapping agent is notified that a data object in the enterprise data model is deleted, it identifies all the base terms that are mapped to the data object and issues a mapping update notification to prompt the data analyst to recreate the semantic mappings for those affected business terms.

In summary, we described an approach that captures the semantics of business rules and proposed an agent-enabled mapping mechanism that links business rules with associated data objects in the enterprise data model. We believe this approach addresses the challenges raised earlier. It bridges the semantic gap between business rules and the enterprise data model by mapping business terms in business rules to data objects in the enterprise data model. It allows decision makers to formulate and specify business rules using business terms rather than terms in the data sources. Additionally, its loosely coupled structure allows changes in business rules to be easily accommodated without having to change the whole system.

Figure 7: Architecture of the Business Rule Sourcing System



SYSTEM FOR BUSINESS RULE SOURCING

In this section, we present the architecture and implementation of a prototype system designed to manage business rules and to facilitate semantic mapping between business rules and data objects in the enterprise data model. This prototype system was developed to demonstrate the feasibility of our approach presented in the preceding section. The prototype is implemented in Java and is accessible via a Web-based, graphical user interface.

Architecture of the Business Rule Sourcing System

Figure 7 illustrates the overall architecture of the system, which consists of two subsystems: the business rules management subsystem and the business terms management subsystem. The former allows the specification of business rules while the latter enables mapping of the terms to the enterprise data model. The system provides an integrated and collaborative facility for managing business rules and mappings between rules and data objects with each subsystem aiming to meet the specific needs of various user groups (e.g., business experts and data analysts).

The business rules management subsystem contains a business rule generator module that allows business experts to specify business rules. This module provides a business rule template that displays a list of existing business terms and directs business experts to formulate business rules in business terms. The business rules specified by business experts are then stored in the business rules repository.

The business term management subsystem encompasses a business term gen-

erator and a semantic mapper. The business term generator captures basic information regarding business terms and stores the terms in a business terms dictionary. The semantic mapper facilitates creation and maintenance of mappings between business terms and associated data objects in the enterprise data model. This semantic mapper is coupled with the mapping storage repository in which the mappings are saved. The various repositories are all implemented using a state-of-the-art, relational database management system.

Operation of the Business Rule Sourcing System

Our system is designed to support the tasks of formulating business rules using business terms and mapping business terms to data objects in the enterprise data model. In this subsection, we provide a brief summary of the entire process, from specifying business rules and terms to mapping terms to data objects, using the system.

Figure 8 illustrates how a business term, order-size in this example, is defined using our system. The first window with the title "Open Term Dictionary" displays business terms stored in the business term dictionary and serves as the starting point for adding new business terms or updating existing terms. The window in the middle is for capturing general information about a term including its brief description and the authorizing business expert. The data analyst acquires knowledge of the business term by consulting with business experts and decides on the classification of the term at this stage. After the business term is generated and saved in the business term dictionary, the third

Figure 8: User Interface for Business Term Management

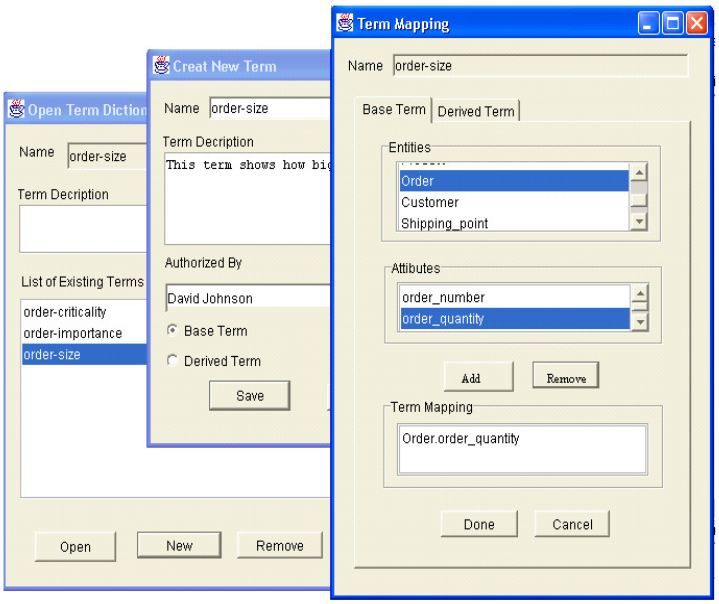
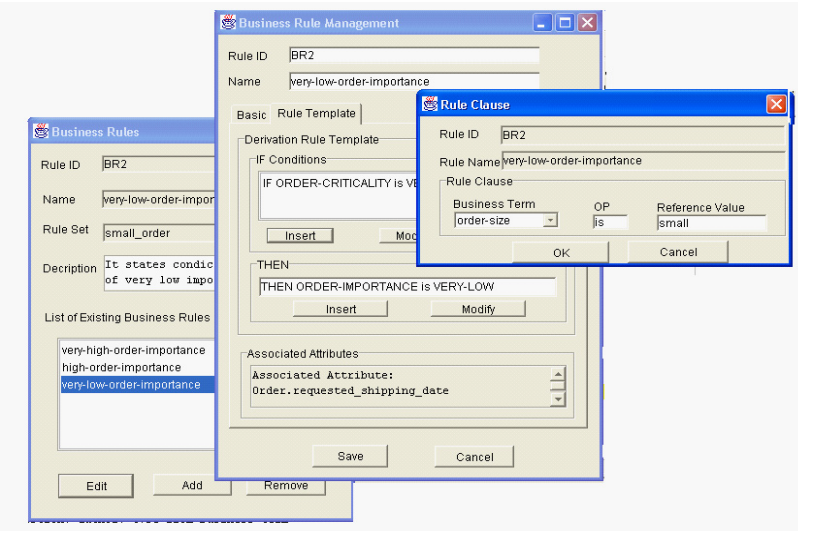


Figure 9: User Interface for Business Rule Management



window appears to display all the entities and their associated attributes contained in the enterprise data model. It prompts the data analyst to map the business term to one or more data objects in the enterprise data model. In our example, via a single mouse click, the data analyst creates a mapping between the business term order-size and the attribute Order.order_quantity in the enterprise data model.

As described earlier, the business expert invokes the business rule generator to specify business rules. The user interface of the business rule generator is shown in Figure 8. The core of the business rule generator is a rule template shown in the second window in Figure 9. As discussed in the preceding section, effective expression of businesses rules at the business level requires formative guidelines to preclude naïve use of natural language. Thus, the rule template that specifies sentence patterns of business rule statements is an important part of our system. It guides the business expert to formulate a business rule in the required syntax. As shown in Figure 9, whenever the business expert wants to insert an "if" or "then" statement, the rule template is overlaid by a window that allows him or her to form a rule clause by selecting a term defined by data analysts in the business term management subsystem. In our example, the business rule named very-low-order-importance consists of business terms order-criticality and order-size. In this case, since these business terms have already been mapped to data objects in the enterprise data model, the system automatically displays the attributes associated with the business rule.

CONTRIBUTIONS & COMPARISON WITH OTHER APPROACHES

As mentioned in "Related Work," our work draws upon research done by Ram and Khatri (in press) and Rosca et al. (2002). Our approach is similar to these approaches in that it captures the semantics of business rules and links them to the enterprise data model which is represented using the USM (Ram, 1995). However, our approach is different from their approaches. While Ram and Khatri focus on formalizing set-based database integrity rules, we focus on business rules that represent policies and decisions of an enterprise. The business rules we investigate in the previous sections are classified as derivation rules in the literature. Our approach is different from the one proposed by Rosca et al. in that it maps business rules to the enterprise data model to provide business rule sourcing, while Rosca's approach focuses on the traceability (Kilov & Simmonds, 1997) of business rules to the components of the IT system by attaching business rules to data objects and processes in the enterprise model.

Our approach also differs from approaches such as the Ross method that incorporate business rules directly into a data model. As mentioned in the section titled "Our Semantics-Based Approach," our approach adopts a loosely coupled structure. Instead of modeling business rules of an enterprise in a data model, our approach integrates the two with a semantic-mapping mechanism, loosely coupling business rules and the enterprise data model. The loosely coupled structure leads to several advantages. First, incorporating business rules into a data model constrains the capability of the conceptual model to adapt to changes in business rules. In this paper, we have highlighted the special nature of business rules as being subject to constant changes. Therefore, it is not appropriate to build them routinely into a database model (Business Rule Group, 2000). As illustrated in the previous two sections, our approach provides high flexibility because it accommodates possible changes to business rules and data objects.

Second, given a large number of business rules for an enterprise, it is extremely difficult, if not impossible, to construct and manage such a unified conceptual model of business rules. In contrast, our framework keeps the enterprise data model intact and integrates business rules with the enterprise data model with a flexible mapping mechanism described above, thus making it easier to implement than the Ross method.

Third, a conceptual model is used to describe the structure of the application domain in a way that is easily understood and validated by the domain experts (Halpin, 2002). Accordingly, a conceptual business rule model needs to be understood and validated by business experts who have little knowledge about the existing enterprise data model. The semantic gap highlighted above makes it impossible to directly incorporate business rules expressed at the business level into the enterprise data model. In comparison, our framework successfully bridges the busi-

ness and system design of business rules services. It relieves business experts from learning the enterprise data model and data analysts from transforming business rules to meet design requirements.

Last, but not least, because business rules are expected to change, externalizing them is a strategy espoused by many researchers for supporting the evolution of business rules. A single data model that incorporates business rule semantics does not conform to the principle of externalizing business rules, thus making it unsuitable for the development of a business rules management system. In contrast, by loosely coupling the business rule model and the enterprise data model, our approach is more intuitive in that it adopts the common design of current information systems featuring a centralized rule system operating on top of enterprise data management systems.

CONCLUSION & FUTURE WORK

In conclusion, our approach addresses limitations of other approaches to modeling business rules. The approach of integrating business rule semantics with the enterprise data model using a mapping mechanism benefits the business expert, the data analysts, and the end user in many ways: (a) Instead of forcing the business expert to formulate business rules in terms of data objects, they use concepts and terms that are familiar to them, (b) the data analyst maps business terms to data objects in the enterprise data model without the need to learn details of the business rules, and (c) many enterprises operate business rules on legacy systems — our framework does not invalidate existing conceptual schemas and any investments made by the enterprises.

In ongoing research, we are investigating techniques to provide automated support for rule evolution that includes the remapping of business terms when the enterprise data model evolves or when the business rules change. We are also involved in a large-scale deployment of the proposed approach and a case study to document the benefits of the approach.

REFERENCES

Boudriga, N., & Obaidat, M.S. (2004). Intelligent agents on the web: A review. *Computing in Science & Engineering, 6*(4), 35-42.

Business Rule Group, The. (2000). *Defining business rules: What are they really?* (Final Report, Revision 1.3).

Ceri, S., Gennaro, C., et al. (2003). Effective scheduling of detached rules in active databases. *IEEE Transactions on Knowledge and Data Engineering, 15*(1), 2-13.

Chen, P. (1976). The entity-relationship model: Toward a unified view of data. *ACM Transactions of Database Systems, 1*(1), 9-36.

Date, C.J. (2000). *What Not How: The Business Rules Approach to Application Development.* Addison-Wesley.

Halpin, T. (2001). *Information Modeling and Relational Databases: From Conceptual Analysis to Logical Design*. San Francisco: Morgan Kaufmann Publishers.

Halpin, T. (2002). *Join constraints.* Seventh CAiSE/IFIP8.1 International Workshop on Evaluation of Modeling Methods in Systems Analysis and Design, Toronto, Ontario, Canada (pp. 121-131).

He, M., Jennings, N.R., & Leung, H. (2003). On agent-mediated electronic commerce. *IEEE Transactions on Knowledge and Data Engineering, 15*(4), 985-1003.

Herbst, H. (1996). Business rules in systems analysis: A meta-model and repository system. *Information Systems, 21*(2), 147-166.

Herbst, H. (1997). *Business Rule-oriented Conceptual Modeling.* Heidelberg: Physica-Verlag; New York: Springer Verlag.

Jennings, N. (2001). An agent-based approach for building complex software systems. *Communications of the ACM, 44*(4), 35-41.

Jennings, N., Sycara, K., & Wooldridge, M. (1998). A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems, 1*(1), 7-38.

Kilov, H., & Simmonds, I. (1997). *Business rules: From business specification to design*. The ECOOP'97 Workshop on Precise Semantics of Object Oriented Modeling Techniques, Jyväskylä, Finland (pp. 188-194).

Leite, J. C. S. d. P., & Leonardi, M. C. (1998). *Business rules as organizational policies.* Ninth International Workshop on Software Specification and Design (IEEE IWSSD9), Iseshima, Japan (pp. 68-76).

Martin, J., & Odell, J. (1997). *Objected-oriented Methods: A Foundation.* Upper Saddle, NJ: Prentice Hall.

Peckham, J., & Maryanski, F. (1988). Semantic data models. *ACM Computing Surveys, 20*(3), 153-189.

Ram, S. (1995). Intelligent database design using the unifying semantic model. *Information and Management, 29*(4), 191-206.

Ram, S., & Khatri, V. (in press). A comprehensive framework for modeling set-based business rules during conceptual database design. *Information Systems*.

Rosca, D., & D'Attilio, J. (2001). *Business rules specification, enforcement and distribution for heterogeneous environments.* The 25th International Computer Software and Applications Conference (COMPSAC 2001), Chicago, IL (pp. 3-9).

Rosca, D., Greenspan, S., & Wild, C. (2002). Enterprise modeling and decision-support for automating the business rules lifecycle. *Automated Software Engineering, 9*(4), 361-404.

Ross, R. (1997). *The Business Rule Book: Business Rule Solutions*. Boston: Database Research Group.

Sandifer, A., & Von Halle, B. (1993). Business rules: Capturing the most elusive information asset. In B. Von Halle & D. Kull (Eds.), *Handbook of Data Management* (pp. 453-486). Philadelphia: Auerbach Publishing.

Shao, J., & Pound, C. J. (1999). Extracting business rules from information systems. *BT Technology Journal, 17*(4), 179-186.

Von Halle, B. (2002). *Business Rules Applied.* New York: Wiley & Sons.

Whitten, J. L., Bentley, L. D., et al. (1994). *Systems Analysis and Design.* Irwin: McGraw Hill College Division.

Wooldridge, M. (1997). Agent-based software engineering. *IEE Proceedings of Software Engineering, 144*, 26-37.

Sudha Ram is Ella professor of MIS in the Eller School of Management at the University of Arizona. Dr. Ram's research deals with issues related to enterprise data management. Specifically, her research deals with interoperability among heterogeneous database systems, semantic modeling, bioinformatics and spatiotemporal semantics, business rules modeling, Web services discovery and selection, and automated software tools for database design. Her research has been funded by organizations such as IBM, Intel Corporation, Raytheon, the US Army, NIST, NSF, NASA, and the Office of Research and Development of the CIA. Dr. Ram has published articles in such journals as Communications of the ACM, IEEE Expert, IEEE Transactions on Knowledge and Data Engineering, Information Systems, Information Systems Research, Management Science, *and* MIS Quarterly. *Dr. Ram serves on the editorial board for such journals as* Decision Support Systems, Information Systems Frontiers, Journal of Information Technology and Management, *and is an associate editor for* Information Systems Research, Journal of Database Management, Journal of Systems and Software, *and the* International Journal of Intelligent Information Technologies. *She has chaired several workshops and conferences supported by ACM, IEEE, and the Association for Information Systems (AIS). She is a cofounder of the Workshop on Information Technology and Systems (WITS) and serves on the steering committee of many workshops and conferences including the Entity Relationship Conference (ER). Dr. Ram is a member of ACM, the IEEE Computer Society, INFORMS, and AIS. She is also the director of the Advanced Database Research Group based at the University of Arizona.*

Jun Liu is a doctoral candidate in management information systems in the Eller School of Management, University of Arizona. His research interests include semantic modeling, business rules management, ontology, intelligent data retrieval, heterogeneous database integration, and agent-based electronic business transactions. He has an MS in management information systems from the University of Arizona.

Memes and Mutation:
Societal Implications of Evolutionary Agents in Push Technologies

Kenneth E. Kendall, Rutgers University, USA
Julie E. Kendall, Rutgers University, USA

ABSTRACT

Push technologies are rapidly moving toward autonomous and evolutionary intelligent agents for seeking, organizing, and creating information via the Web and other pervasive and innovative information technologies. We describe and define autonomous and evolutionary agents designed for push technologies. Memes, which are messages one agent broadcasts to another, causing the agent to evolve are introduced and we explore how memes will influence evolutionary agents. We develop the social implications of meritorious and malevolent memes exchanged by evolutionary agents. In the conclusion we explore the interactions among humans, evolutionary agents, and memes in order to reflect upon the future. Finally, we raise a series of future research questions regarding genetic determination of evolutionary agents, whether it is possible to predict if a meme will be meritorious or malevolent, and ask whether it is desirable to legislate the evolution of agents that are evolved from malevolent memes. Our contribution is to raise awareness of the movement toward push technologies deploying evolutionary agents and its promise and caveats, as well as to provide future research directions.

Keywords: agent technology; autonomous agent; evolutionary agent; memes; push technology; virus

INTRODUCTION

In Web-based push systems, intelligent agents are used to personalize and filter out unsolicited messages. The agent is able to employ user profiles to screen out unwanted content based on a set of predetermined criteria. When agents act on their own without immediate, direct input from users, they are called autonomous agents. When agents can change based on what they observe, they can be called evolutionary agents.

In this paper we explore the world of autonomous versus evolutionary agents that support and aid information seekers

on the Web. We accomplish this by the use of illustrative examples of push technologies. We then introduce the concept of memes, messages that evolutionary agents use to communicate with each other, and how they influence evolutionary agents. The heart of the paper describes the social implications of meritorious and malevolent memes and their interplay between humans and evolutionary agents. Finally, the conclusion explores what will happen in the upcoming generations of evolutionary agents, humans, and memes, and future research directions are provided.

AUTONOMOUS AGENTS AS USED IN PUSH TECHNOLOGIES

Autonomous agents develop their rules through the use of user profiles and data mining. User profiles are general preferences set up far in advance; for example, a user might prefer an aisle seat. The agent must then determine whether the seat preference or the airline preference be given more weight. Data mining (Codd, 1995; Gray & Watson, 1998; Watson & Haley, 1997) uses a wide range of methods to determine patterns that may predict a user's preference.

Autonomous agents are used not only to automate information gathering, but ultimately to steer the user in the direction it thinks the user wants to go. For example, Amazon.com, the large online bookseller, uses agents to track customers' preferences and report the results using collaborative filtering. Recommendation systems are used to suggest which Web site a user visits (Stohr & Viswanathan, 1999).

Autonomous agents can find cheaper airfares, worthwhile books to purchase, parts to upgrade your computer, stock tips, and other opportunities. Agents may also advise stock brokers and travel planners to call you and generate e-mail messages that include suggestions for books and computers to purchase. Some of these recommendations may be perceived to be desirable while others may not.

Autonomous agents attempt to match what the content provider wants to send and what it thinks users want to receive. This assumes that an agent can actually *know* what the user wants. Agents require complex models and insights about human behavior to obtain, systematically analyze, and act on information collected from and about users.

EVOLUTIONARY AGENTS AS USED IN PUSH TECHNOLOGIES

Evolutionary agents can observe the behavior of an individual over time and notice, not only those patterns observed by the autonomous agent, but the changes in the patterns.

Individuals change. They receive education that transforms their thinking, they hear opinions of friends and family, and they may also be evangelized or recruited so that they see everything in a new light. Evolutionary agents notice these changes and act accordingly. Evolutionary agents deliver what the user *needs*, not merely what the user *wants*.

Table 1: Elements of Autonomous and Evolutionary Agents

Type	Analysis	Design
Autonomous Agent	Determining what the user really wants	Providing content that the user wants to the user
Evolutionary Agent	Learning what the user really needs	Pushing content that the user needs but may not articulate

This is the basic difference between the autonomous agent and the evolutionary agent. Table 1 shows the analysis and design issues associated with autonomous and evolutionary agents.

Alexa is one of the first user-oriented products to attempt to create an evolutionary agent, although Alexa falls somewhat short. First described as a "navigation utility," an Alexa tool bar works with a browser to accomplish what the search engines do and more. When a user visits one Web page, Alexa suggests other pages to visit. In effect, Alexa pushes information on a user when they try to pull similar information from the Web. It accomplishes this in three ways: by examining typical usage paths of the Alexa community, by using data-mining techniques to identify patterns in particular clusterings of sites that are similar, and by listening to user recommendations. Alexa, however, is considered to be spyware. Here lies one of the problems associated with evolutionary agents.

Every time a user reads a story from the Web, an evolutionary agent will be taking notes. Then the agent will evolve based on the pages visited, the time spent on each page, and the number of times the page is revisited. The agent will also evolve when it discovers patterns in the Web sites a user visited. Finally, it will show its results to other agents and these agents will evolve based on how other agents judge the results. Advances in genetic agents in AI are making this possible. Evolutionary agents can change rapidly. An entire generation occurs each time the user visits another Web site to make a purchase, signs up for a newsletter, or registers for a warranty.

Evolutionary agents can protect users from seeing what they do not want to see. By this we do not mean that users are protected by spam filters. What we mean is that users receive information they are receptive to and are spared information that will not influence the user. Data mining tries to screen out information about products that users have no intention of buying.

This works with themes as well. A liberal might prefer not to be bombarded with conservative viewpoints and the opposite probably also holds true. An evolutionary agent would be highly effective in screening out propaganda harboring a particular bias and the user might be very pleased with the results. Only if the user strayed by pulling the opposite political views from the Web will the agent evolve.

This presents a paradox because it might be in the best interests of the person to see *both* sides of a political issue. In order to accomplish this, the agent must be evolutionary, sensing that a user actually needs a political view they might *not*

normally seek themselves. This implies that an evolutionary agent be more aware of the needs of a person than the person himself/herself.

DIFFERENCES BETWEEN AUTONOMOUS AND EVOLUTIONARY AGENTS

In order to illustrate the differences between autonomous agents and evolutionary agents, we borrow from a subset of design characteristics first proposed by Sloman (as cited in Franklin, 1995) and reexamined by Franklin who selected 10 of the more critical design considerations. We rephrase Franklin's ideas and extend his concepts to illustrate the differences in Table 2.

The uniqueness of the evolutionary agent can be expressed as the following.

1. The evolutionary agent is capable of assessing the merits of different, inconsistent objectives or motives, while the autonomous agent's actions are controlled by the most recently generated motive.
2. The evolutionary agent is capable of simultaneously storing and comparing different motives, while the autonomous agent takes on singular motives sequentially.
3. The evolutionary agent is capable of changing under the influence of genetically determined factors, while the autonomous agent can change only in the light of interactions and the inferences drawn.

Table 2: Differences Between Independent Agents and Evolutionary Agents Based on Some of Sloman's Design Considerations

Attribute	Autonomous Agent	Evolutionary Agent
Ability to compare motives	Has only one motive at a time	Can simultaneously store and compare different motives
Ability to handle multiple motives	Includes independent sources of motivation rather than a single top-level goal	Same
Ability to modify motive generators	Modification of its motive generators in the light of experience	Same
Ability to change	Can change only in the light of interactions	Changes under the influence of genetically determined factors
Ability to be examined and changed	Motive generators and comparators are merely uncontrolled side effects, not open to free will	Motive generators and comparators accessible to explicit internal scrutiny, analysis, and change
Ability to change motives under scrutiny	Influenced by how things affect it rather than the judgment of other agents	Motive generators can be influenced by likes and dislikes, or approval or disapproval, of other agents
Agent's ability to think	Agents are able to extend formalisms they use for thinking	Same
Ability to process motives	Controlled by the most recently generated motive	Assesses the merits of different, inconsistent motives
Ability to generate goals and subgoals	Independent subsystems can generate different goals	Same
Ability to use qualitative rules in solutions	Conflicts are resolved on the basis of qualitative rules	Same

4. The evolutionary agent's motive generators and comparators are accessible to explicit internal scrutiny, analysis, and change, while the autonomous agent's motive generators are merely uncontrolled side effects.
5. The evolutionary agent's motive generators and comparators change under the influence of the likes and dislikes, or approval or disapproval, of other agents. Autonomous agents, however, are influenced solely by how things affect them (Kendall, 1996).

In essence, the evolutionary agent is different from an autonomous agent in the way it is able to change.

MEMES: THE MESSAGES SENT TO AGENTS TO EVOLVE

In the previous section, we discussed the evolutionary agent and its ability to change. Three differences between the autonomous agent and the evolutionary agent are critical and need to be considered in the design of evolutionary agents. It is these differences that emphasize how evolutionary agents can and do change.

1. Evolutionary agents can change under the influence of genetically determined factors.
2. The motive generators and comparators of an evolutionary agent are accessible to explicit internal scrutiny, analysis, and change.
3. The motive generators can be influenced by the likes and dislikes, or approval or disapproval, of other agents.

The first principle above implies that an agent can change on its own, based on some preprogrammed set of rules. The second difference implies that an agent can change by self-examination. This, however, suggests that an agent can evolve if it receives information or judgment from external sources. Most likely, those sources would be other agents. Note that these three differences are similar to the three key influencing factors found in Alexa technology.

In order for these evolutionary agents to radically change, a message must be sent to the agent. In this section we explore one type of message that is broadcast to other agents. That message is called a meme.

A meme is a message that finds its way into a mind (in this case, the "mind" of an evolutionary agent) and influences events. But memes tend to make copies of themselves and manage to find their way into other minds. It is in some way like a more basic form of a virus.

A meme can include memory aids and may also include memory traps. For humans, memes include things like catchy tunes and advertising jingles, catchphrases, sales slogans, and fashion faux pas. Memes also include infectious ideas. It is obvious where the danger lies.

A successful meme is one that spreads easily. In his book, Brodie (1996, p. 36) refers to a meme as a "virus of the mind," but then goes on to say it is something that influences the infected people's behavior in a way such that they themselves help to perpetuate the virus.

Brodie (pp. 40-44) also introduces us to four types of memes:

1. instincts and programming,
2. distinction-memes,
3. strategy-memes, and
4. association-memes.

If we critically examine the foregoing types of memes, it is clear that some of them are not only useful, they are necessary for the success in our life, our careers, and our survival. It is unfortunate that Brodie grouped instincts and programming into one category since there is a considerable gulf between the two.

Instincts are inborn behavior. We were not taught to breathe, for example, we just do. Instincts are common in all animals: Birds fly south for the winter and salmon swim upstream to spawn. However, if we put our hand on a hot burner, we *learn* not to touch a burner again. In this example, programming is something we are taught or teach ourselves.

Distinction-memes are useful. These allow us to classify, categorize, organize, and label things. It is one level up from instincts and programming in terms of sophistication. Not every animal can do this.

Further up the hierarchy of order and thought are strategy-memes. We use these memes to accomplish things, like driving a car. Every time a driver sits down behind the wheel of a car, he or she does not have to figure out how to drive. Instead, the driver remembers some things (strategy-memes) that organize his thought, helping him to recognize that some ways of driving are superior to others, and then instructing him or her on how to use that knowledge to drive away.

Finally, there are association-memes. These memes link two or more memes together. If, for example, you smell something familiar, it may remind you of many things including a place, a person, a season, or a past event. We all have our own linkages. Advertisers attempt to link memes together, sometimes successfully, sometimes not. Advertisers will try to link sex with their products, for example.

If Brodie considers memes to be viruses of the mind, then is there any distinction between the two? Why are we not simply discussing viruses if that is the case? One can argue a virus is something that follows a predetermined set of instructions using some external copy machine that helps it make copies of itself. In the computer world, we use the term virus to represent something vicious or evil.

A computer virus, for example, is a program that is designed to make copies of itself on other computers. A meme, one can argue, can mutate, either in a good way or a bad way, and replicate itself without the external copying mechanism. It is therefore the basic characteristic of the evolutionary agent.

That is because evolution is not satisfied with replication. Evolution requires both replication and innovation. Innovation is a way of positively describing mutation, of course, and evolution definitely has a strong positive connotation itself.

Evolution, in the Darwinian sense, has much to do with the concept of "fitness" as in the "survival of the fittest." Memes therefore need to be designed to possess fitness, or the likeliness to be copied. Dawkins (1987) extends these concepts and introduces the concept of the "selfish gene." Dawkins believes that evo-

lution occurs because it is in the interest of a gene and its well-being, not our well-being.

A meme normally would require some sort of stimulus to evolve. The evolution of a species might occur because of a trend in the temperature, so a meme would also need something before it decides to change. In the case of a meme, it could evolve in response to some sort of crisis, problem, danger, or opportunity. Alternatively, it could evolve based on self-examination or observations of the likes and dislikes, or even the approval or disapproval, of external judges. (Note: An evolutionary agent might react to other evolutionary agents.)

USING MEMES IN WEB-BASED AGENTS

Web agents can use memes to mutate and viruses to replicate and spread content. Successful reproduction of messages can alter behavior in a way that a push provider desires. For example, advertisers that sponsor Webcasts use memes in push technology to formulate "catchy" songs, ideas, and slogans that stay with users of the media long enough for them to purchase a product or service that has been crystallized in the meme.

Higher forms of push technologies (like the delta-push technologies described by Kendall & Kendall, 1999) can be thought of as evolving memes. Preparing large segments of a population for an impending emergency such as a forecast hurricane is an example of an advantageous use of evolutionary agents passing memes. Mounting a concentrated effort for solving a large-scale transportation strike, or unifying negotiators in looking for a solution to a peace process within a short time frame might also be worthwhile uses of agents.

As evolutionary agents evolve, they will do more of the work. At one time, it was a chore to "surf" the Web to find the best airline fare, to find information about an event, or to order a spare part. Evolutionary agents understand what a user needs and blurs the line between Web push and pull technologies. As Reid (1996, p. 11) suggests, "The efficiency of pushed channel *misses*, priced-out content, and available-but-not-ubiquitous pushed information can be reduced by placing content on a shelf from which anybody can *pull* it.…The Web is precisely that shelf, its audience's interests in it becomes its distributor, and the frictionless ether its channel."

A meme can represent meritorious or malevolent values. Furthermore, one may believe a meme to be meritorious, while in reality it turns out to be evil. An example of this is programming one's mind to believe in some sort of political agenda, one that turns out to be hateful. This propaganda meme can be destructive.

The story of the sorcerer's apprentice is illustrative of what happens when push technologies go astray in the perpetuation of harmful memes. The sorcerer has the power to command an intelligent agent to perform a certain behavior. Without feedback, the apprentice continues unchecked until it has carried enough water to fill an entire room. Then, not only does the apprentice accomplish a task, but continues executing a task to the point

where the results of engaging with the technology actually becomes harmful.

Users can get infected through conditioning, using push systems such as newsletters, or by pulling seemingly useful items and ideas from the Web (Trojan horses). Spreading can take place by either pushing the message to another casually, or pulling the user to an attractive site to evangelize the user.

APPRECIATING EVOLUTIONARY AGENTS AND MERITORIOUS MEMES

How will individuals respond to evolutionary agents? For some, evolutionary agents will be looked on as saviors. Evolutionary agents are, of course, expected to deliver only what a user needs. But there is a paradoxical, almost ironic air to the problem of information overload, for even when users complain bitterly about being overwhelmed by the spam in their e-mail, they are at the same time pursuing steadily increasing amounts of information via the Web. Heim (1993), a modern-day philosopher, characterizes this paradox of greater information coupled with the loss of significance by stating, "With a mindset fixed on information, our attention span shortens. We collect fragments. We become mentally poorer in overall meaning."

Happily, evolutionary agents may be able to reduce the amount of unwanted information. Reid (1996, p. 10) reminds us that "[a]ll pushed information inevitably reaches many people who just aren't interested in it. Consider the Sunday newspaper…The many misses of pushed communication mean that every spot on *hit* is expensive to reach, and this expense prices a tremendous amount of would-be content out of the distribution channel." Reid has also characterized the steady barrage of broadcast information (minus customization or personalization) as a "wasted expression." Perhaps the important meritorious memes will succeed and individuals will be spared of useless information.

ADDRESSING THREATS POSED BY EVOLUTIONARY AGENTS AND MALEVOLENT MEMES

As we learned from the foregoing discussion, memes are a basic unit of cultural transmission or imitation (Dawkins, 1987, p. 27). There would be no problem if (a) the content and process involved in memes was pure in heart and (b) the virus reproduces memes perfectly. Unfortunately, memes can be malevolent (and anything within a spectrum of behaviors from meritorious to malevolent). Furthermore, the virus, or virtual copy machine, so to speak, is not perfect. Mutations occur and the copy is modified. To be even-handed in this regard, however, it should be duly noted that mutations can make a meme defective, or alternatively, mutations can actually improve a meme. In any case, neither the meme nor the virus is perfect.

Memes can be destructive. The future of push technologies can be charac-

terized by examining what happens when evolutionary agents push memes that mutate, affecting and infecting other agents, other users, corporations, and societies that are susceptible, with no known antidote for the meme. A dark scenario would feature an evolutionary agent that propagated a malevolent message of racial hatred, aimed at propelling a country into war against any number of targets, perhaps even to the extent of attacking its own citizens. Push technologies using evolutionary agents may create messages, and the situations for reception of those messages, that are so memorable as to eradicate all other memes or other forms of thinking.

There are many tactical, rather than strategic, ways to address the problems discussed earlier as part of the threats of evolutionary agents pushing malevolent memes. Remedies can be fashioned that create a culture of awareness of the effects, attributes, problems, and responsibilities inherent in the deployment of evolutionary agents via push technology. We recommend that individuals and organizations start small and work on many tangible problems, the artful solving of which will demonstrate tangible results and depict the possibilities available for tackling larger themes. As adoption of remedies and resolutions become more widespread, they can also be supported through systematic adoption.

A society can seek to address the experiences of users unaccustomed to advanced, evolutionary agents who push memes (for good or ill). They include underscoring the value of such as technology and having cultural education that reframes users' perceptions of how information technology is revolutionizing the entire society (Kendall & Kendall, 2000).

In very basic terms, this first remedy might be called "know the foe." That is, in order to mitigate the worrisome aspects of push technologies, the society at large should foster greater understanding of and appreciation of the situation at hand. Postman (1985, p. 157) states, "To be unaware that a technology comes equipped with a program for social change, to maintain that technology is neutral, to make the assumption that technology is always a friend to culture, is, at this late hour, stupidity plain and simple."

Another researcher describes the problem as that of curtailing user control without a giveback of linear narrative found in the content and technology of traditional broadcasting (Johnson, 1997). How this actually translates into everyday life is that the technology of push, including evolutionary agents, is ubiquitous, but is not disruptive of routine tasks. So it can sit on a "desktop," present but not engaging the user with a narrative structure. The very ubiquity of push media means that their capacity for good story telling is diminished. While many adults might argue that stories are for children, persuasive sources point to the importance of narrative in all we do. For more reflections on the adoption of new push technologies and how they might impact narrative structure in organizations, see Kendall and Kendall (1995).

The second remedy is centered on proper education. As many users as possible need to know about the design, use,

and impacts of evolutionary agents in push technology. Educating people about how the adoption and use of new media will change the structure of our discourse is imperative. The process should begin with educating those who hold decision-making power in organizations and other users about the pressing need for balance in the use of new media. Another group that requires education on impacts of evolutionary agents is the current push providers. They need to fully comprehend the power of memes as well as the criticality of pushing coherent stories that come together in a meaningful way, rather than flinging fragments about at the bottom of a newscast or on the side of a title slide.

CONCLUSION

In this paper we have introduced the concept of evolutionary agents embedded in push technologies that are capable of evolving rapidly. We have compared and contrasted the differences and similarities between autonomous agents and evolutionary agents, and highlighted some of the key areas of departure between the two. Chief among these differences are five attributes that evolutionary agents handle in markedly different ways than autonomous agents. They can simultaneously store and compare different motives; they can change under the influence of genetically determined factors; their motive generators and comparators are accessible to explicit internal scrutiny, analysis, and change; their motive generators can be influenced by the likes and dislikes, or approval or disapproval, of other agents; and evolutionary agents can assess the merits of different, inconsistent motives. We can conclude that evolutionary agents have indeed arrived and are being embraced by designers of push technologies and content providers. Evolutionary agents are here to stay.

Evolutionary agents evolve by their encounters with messages they exchange with other agents. Messages can also be called memes. Recall that memes were defined as a basic unit of cultural transmission or imitation. Memes can be either meritorious or malevolent.

One possible approach to take in dealing with evolutionary agents is to redefine the memes that evolutionary agents exchange so that they can become highly relevant to managers who encounter them in their push systems. It is this aspect of emblematic usefulness, along with the capacity for *both* human and machine control, that is one of the most important implications of memes for managers and how they work.

According to Lissack (2004),

> *If an environmental niche has an important managerial role, paying attention to its symbols and affordances can also be important. Memes are stripped of their casual role and instead become semantic tokens capable of evoking ascribed meanings. It is the process of evoking and the efficacy of the meme as the trigger for attention, recall, and petition of the ascribed meaning that gives memes relevance to managers. (p.2)*

The very nature of memes translates into a phenomenon that does not respect these traditional lines of authority and re-

sponsibility (which have typically been the building blocks of analysis for this type of study). Rather, memes permeate what seems impermeable, running freely across well-entrenched boundaries, making new acquaintances and rudely intruding on others.

The trick for humans is to still retain the possibility of control of memes that could easily mutate out of control. So, the law of requisite variety, familiar from general systems theory, comes into play for designers of evolutionary agents. That is, for every way for a meme to go out of control, there must be an equal number of ways to control it. These can and should consist of both human- and machine-centered approaches. Indeed, the most potent combination for a solution would be one that takes advantage of both human and machine capabilities to stop a malevolent mutation of a meme from spreading. The intelligent agent may, after all, know best, but the human is the one being served through most of their interactions with the evolutionary agent, and not vice versa.

In the last section, we discuss likely research directions as well as pressing research questions arising out of the advent of evolutionary agents and the memes that they exchange.

FUTURE RESEARCH DIRECTIONS

There are many areas that we have introduced in this article that deserve further reflection, explication, and research. We have, out of necessity, raised more questions here than could possibly be answered. Other researchers will want to undertake the study of these questions using both quantitative and qualitative methods. Indeed, interdisciplinary approaches to these problems that address complex computational questions as well as ethical and social dilemmas are clearly mandated by the intricacies of the scenarios we are seeking to understand.

One of the most basic questions raised here that will need further inquiry is, *How can one tell whether a meme is meritorious or malevolent?* A corollary of this question is, *How can you tell the difference between them?*

Research also needs to be done to find out if one can predict whether a meme will evolve into being either meritorious or malevolent. This will also help inform the question of whether one can control the evolution of memes, or whether this will be random.

Another research question arising out of this article is how to design evolutionary agents. In our paper we identified three key design elements of evolutionary agents. Therefore, we need to ask, *How much of an evolutionary agent's behavior is genetically determined?* A different way to phrase this is, *How much does the evolution of an intelligent agent depend on its original design?*

How much should the evolution of evolutionary agents depend on internal scrutiny, analysis, and change? And how much does their evolution depend, instead, on their interaction with other evolutionary agents?

The all-too-familiar problems of spam and e-mail viruses provide the basis for some further research questions about memes and evolutionary agents. Un-

wanted spam and viruses are currently being invoked by humans. We can create legislation that stops humans from doing this, but what if evolutionary agents take on this activity? Can we stop them from evolving with malevolent memes? Would it be desirable or even possible to pass legislation that controls the evolution of memes and evolutionary agents? Beyond that, what if an evil evolutionary agent disguises itself as a friendly agent? This could then belie a darker scenario.

As can be seen from the foregoing discussion, there are many important research directions to pursue in examining the relationships of evolutionary agents and memes to society. Interdisciplinary, multiple-method approaches that examine these problems systemically and fully in their many facets are bound to be those that are the most fruitful.

REFERENCES

Brodie, R. (1996). *Virus of the Mind*. Seattle, WA: Integral Press.

Codd, E.F. (1995, April 13). Twelve rules for on-line analytic processing. *Computerworld*, 84-87.

Dawkins, R. (1987). *The blind watchmaker: Why the evidence reveals a universe without design*. New York: W. W. Norton & Company, Inc.

Franklin, S.P. (1995) *Artificial minds*. Cambridge, MA: Bradford Book, MIT Press.

Gray, P., & Watson, H.J. (1998). *Decision Support in the Data Warehouse*. Upper Saddle River, NJ: Prentice Hall.

Heim, M. (1993). *The Metaphysics of Virtual Reality*. New York: Oxford University Press.

Johnson, M. R. (1997). *Webcasting*. Retrieved July 1, 1998: http://www.mindspring.com/-cityzoo/mjohnson/papers/webcasting/introduction.html

Kendall, J. E., & Kendall, K. E. (1995). The impact of using hypertext-based information systems on organizational narrative: Changing the structure of stories. In D. M. Olson (Ed.), *Proceedings of the Third International Meeting of the Decision Sciences Institute*.

Kendall, J. E., & Kendall, K. E. (1999, April 23). Information delivery systems: An exploration of Web push and pull technologies. *Communications of AIS, 1*.

Kendall, J. E., & Kendall, K. E. (2000). Individual, organizational, and societal perspectives on information delivery systems: Bright and dark sides to push and pull technologies. In R. Baskerville, J. Stage, & J. I. DeGross (Eds.), *IS2000: The social and organizational perspective on research and practice in information technology* (pp. 179-193). Boston: Kluwer Academic Publishers.

Kendall, K. (1996, Fall). Artificial intelligence and götterdämerung: The evolutionary paradigm of the future. *The DATA BASE for Advances in Information Systems, 27*(4), 99-115.

Lissack, M. R. (2004). The redefinition of memes: Ascribing meaning to an empty cliché. *Journal of Memetics: Evolutionary Models of Information Transmission, 8*. Retrieved July 8, 2004: http://jom-emit.cfpm.org/2004/vo18/lissack.html

Postman, N. (1985). *Amusing Ourselves to Death: Public Discourse in the Age of Show Business*. New York: Penguin Books.

Reid, R. H. (1996). *Architects of the Web: 1,000 Days that Built the Future of Business*. New York: John Wiley & Sons.

Stohr, E. A., & Viswanathan, S. (1999). Recommendation systems: Decision support for the information economy. In K. E. Kendall (Ed.), *Emerging Information Technologies: Improving Decisions, Cooperation, and Infrastructure.* Thousand Oaks, CA: Sage Publications, Inc.

Watson, H. J., & Haley, B. (1997). A framework for data warehousing. *Data Warehousing Journal,* (2)1, 10-17.

Kenneth E. Kendall, PhD, is a professor of e-commerce and information technology in the School of Business at Rutgers University in Camden. He is one of the founders of the International Conference on Information Systems (ICIS). He is an associate editor for Decision Sciences *and the* International Journal of Intelligent Information Technologies*, and he is on the senior advisory board of* JITTA. *Dr. Kendall was named one of the top-60 most-productive MIS researchers in the world, and awarded the Silver Core from IFIP. Ken is a fellow of the Decision Sciences Institute. His books include* Systems Analysis and Design*, sixth edition, published by Prentice Hall and* Project Planning and Requirements Analysis for IT Systems Development*, second edition. Dr. Kendall has published in* MIS Quarterly*,* Management Science*,* Operations Research*,* and Decision Sciences.

Julie E. Kendall, PhD, is an associate professor of e-commerce and information technology in the School of Business at Rutgers University in Camden. Dr. Kendall is the chair of IFIP Working Group 8.2. She has published in MIS Quarterly*,* Decision Sciences*,* Information & Management*,* Organization Studies*, and many other journals. Additionally, Dr. Kendall has coauthored a college textbook with Kenneth E. Kendall,* Systems Analysis and Design*, sixth edition, published by Prentice Hall. She is also a coauthor of* Project Planning and Requirements Analysis for IT Systems Development*, second edition. She coedited the volume* Human, Organizational, and Social Dimensions of Information Systems Development *published by North-Holland. Dr. Kendall is a senior editor for* JITTA*, was a functional editor of MIS for* Interfaces*, and has served as an associate editor for* MIS Quarterly.

adapt@Agent.Hospital:
Agent-Based Optimization & Management of Clinical Processes*

Christian Heine, University of Hohenheim, Germany
Rainer Herrler, University of Würzburg, Germany
Stefan Kirn, University of Hohenheim, Germany

ABSTRACT

There are several continuing challenges within the health care domain. On the one hand, there is a greater need for individualized, patient-oriented processes in diagnostics, therapy, nursing, and administration, and on the other hand, hospitals have extremely distributed decision processes and strong local (individual) autonomy with a high degree of situational dynamics. This research focuses on developing an information system that can substantially increase the efficiency of hospital process management. We present a system for agent-based simulation and support for clinical processes. In particular, we describe the architecture of the system and its functionalities and the integration of existing Foundation for Intelligent Physical Agents (FIPA), DICOM, and HL7 standards. We discuss an example scenario for clinical trials to illustrate how the system supports distributed clinical process management. This system interacts with other multiagent systems within the Agent.Hospital framework and hospital information systems (HISs) in the eHealth Lab. This research is part of the German Priority Research Program (SPP) 1083, Intelligent Agents and Their Application in Business Scenarios.

Keywords: clinical trials; health care; multiagent systems; simulation

INTRODUCTION

Over the last few years, the cost of health services in Germany has become an important area of economic analysis. The budget for German health care increased to 218 billion Euro during the year 2000 with a growth of 5% to 10 % every year after that. Thus, the current expense of health care is forcing the German government to take a closer look at the macroeconomic cost structures within this industry and find new ways to minimize the overall cost. Information technology is

expected to play a major role in the reduction of expenses and the stabilization of health care budgets. With a view to curb expenditure, recently, the biggest IT project for German health care was started, which has introduced a new electronic eHealth-Card until 2006[1]. One of the major problems that need to be addressed is hospital logistics, which is rather complex and highly distributed. Considerable research is underway in this area that is designed to analyze the domain as well as develop software systems, particularly for management and control of information flows and business processes. The approaches to this domain are twofold. On the one hand, simulation systems are used for analyzing the domain, planning, and reengineering of business processes. On the other hand, development of management and control systems for information flow and business processes is of interest. The increasing importance of information systems is interlinked with specifics of the (German) health care domain, such as highly fragmented, disparate professional groups, which complicates or particularly blocks the integration of innovative service processes as well as new and more-efficient information systems.

Several important questions have been examined by the German Priority Research Program (SPP 1083)[2] since 2000. The main goal of SPP 1083 is to investigate the utilization of agent technologies within large realistic business scenarios and the identification of further research needs. Basic supposition of the SPP is that agent-based development and connection of decentralized information systems generates important benefits by supporting interorganizational business processes and organizational flexibility. Examination of these hypotheses is supported by the development of large agent-based software systems.

In this paper we introduce the essential questions and the research approach addressed by the ADAPT project, which is part of the previously mentioned SPP 1083. We discuss the systems developed for simulation and application and their integration into the Agent.Hospital framework (REF), which is part of the Agentcities research network. In addition we address questions on how to realize large-scale, agent-based distributed service networks that support health care standards like HL7 or DICOM. An example scenario in clinical trials illustrates our approach. At the end of the paper we discuss the shortcomings and give an outlook on further research steps and implementation activities.

RESEARCH PROBLEM & MAIN OBJECTIVES

The overall objective of our research is to substantially increase the efficiency of hospital processes. The main problem in optimization is that the processes, as well as the systems, are inherently distributed and the effects of changes are difficult to foresee. Under these circumstances, agent technology is the appropriate solution. Optimization by agents can basically be reached in two ways: (a) Optimization potential can be identified by agent-based simulation experiments, and (b) flexible, agent-based information systems can optimize the real system. Inte-

grating both of these components, simulation models can be used as a test bed for the development of agent applications. Our empirical studies have identified the following critical success factors that help achieve the objectives of being able to optimize the system.

- Completeness and topicality of information available to (human) actors
- Knowledge about the actual patient status helps trigger other changes
- Adequate representation of the variety and the dynamics of inter and intra process interactions

Based on our experience, we contend that a lot of domain knowledge and interdependencies are necessary to create an appropriate agent-based system. On the other hand, even more knowledge would be needed by conventional information systems. The adaptive abilities of agents allow dealing with even uncertain conditions and distributed knowledge.

In cooperation with the clinical partners, the ADAPT project (Heine, Herrler, Petsch, & Anhalt, 2003) tried to optimize the processes as well as the information flow between the participating oncology and radiation therapy departments. As a result of this, resource allocation, time scheduling, and tactical planning should improve with respect to efficiency and control. The experiences gained should allow us to draw conclusions about the usability as well as the advantages of agent-based software. The specific motives and goals of this research are the following.

- **Improvement of distributed appointment scheduling.** Appointment scheduling for treatment and examination tasks in hospitals is inherently distributed between various organizational units and there are a lot of interdependencies between the processes and the actors. This fact makes it very difficult to determine the effects of single action and to optimize the process. The scheduling and negotiation strategies used in general practice are mainly dominated by simple ad hoc solutions, which can be optimized. Therefore, simulation and analysis of scheduling scenarios and finding more sophisticated processes is an important objective of the ADAPT project. Using clinical trials on chronomodulated medication, an agent-based system was designed and implemented in order to support the trial scheduling and coordination of patients, medical staff, and equipment.
- **Support for decisions about participation in clinical trials.** Clinical trials obligate the participating hospitals to perform prescribed treatments and examinations. As one of the benefits, the cost calculation should show whether the incurred liabilities are covered or not. According to this, the ADAPT system should allow cost-benefit estimations based on simulations which take into account medical considerations and individual preferences. Agent technology is used to handle the high level of environmental dynamics as well as the complex and restrictive requirements of clinical trials. Utilization and trial-specific statements are supported by realistic simulation results.

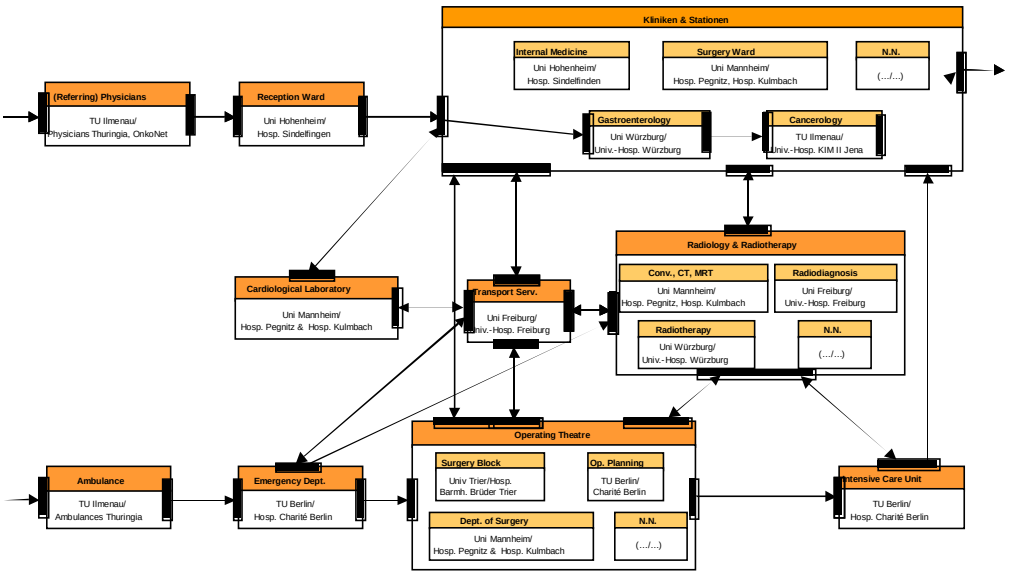
- **Operation of study protocols.** Due to the complexity and regimentation of trials, a tremendous effort is needed for the coordination of all the actors and equipment involved. A continuing goal of the ADAPT project is to advance the agent-based simulation system towards a real-time assistance system that enables online resource allocation. Supporting the operational processes by coordinating the flow of information (adapted to the individual medical pathway) using agent technology should result in a higher level of process efficiency.

AGENT.HOSPITAL FRAMEWORK

To evaluate the agent systems that have been developed and to show that they interoperate in a sophisticated way, a complex and realistic evaluation scenario is needed. Therefore, the hospital-logistics working group of SPP 1083 including the ADAPT project developed an extensive model called *Agent.Hospital (Kirn, Heine, Herrler, & Krempels, 2003)*. This model is based on an open framework for numerous health care actors. The basic architecture of the framework is shown in Figure 1. The *Agent.Hospital* framework contains several detailed health care models, different kinds of service agents, and agent-based platforms. Within *Agent.Hospital*, one can evaluate modeling methods or examine configuration problems, as well as agent-based negotiation strategies and coordination algorithms.

In the context of SPP 1083, the research partners involved in the hospital-

Figure 1: Application Diagram of the Agent.Hospital Framework with Several Selected Supply Chains



logistics working group integrated the partial hospital-logistics models created by the individual projects. Due to the high number of participating research groups, a wide spectrum of relevant clinical processes could be offered. Relevant organizational structures, processes, and necessary data models were analyzed, formalized, and modeled at several hospitals. To be able to integrate a number of separate, partial models, it was necessary to define numerous gateways between these models (see Figure 1) and to develop a common ontology for interaction through these gateways. Additionally, basic process patterns had to be defined (for instance, planning and execution of clinical trials with oncological patients). At the conceptual level, *Agent.Hospital* consists of partial models, process patterns, gateway specifications, and shared ontologies. The *resulting framework* is an overall conceptual model *based on t*he integration of project-specific, partial models.

In *Agent.Hospital*, several clinics, departments, and wards are defined as service provider and consumer units. In our framework, the services offered and the services requested are denoted simply as *services*. For a meaningful interaction between the service provider and the service consumer unit, there needs to be a well-defined interface based on existing agent communication languages, interaction protocols, content languages, as well as common ontologies.

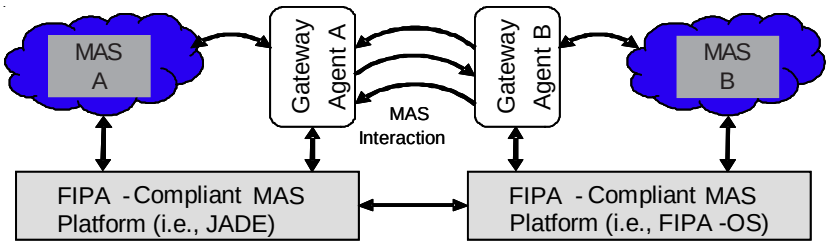
Currently the following supply chains of the overall framework shown in Figure 1 have been implemented: clinical trials and radio therapeutics (ADAPT), rescue patient (AGIL), lung cancer treatment (ASAinlog), angina pectoris (MedPAge), gallstone therapy, surgery processes (Policy Agents), and radiological service processes (EMIKA). More details and references to these projects are provided via the RealAgentS Web site, the public groupware tool of the SPP (http://www.realagents.org).

For the successful development of the Agent.Hospital framework, the use of existing and established standards was very important. In agent research, the Foundation for Intelligent Physical Agents (FIPA) plays the leading role. During the last few years, the FIPA organization has proposed primary standards for the design of agent platforms as well as standards for communication and interaction. To be able to integrate the different agent systems, it was necessary to define numerous gateways using these standards (Figure 2) and to develop common ontologies for interactions among the respective gateway agents.

Intelligent software agents in any application have to be integrated into an existing (often-proprietary) information system infrastructure. Therefore, the aim of the hospital-logistics working group is the implementation, evaluation, and documentation of agent-based health care services, which are supposed to be the foundation for future FIPA *application specifications*. Since 2002, the SPP 1083 has been a member of FIPA, and was tasked to develop and evaluate examples and to refine the existing application specifications in health care.

Besides working with FIPA, the SPP 1083 has initiated cooperation with the Association of Information System Produc-

Figure 2: Gateway Agent Concept (Krempels, Nimis, Braubach, Pokahr, Herrler, & Scholz, 2003)



ers in Health Care (VHGit) to develop specifications for interfaces between specialized applications (practice systems, patient record, etc.). Furthermore, the SPP contributes to *Agentcities* (Willmott, Constantinescu, Dale, Somacher, Marinheiro, Mota et al., 2003), an open, worldwide net of FIPA-compliant agent platforms, using the different agent-based services — often in the development stage — that are provided. Currently there are more than 140 agent platforms that are members of *Agentcities*, with a high concentration in Europe and also several platforms in the USA, Australia, and the Near East.

Within Agentcities, a lot of different application domains exist, for instance, eHealth, manufacturing control, digital libraries, travel services, and so on. The ultimate aim of this initiative is to support the commercial, as well as academic, efforts for the development of agent-based applications, and to make it possible to compose dynamic, intelligent and autonomous agents as well as complex service agents. *One goal of Agent.Hospital* was to become a part of *Agentcities*. Therefore, new *Agentcities* platforms have been set up in five German cities. They are connected by a central directory service in Aachen, which registers all participating service agents.

Different infrastructural services and tools have been set up or developed to realize Agent.Hospital.

- **Agent.Hospital Directory Facilitator (AHDF):** a common and centralized yellow pages service for the agents within Agent.Hospital. It provides a Web interface for online monitoring and several other extended functions like availability testing.
- **Agent.Hospital Event Service (AHES):** an event service realized by an agent. It can be used for event-based simulation of multiagent systems in the health care domain.
- **Agent.Hospital Actor Agent (AHAA):** another common component to all the involved projects is the actor agent. Instances of the actor agent represent real actors in Agent.Hospital scenarios.
- **Agent.Hospital Simulation Model (AHSim):** an executable simulation model describing a typical patient pro-

cess providing interfaces for agent-based application systems.
- **Agent.Hospital Ontology Repository (AHOR):** a repository providing health-care-specific task ontologies (Becker, Heine, Herrler, & Krempels, 2002) on demand. The shared ontologies are a base for mutual understanding of the systems.
- **Agent.Hospital CVS (AHCVS):** a concurrent-version system containing the source files of the service and interface agents.

TECHNICAL BASE & PROTOTYPICAL SOLUTIONS

To optimize scheduling and coordination of clinical processes, intelligent agents for clinical care were developed within the ADAPT project. A set of personal assistance agents was developed for solving coordination problems of different actors (physicians, patients, nurses) in the hospital. These agents support their specific users in different tasks such as the administration of a personal date book, making decisions, or scheduling of shifts or medical actions.

The basic foundation for the A4Care assistance system is the FIPA-compliant platform JADE. The central directory service of Agent.Hospital supports the process of finding suitable interaction partners. The agents of the assistance system have the following responsibilities.

- **Physician agent.** A physician agent can compose single tasks to a treatment plan and assign it to a specific patient. In doing so, it can specify order restrictions and possible parallel sequences. As an extension, we plan to integrate a library of predefined "clinical pathways."
- **Patient agent.** This assistance agent takes over the automatic or semiautomatic scheduling of the assigned plan. It respects the wishes of its owner and presents the results in a personal date book. The user can do manual rescheduling and define new policies including personal restrictions and preferences. The actual negotiation is then processed in the background by connecting the assistance agents to the functional units.
- **Functional unit agent.** This agent is a proxy "person" for a functional unit. It coordinates the incoming appointment requests and sends appointment proposals automatically. Here the current schedule, opening times, and additional restrictions (e.g., for reduced personal visits during lunch) are respected. A collateral monitoring of appointment execution starts an automatic rescheduling if discrepancies between the planned and actual schedules occur.

Users can grant different grades of autonomy to their agents, and decide when an agent is allowed to act autonomously and when it has to make further enquiries to the users. A high grade of autonomy on the one hand increases the overall performance of the system; on the other hand, it requires that preferences of the users be formalized very exactly and updated regularly. Additional agents to the ones de-

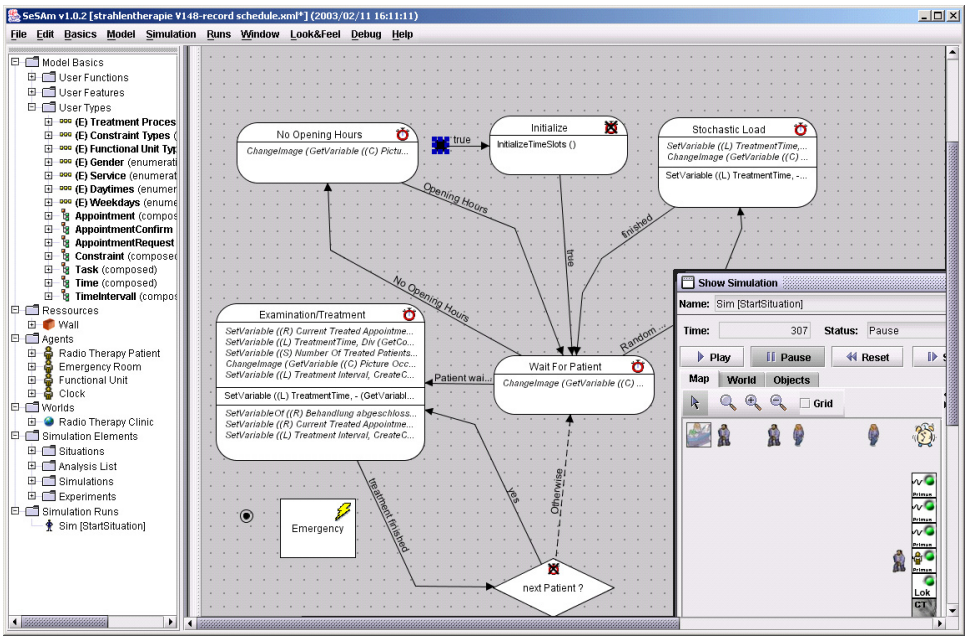
scribed above are realized for the medical staff and wards to solve the problem of making fair shift schedules. In addition to the development of these agent-based services running on FIPA platforms, we need to develop a realistic environmental model. This model is well populated with agents, this time, agents representing human actors.

An integral part of the ADAPT project is the development of the agent-based simulation system called the "shell for simulated agent systems" (SeSAm; Klügl et al., 2001). Originally developed for the simulation of social insects (Klügl, Puppe, Raub, & Tautz, 1998), it was used to implement environmental simulations for the ADAPT project as well as for comprehensive Agent.Hospital framework. SeSAm is an integrated environment for modeling and simulating multiagent systems and provides powerful modeling functions for the easy construction of complex models. As a special feature, it offers visual modeling of all the tasks to support domain experts that do not have programming skills in creating simulation models. It allows the visual construction of agents and their environment respecting their properties, abilities, and behavior. Different agent classes can be created and their properties and abilities can be defined by just adding agent-specific attributes and modular features.

There are several useful functions available within SeSAm, for example, movement, evolution, and scheduling, and the given set can be extended by the user. Using these features and the basic set of primitive actions (e.g., Move, ChangeImage, SendMessage) and functions (e.g., GetAllObjectsOnPosition), the behavior of the agent can be described. The basic representation of the agent's behavior is the UML activity diagram (Figure 3). It is easily understandable and uses common notations well known to system developers. Activity diagrams consist of activities and transition rules. The activities can be seen as states containing a series of actions, which are executed until an exit rule activates the next activity. After modeling the agent classes, simulation situations can be created on a two-dimensional map. For the simulation of a scenario, this model representation is compiled and can be run directly from within the environment. During simulation it is possible to watch the animated map, pause, and resume the simulation, and to control agents' states.

Recently, several additional features to support handling of more-complex models like health care scenarios were integrated in the environment (Oechslein, Klügl, Herrler, & Puppe, 2002). Among various other features, the environment allows hierarchical modeling at the level of function description, as well as at the level of behavior description. Basic function primitives can be composed to more-complex user-defined primitives, and parts of the activity graph can be combined together to form compound activities. To simplify testing and finding errors in the model, it is possible to create rules for invariant states, similar to assertions in modern programming languages. Even refactoring methods are supported, which are also quite popular in traditional pro-

Figure 3: Visual Modeling & Simulation with the SeSAm Environment



gramming. Experienced modelers can get the power of a programming language without having to learn any syntactic notations.

To support the connection to external agent systems, additional extensions were required. Therefore, new features to support communication as well as ontologies were added. Here, general standards were adhered to in order to ensure interoperation with other systems. The new plug-ins provide FIPA-compliant communication primitives and the possibility to import ontologies modeled with the ontology tool Protege 2000.

In our view, agent-based applications are part of an environment containing real-world agents like users and other systems (possibly agent systems). In agent software development, isolated tests are usually very difficult; instead, tests have to be carried out under the conditions of this environment. As mentioned above, our goal was to extend SeSAm in a way that it can be used as a test bed for agent-based software without remodeling the agent for interacting with the simulated environment (Klügl, Herrler, & Oechslein, 2003). Then, creating models of the global system becomes rather intuitive. Figure 4 shows a sketch of this approach. An agent-based application is integrated in an environment consisting of additional information systems and users, which can also be seen as agents. All agents from the real-world environment are replaced by simulated agents in SeSAm. The level of detail in the simulation is dependent on

Figure 4: Engineering Agent Systems in a Simulated Environment vs. Application in Real World

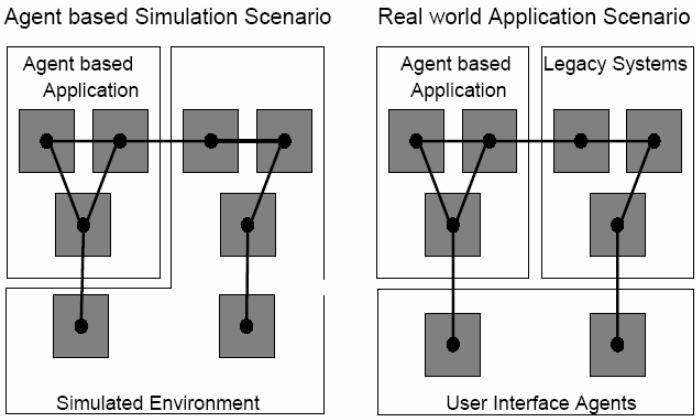
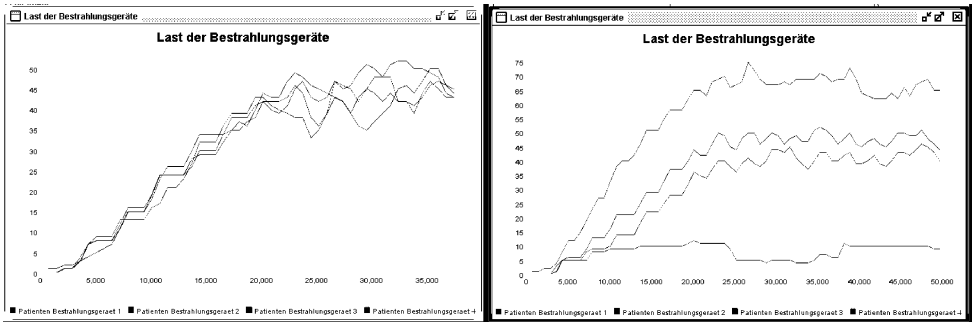


Figure 5: Online Graphical Analysis of Simulated Processes



the requirements of the agent system. Due to the representation of the real world in the simulated test bed, the developed components can be tested both under realistic as well as extreme conditions. This usually cannot be done in the real world for reasons of security and prohibitive costs. Especially in the health care domain, practical testing of prototypical software is usually dangerous. In this context, SeSAm offers more-powerful capabilities for experimentation support, like the distribution of simulation runs on several hosts and online graphical analysis of the simulation results like the ones shown in Figure 5.

A SCHEDULING SCENARIO

The example scenario discussed here relates to scheduling in the radiation therapy units. Based on empirical investigations, we have implemented a detailed SeSAm model describing the patient scheduling in a clinic for radiation therapy.

The clinic consists of a reception area, where the patients' appointments are planned, waiting rooms and treatment rooms, as well as radiation devices. Functional units and devices have specific opening times and a current schedule. One of the typical "patient processes" for radiation therapy is assigned by the doctor after a patient's arrival. Often, these processes are also referred to as clinical pathways. Once the pathway is assigned, the necessary medical actions (treatment and diagnosis) have to be planned and the patient has to visit the functional units. A typical problem is that the schedule of the real treatment execution always differs from the preplanned schedule. Here we wanted to find new scheduling strategies to improve the process, taking into consideration predefined quality criterions like the waiting time of patients and the utilization ratio of resources. Extensions of the model can also simulate late patients and uncertain treatment times.

Different settings representing the situation in other clinics can be created very easily. The number and kinds of functional units as well as their opening times can be varied (Figure 6). Alternative clinical pathways can be assigned to patients. The probability of a new patient's arrival, as well as potential tardiness, is adjustable. All this makes the model a flexible testing environment

Figure 7 shows a screenshot of a running simulation. Some patients are waiting for admission. The selected patient is waiting in the queue in front of the tumor localization room (called "Lok"). This unit currently has opening hours (shown by the green light) whereas the radiation therapy units are still closed (indicated by the red light). Figure 7 also shows the schedule of the selected patient and the appointment for "Vorlok," which is also known as the localization room.

Based on this model, it is possible to evaluate different strategies, for example, a first-come, first-served or preplanned schedule that takes into consideration priorities of examinations. Different hypotheses of dynamic changes (patients that come too late, additional emergencies) can be evaluated with their effects on evaluation criterions and plan stability.

As a first step, our experiments were able to show plausible reactions of the system that are similar to current real-world behavior. In the next step, we want to connect the simulation engine to the negotiation-based, reactive planning system A4Care, which is very effective in a distributed dynamic environment (Herrler, Heine, & Klügl, 2002).

A CLINICAL-TRIAL SCENARIO

An important field in health care is the medical treatment of cancer. Typically, these treatments are very expensive, and new clinical trials are conducted to test the effectiveness of various new drugs and treatment protocols. Clinical trials are carefully controlled studies in which oncology experts evaluate better and cost-effective ways to treat, prevent, or diagnose cancer. Increasing the number of qualified patients who are enrolled in clinical trials, as they represent an opportunity for patients with cancer to receive the best possible care, is really important.

Figure 6: Flexible Configuration of the Simulation (number of agents and opening times of the functional units may vary)

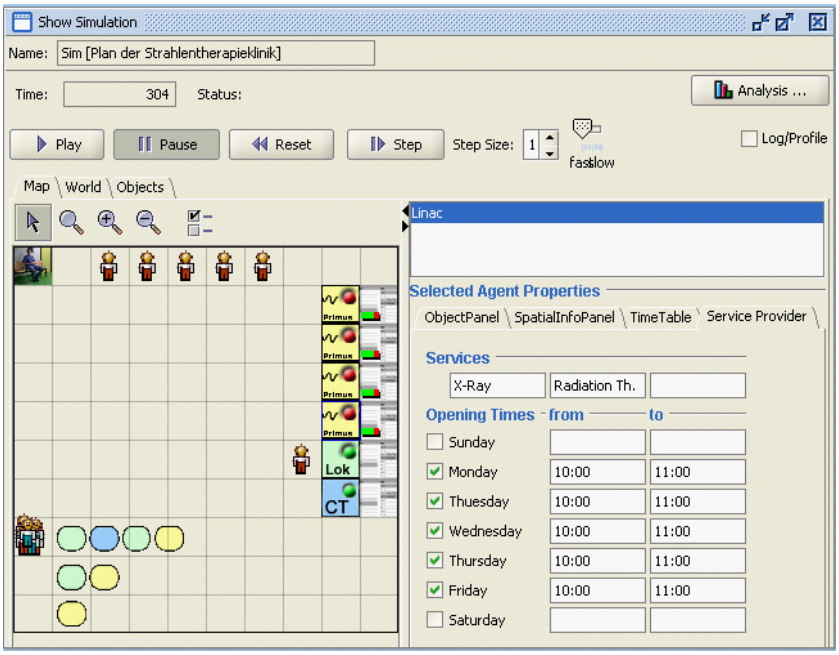
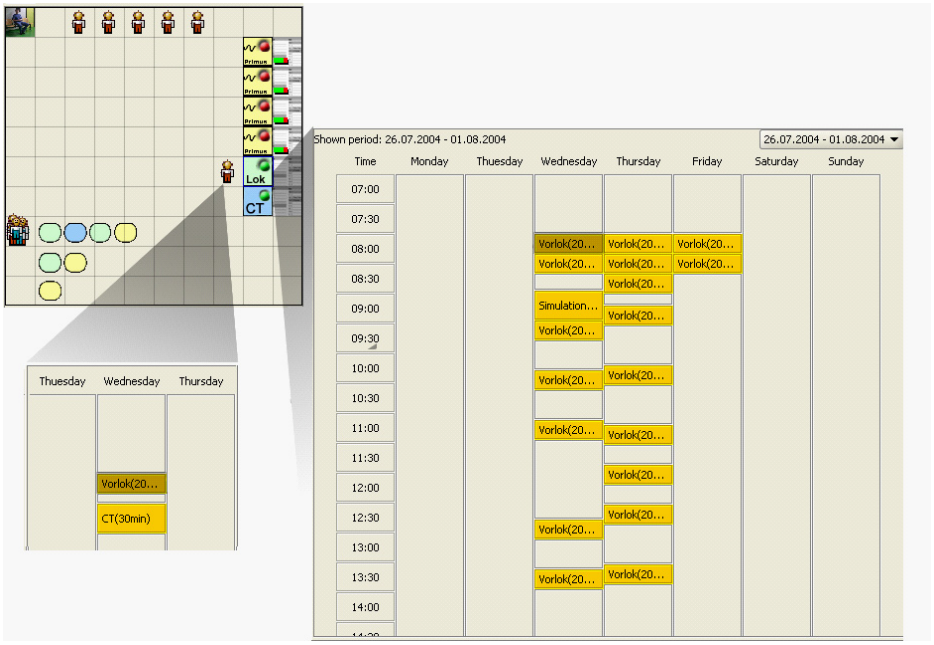


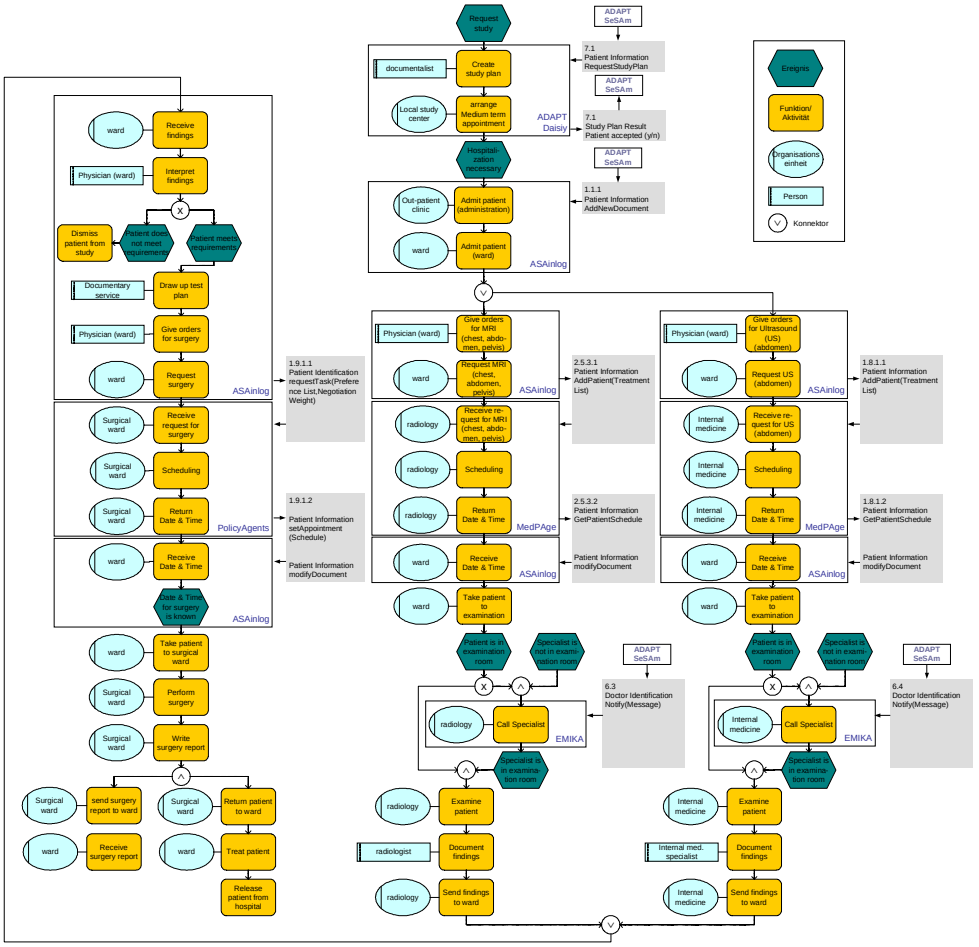
Figure 7: Weekly Schedules of Patients & Radiation Therapy Dbevices



Clinical trials are also extremely valuable because they answer important questions that will help to continually improve cancer care and decrease the risk of cancer development. Before a new therapy or medical treatment can be put into daily clinical practice, a lot of trials have to be run. These trials are detailed plans for medical treatments, for instance, a clinical trial protocol can describe at which point in time, in which quantum, and how medications or therapies have to be executed. But clinical trials are not performed in linearity; instead, their concrete structures depend on specific patient constitutions, laboratory results, and so forth. Clinical trials can be described using the graph metaphor, with each node representing a particular state of a particular study. Typical sample sizes for clinical trials require about 150 to 200 patients, which makes it necessary in general that several hospitals cooperate in performing such trials (multicentric trials). Figure 8 exemplifies the process aspects of the integrated clinical-trial scenario. The simulation according to the process description is directed by SeSAm and utilizes several agent-based services provided by the SPP 1083. At the beginning of a clinical trial, various tasks for diagnosis and treatment have to be coordinated, and resources have to be scheduled and, if necessary, communicated to the corresponding units. Figure 8 shows an example with computed tomography (CT) and magnetic-resonance-tomography (MRT) examinations, as well as the execution of a surgery.

At first, the suitability of the patient for the clinical study will be checked (age, gender, blood count, etc.). If the patient fulfills the preconditions for the trial, for instance, an MRT examination is necessary. Planning of this examination means that different service agents try to negotiate mandatory appointments (see, e.g., the gateways 2.5.3.1 and 2.5.3.2 in Figure 8). But these agent-based negotiations are constrained by the existing timetable of the corresponding HIS. To interact with this information system (precondition is HL7 compliance) the HL7 agent generates specific HL7 schedule messages like SRM. There are several health-care-standards development efforts currently underway throughout the world. Health Level Seven is one of several of the American National Standards Institute's (ANSI) accredited Standards Developing Organizations (SDOs) operating in the health care arena. The application level addresses definition of the data to be exchanged, the timing of the interchange, and the communication of certain errors to the application. It supports such functions as security checks, participant identification, availability checks, exchange mechanism negotiations, and, most importantly, data exchange structuring[3]. Figure 9 shows the HL7 SRM structure. Based on this message structure, the actual schedule message (SRM^S01) is generated, which is shown in Figure 10. The responsible HL7 agent is able to receive, edit, process, and send such HL7 messages to standards-compliant HISs. Currently, detailed tests with different HISs have been planned. The agents communicate in a FIPA-compliant way and transmit the HL7 as well as the DICOM messages as content of the FIPA agent-communication-language (ACL) messages. The American College

Figure 8: Exemplary Part of the Clinical Trial Scenario with Interactions Between Different Multiagent Systems



of Radiology (ACR) and the National Electrical Manufacturers Association (NEMA) formed a joint committee to develop a standard for digital imaging and communications in medicine. This standard is developed in liaison with other standardization organizations including CEN TC251 in Europe and JIRA in Japan, with review also by other organizations including IEEE, HL7, and ANSI in the USA. The DICOM standard pertains to the field of medical informatics. The field addresses the exchange of digital information between medical imaging equipment and other systems. Because such equipment may interoperate with other medical devices, the scope of this standard needs to overlap with other areas of medical informatics. However, the DICOM standard does not address the breadth of this field[4].

Figure 9: HL7 Schedule-request Message (SRM) Structure in Clinical-trial Scenario

```
SRM^S01 -S11^SRM_S01      Schedule-Request Message
MSH                       Message Header
ARQ                       Appointment-Request Information
[ APR ]                   Appointment Preferences
[ { NTE } ]               Notes and Comments
[ { PID                   Patient I dentification
    [ PV1 ]               Patient Visit
    [ PV2 ]               Patient Visit  - Additional Info
    [ { OBX } ]           Observation/Result
    [ { DG1 } ]           Diagnosis
  }
]
{ RGS                     Resource Group Segment
  [ {  AIS                Appointment Information   - Service
       [  APR ]           Appointment Preferences
       [ { NTE } ]        Notes and Comments
     }
  ]
  [ {  AIG                Appointment Information   - General Resource
       [  APR ]           Appointment Preferences
       [ { NTE } ]        Notes and Comments
     }
  ]
  [ {  AIL                Appointment Information   - Location Resource
       [  APR ]           Appointment Preferences
       [ { NTE } ]        Notes and Comments
     }
  ]
  [ {  AIP                Appointment Information   - Personnel Resource
       [  APR ]           Appointment Preferences
       [ { NTE } ]        Notes and Comments
    }
  ]
}
```

Figure 10: SRM for an MRT Examination

```
MSH|^~\&|||||20030101||SRM^S01|1234|1234|2.4
ARQ|0001|0002|||||0010^radiological
examination^ROUTINE|NORMAL|120|min|200303150800^200303201700|P
riorität|||4711^Mayer^Hans^^Dr.||||4712^Otto^Karl^^Dr.
PID|1111|2222|3333|4444|Mustermann^Hans|Mustermann|19500521|M
DG1|001|I9|786.5|CRANIAL PAINS|199401010730|W
DG1|002|I9|412|OLD MYOCARDIAL INFARCTION|199401010730|W
AIL|1234|||0100^MRT
```

A screenshot of the HL7 message agent interface is shown in Figure 11. Relevant message segments and fields can be edited by the user and sent to the next responsible agent or HIS.

After the MRT examination, several HL7 messages have to be sent pertaining to the results of the examination in the current scenario. But the transfer of radiological image data is historically grown and still not supported by HL7. The transfer of image data, for instance, from the MRT modality to viewing workstations, digital archives, or a remote radiologist for a second opinion is supported by the DICOM standard (our basic assumption is that all

Figure 11: HL7 Message Agent for Interactions with HL7-compliant HIS

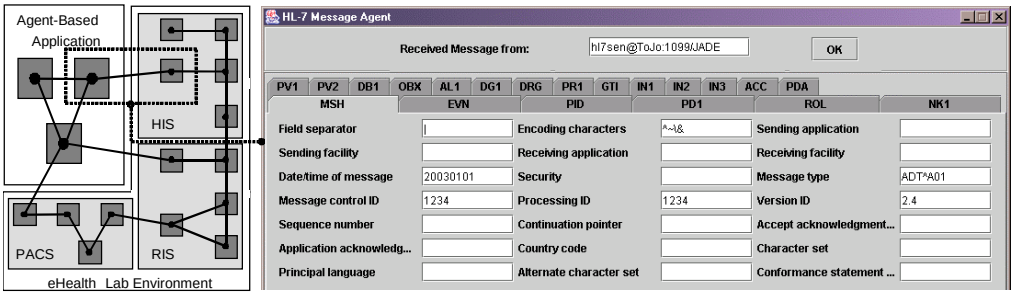


Figure 12: Simplified Interaction Diagram in the Integrated Clinical-trials Scenario

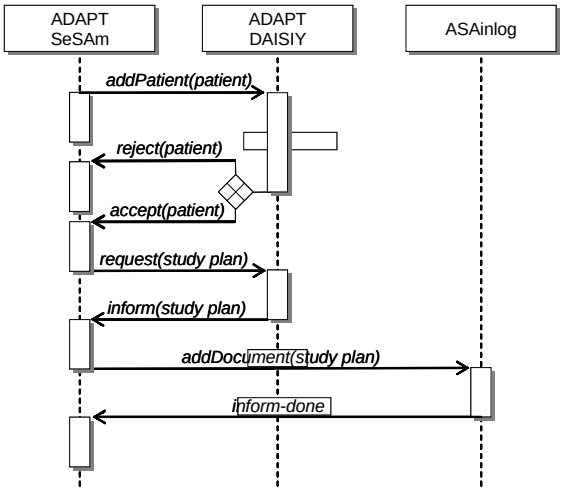
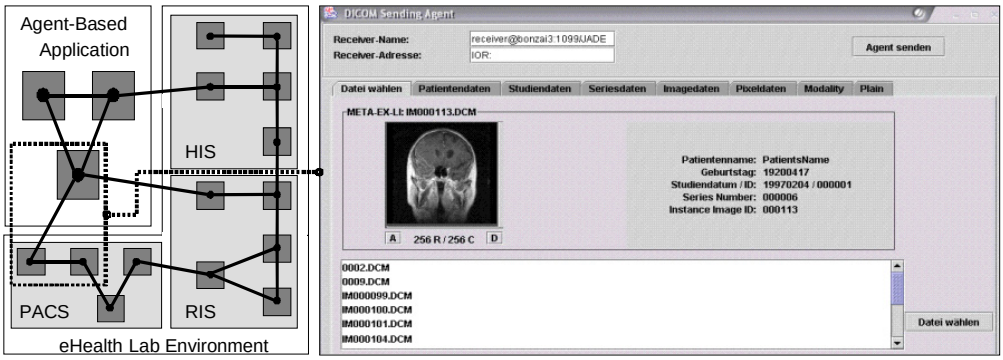


Figure 13: DICOM Message Agent for Interaction with DICOM-compliant PACS, RIS, etc.



systems are DICOM compliant with specific conformance statements). The following interaction diagram (Figure 12) shows a simplified process of the system interactions within Agent.Hospital. This diagram depicts just the essential steps, and the interactions for error handling are not included (e.g., not-understood, refuse, failure, etc.).

The screenshot of the DICOM message agent interface shows the viewing and editing (if allowed) functions (Figure 13). The DICOM image data with all header information can be sent by the agent to the next responsible agent or archiving system.

All these interactions have first been tested in a simulated environment provided by SeSAm and in the second step, implemented and tested in connection with real HISs in the secure laboratory environment called eHealth Lab[5].

SUMMARY & OUTLOOK

Agent technology is able to handle the high level of environmental dynamics as well as complex and restrictive requirements of clinical processes. The aim of ADAPT is to optimize the planning and management of clinical trials as well as provide decision support in this context. In order to accomplish this, realistic simulation results are needed. In this research, we have employed a multiagent-based simulation because conventional simulation has been shown to be insufficient (Klügl, Oechslein, Puppe, & Dornhaus, 2002). As far as multiagent modeling is concerned, we have improved existing modeling techniques to handle complex models with intelligent actors. Another unique aspect of our approach to materializing realistic simulation scenarios is the integration of partial models within the Agent.Hospital initiative. FIPA compliance and subsequent use of ontologies facilitate this interoperation and also enables the development of agent systems, which can be deployed from the simulation environment to an existing information system infrastructure.

The prototype system development has been completed. There are already demo systems, which can explicate the possibilities of our approach. While the initial feedback is very encouraging, we still have to demonstrate the practical use of our system. Our immediate goal is to bring the agent systems to a controlled real-life experiment and show that the results of simulation and the system can be practically used in the management and control of clinical trials. We are at the point where we have gotten the initial results from the simulation, and now we can evaluate the experimental agent systems and deploy them to existing information system infrastructures in the near future. Both applications (simulation and management) will be evaluated with respect to specific advantages of agent-based software over traditional systems.

REFERENCES

Becker, M., Heine, C., Herrler, R., & Krempels, K.-H. (2003). OntHoS: An ontology for hospital scenarios. In A. Moreno & J. Nealon (Eds.), *Applications of software agent technology in the health care domain* [Whitestein

Series in Software Agent Technologies (WSSAT)] (pp. 87-103). Birkhäuser Publisher Basel.

Heine, C., Herrler, R., Petsch, M., & Anhalt, C. (2003). *ADAPT: Adaptive Multi-agent Process Planning & Coordination of Clinical Trials*. AMCIS 2003 Proceedings, Tampa, FL.

Herrler, R., Heine, C., & Klügl, F. (2002) *Appointment Scheduling Among Agents: A Case Study in Designing Suitable Interaction Protocols*. AMCIS 2002 Proceedings, S. 1456-1463, Dallas.

Kirn, S., Heine, C., Herrler, R., & Krempels, K.H. (2003). Agent.Hospital: Agent-based open framework for clinical applications. In G. Kotsis & S. Reddy (Eds.), *Twelfth IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET ICE 2003 Post-Proceedings)*. Los Alamitos, CA: CS Press.

Klügl, F. (2001). *Multagenten-simulation: Konzepte, Werkzeuge, Anwendungen* [Multi-agent simulation: Concepts, tools, application]. Addison Wesley.

Klügl, F., Herrler, R., & Oechslein, C. (2003). From simulated to real environments: How to use SeSAm for software development. In M. Schillo et al. (Eds.), *Multiagent System Technologies: First German Conference MATES'03*, (LNAI 2831).

Klügl, F., Oechslein, C., Puppe, F., & Dornhaus, A. (2002). Multi-agent modeling in comparison to standard modeling. In F. J. Barros & N. Giambiasi (Eds.), *AIS2002 (Artificial intelligence, simulation and planning in high autonomy systems)*. SCS Publishing House.

Klügl, F., Puppe, F., Raub, U., & Tautz, J. (1998). Simulating multiple emergent behaviors: Exemplified in an ant colony. In C. Adami, R. Belew, H. Kitano, & C. Taylor (Eds.), *Proceedings of Artificial Life VI*, Los Angeles, June 26-29. MIT Press.

Krempels, K.-H., Nimis, J., Braubach, L., Pokahr, A., Herrler, R., & Scholz, T. (2003). Entwicklung Intelligenter Multiagentensysteme: Werkzeugunterstützung, Lösungen und Offene Fragen [Development of intelligent multi-agent systems: Tool-support, solutions and open questions]. In K. Dittrich, W. König, A. Oberweis, K. Rannenberg, & W. Wahlster (Eds.), *INFORMATIK 2003 Innovative Informatikanwendungen* (Vol. 1, p. 34)*: GI-Edition Lecture Notes in Informatics*. Köllen Publisher, Bonn.

Oechslein, C., Klügl, F., Herrler, R., & Puppe, F. (2002). UML for behavior-oriented multi-agent simulations. In B. Dunin-Keplicz & E. Nawarecki (Eds.), *From theory to practice in multi-agent systems LNAI: Second international workshop of central and eastern Europe on multi-agent systems, CEEMAS 2001, Cracow, Poland, September 26-29, 2001, Revised papers*. Heidelberg: Springer.

Willmott, S., Constantinescu, I., Dale, J., Somacher, M., Marinheiro, R., Mota, L., et al. (2003). *Agentcities network architecture recommendation*. Re-

trieved January 27, 2003: http://www.agent cities.org/rec/00001/actf-rec-00001a.pdf

ENDNOTES

* An earlier version of the manuscript received the Best Paper Award for the "Intelligent Agent and Multi-Agent Systems in Business" minitrack at the Americas Conference on Information Systems (AMCIS 2003), Tampa, FL, August 4-6, 2003.

1 http://www.dimdi.de

2 http://www.realagents.org

3 http://www.hl7.org

4 http://medical.nema.org/

5 http://www.ehealthlab.de/index_e.html

Dipl.-Wirtsch.-Inf. Christian Heine obtained a master's degree in information systems at the Technical University of Ilmenau and is a research assistant at the University of Hohenheim (Stuttgart) with the chair for Information Systems II. He successfully headed the PaGS-project (2000-2001) on partial global scheduling of processes and the ADAPT-project (2002-2003) on adaptive multi-agent process planning & coordination. Currently he heads the ADAPT II project (2004-2006) on adaptive coordination of clinical processes with cooperating software agents. He is a member of the Agentcities Healthcare Working Group and specializes in requirements engineering and business process modeling methodologies, especially in the healthcare sector.

Dipl.-Inf. Rainer Herrler obtained a master's degree in computer science at the University of Würzburg (2000). From 2000 until the present, he has been working as a research assistant with the chair of artificial intelligence at the University of Würzburg. His main research areas are distributed scheduling and agent-based simulation. He was responsible for several joint research projects in cooperation with the University of Hohenheim.

Professor Dr. Stefan Kirn obtained master's degrees in management science from the University of Federal Armed Forces Munich Germany (1980) and in computer science from the University of Hagen (1989). He received his PhD from the University of Hagen (1991, distributed AI) and his habilitation at Münster University (1995, organizational theories for multi-agent systems). Since 1995, he has held the chair for Information Systems II, Faculty of Economics and Management, Technical University Ilmenau, Germany. Kirn is the initiator and coordinator of a six-year German Research Foundation priority program on business applications of agent technology, which was started in 2000. In this research program, an agent technology test bed called Agent.Hospital has been created, which was presented to an international audience at Barcelona in February 2003. Since July 2003, he has held the chair for Information Systems II at Hohenheim University (Stuttgart).

PPDAM:
Privacy-Preserving Distributed Association-Rule-Mining Algorithm

Mafruz Zaman Ashrafi, Monash University, Australia
David Taniar, Monash University, Australia
Kate Smith, Monash University, Australia

ABSTRACT

Data mining is a process that analyzes voluminous digital data in order to discover hidden but useful patterns from digital data. However, the discovering of such hidden patterns has statistical meaning and may often disclose some sensitive information. As a result, privacy becomes one of the prime concerns in the data-mining research community. Since distributed association mining discovers association rules by combining local models from various distributed sites, breaching data privacy happens more often than it does in centralized environments. In this work, we present a methodology that generates association rules without revealing confidential inputs such as statistical properties of individual sites, and yet retains a high level of accuracy in the resultant rules. One of the important outcomes of the proposed technique is that it reduces the overall communication costs. Performance evaluation of our proposed method shows that it reduces the communication cost significantly when we compare it with other well-known, distributed association-rule-mining algorithms. Nevertheless, the global rule model generated by the proposed method is based on the exact global support of each item set and hence diminishes inconsistency, which indeed occurs when global models are generated from partial support count of an item set.

Keywords: association rules; data mining; distribute systems; distributed association mining; privacy; randomization; secure multiparty

INTRODUCTION

Modern organizations are distributed in various geographical locations. Various business applications used by such organizations normally store their day-to-day data in their corresponding sites. Discovering useful patterns from such organizations using a centralized data-mining approach is not always feasible due to a large network communication cost, which is imposed when merging all datasets into a

central location (Agrawal & Shafer, 1996). As a result, distributed data mining (DDM) emerges as a new subarea of research in the data-mining domain (Zaki, 1999).

Data-mining algorithms analyze enormous digital data and discover hidden patterns within the dataset, and hence impose a threat that such a discovery may breach the privacy of data. As a result, preserving privacy appears to be a prime concern in the field of data mining. Distributed data-mining algorithms need to share local data and discover patterns beyond the organization boundary. As a result, it threatens the privacy of individual sites more than that of the centralized approach of data mining.

One of the most important fields in DDM is association-rule mining (Agrawal & Shafer, 1996). It has attracted significant attention from numerous research communities. Many interesting and efficient distributed and parallel association-rule-mining algorithms have been proposed in the data-mining literature (Agrawal & Shafer; Cheung, Ng, Fu, & Fu, 1996). The distributed association-rule-mining algorithms share support of each item (i.e., an item that has user-specified support) with other sites. In other words, without sharing the support of each item, distributed association-rule-mining algorithms are unable to produce global association rules.

The support of each item has a significant statistical meaning, and hence exposure of those supports may disclose corporate information or their transaction details. Indeed, such disclosure may include identifiable information or corporate plans, which may threaten corporate business gain. Therefore, the privacy of each site's dataset becomes an important issue. Since the distributed association-rule-mining algorithms share only the support of each item to all other participating sites, the key issue is how to generate global rules without disclosing the exact support of each item.

Most of the existing distributed association-rule-mining algorithms overlook the privacy issues. However, we believe that privacy should be a main concern of distributed association mining; otherwise, the resultant patterns may reveal sensitive information, and hence, participating sites may lose their businesses. To illustrate the above discussion more clearly, consider the following example.

Suppose two supermarkets such as Mart-1 and Mart-2 share their data and would like to find global association rules for mutual benefits. Consider during the mining process that Mart-1 comes to know that 70% of customers of Mart-2 purchase items A and B at the same transaction (i.e., it generates that rule from the exact support of itemsets A and B of Mart-2). Using this information, Mart-1 can offer better deals (e.g., the price of B will be 25% less if the customer purchases items A and B together), and subsequently, Mart-2 may lose their business. The above example shows how exact support of each itemset is important in a distributed association-rule-mining context. Therefore, hiding each item support from other sites is vital to preserve the privacy of each participating site.

Privacy is a key issue of association-rule-mining and becomes an active area of research. We can categorize the existing work into two categories: (a) the distortion or randomization approach (Atallah, Bertino, Elmagarmid, Ibrahim, & Verykios, 1999; Evfimievski, Srikant, Agrawal, & Gehrke, 2002; Rizvi & Haritsa, 2002), and (b) the secure multiparty computation approach (Kantercioglu & Clifton, 2002; Vaidya & Clifton, 2002).

- **Randomization Approach.** The randomization approach intends to discover association rules from randomized datasets of various sites. It focuses on the privacy of the individual site, and it does not reveal original records of one site to other participating sites. To preserve privacy, transactions are randomized by discarding some items and inserting new items into it (Rizvi & Haritsa, 2002). The statistical estimation of original supports and variances given randomized supports allows a central site to adopt the Apriori algorithm (Agrawal & Srikant, 1994) to mine frequent itemsets in nonrandomized transactions by looking at only randomized ones.
- **Secure Multiparty Computation.** The goal of secure multiparty computation is to build a data-mining model from local datasets of various participating sites without revealing individual records of one site to other participating sites. To achieve this, it computes a function $f(x, y)$, where, in this case, two parties hold their inputs x and y, and at the end, all parties know the result of the function $f(x, y)$ and nothing more.

From the above discussion, one may think that we can generate global association rules without exposing the exact support of each item if they apply those approaches. However, such attempts often diminish the objective of distributed association mining because these approaches either increase overall communication cost or may cause discrepancies in the final model. For example, the randomized approach combines all randomized datasets of participating sites into a centralized site, and then adopts the Apriori algorithm (Agrawal & Srikant, 1994) or any other association-rule-mining algorithm to find frequent itemsets. However, if various sites use different randomized functions to distort their dataset, then discrepancies may exist in the resultant model. In addition, sending all randomized datasets to a central site involves large communication costs. Using the second approach, each site sends randomized support counts multiple times to other participating sites to find the exact global support of each item. Due to this, it also incurs large network communication and, consequently, diminishes the objective of distributed association-rule mining in the first place.

The above discussions clearly illustrates that distributed association rule mining algorithms should have efficient methods to hide exact support of each item without increasing the overall communication cost. To achieve the goal, we present a **p**rivacy-**p**reserving, **d**istributed **a**ssociation-**r**ule-**m**ining (PPDAM)

method that generates global association rules without revealing confidential inputs such as statistical properties (i.e., support of each item) of individual sites. The important outcomes of the proposed technique is that it neither discloses the exact support of each item during the mining process or increases the overall communication cost. In addition, it also has the ability to minimize a collusion problem, which occurs when two sites on the chain collude to find the exact support of another site. Nevertheless, it diminishes the reconstruction problem, which is raised when we distort transactions of a dataset by using different randomization techniques.

The rest of the paper is organized as follows. We describe the background of distributed association-rule mining and problems we encountered when considering PPDAM. Related work of PPDAM is then described. Next we present our proposed privacy-preserving, distributed association-rules-generation methodology. A performance evaluation and comparison are then presented, and finally we conclude.

DISTRIBUTED DATA MINING: BACKGROUND

DDM intends to discover rules from different datasets that are distributed across multiple sites and interconnected using a communication network. As combining those datasets in a single site requires massive network communication, DDM offers a new technique to discover knowledge or patterns from loosely coupled distributed datasets. Finally, DDM produces global rules using minimal network communication. Figure 1 illustrates a typical distributed data-mining framework. It shows three participating sites, where each site generates local models from their respective data repository and exchanges local models with other sites to generate global models.

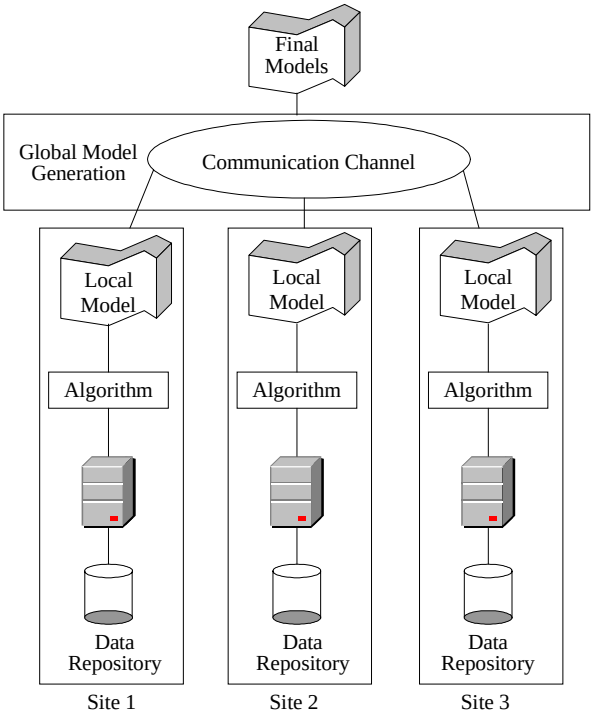
In addition, DDM produces global models in such a way that one would achieve the same model if the datasets from different sites were combined into a single site. It treats all distributed datasets as a single virtual table (Wirth, Borth, & Hipp, 2001). In general, every distributed mining algorithm performs the following three tasks:

- analyzes local data and generates local models,
- exchanges local models with other sites, and
- generates global models by combining all local models.

Distributed Association-Rule Mining

Distributed association-rule mining is a subarea in DDM (Ashrafi, David, & Kate, 2004). Typically, rules generated by distributed association-rule-mining algorithms are considered interesting if they satisfy both minimum global support and confidence thresholds. In order to find interesting global rules, distributed association-rule-mining algorithms generally have two distinct tasks: (a) *global support count*, and (b) *global rules generation*. The global support count of an itemset is an aggregate of all local support counts of

Figure 1: Distributed Data-mining Framework



that itemset, and can be defined as follows.

$$Support(A) = \frac{A \subseteq t_i}{N} \ldots. \qquad \textbf{(1)}$$

Definition. Let D be a virtual transaction dataset comprising $D_1, D_2, D_3, \ldots, D_m$, which are geographically distributed datasets. Let n be the number of items and I be the set of items such that $I = \{a_1, a_2, a_3, \ldots, a_n\}$. Suppose N is the total number of transactions and $T = \{t_1, t_2, t_3, \ldots, t_N\}$ is the sequence of transactions such that $t_i \subset D$. The support of each element of I is the number of transactions in D containing I for a given itemset $A \subset I$; we can define its support as follows:

Itemset A is frequent if and only if $Support(A) \geq minsup$, where *minsup* is a user-defined support threshold. Once the algorithm discovers all global, frequent itemsets, each site generates global rules that have user-specified confidence.

The main objective behind distributed association-rule-mining is to reduce communication costs in such a way that the overall cost will be less than the cost if we combined the datasets of all participating sites into a centralized site. For example, suppose sites $S_1, S_2, S_3, \ldots, S_n$ are involved in the mining task, and each of

the sites has its own dataset D_1, D_2, D_3, …, D_n. Let C_1, C_2, C_3, …, C_m be the communication costs incurred after every iteration. Then, we can calculate the total communication cost for generating global, frequent itemsets: $G_c = \sum_{i=1}^{m} C_i$, where m is the total number of iterations. Suppose the total communication cost of combining all n participating sites' datasets into a centralized site is $D_c = \sum_{i=1}^{n} D_i$.

Then, $G_c < D_c$ will be the primary and desirable property of any distributed association-rule-mining algorithm.

Privacy

Distributed association-rule-mining algorithms should discover association rules beyond the organization boundary (Ashrafi, Taniar, & Smith, 2004). They form the final rule model by combining various local patterns. For example, suppose there are three sites, S_1, S_2, and S_3, and each of them have datasets D_{S1}, D_{S2}, and D_{S3}. Suppose A and B are two items having a global support threshold. In order to find rule A → B or B → A, we need to aggregate the local support of itemsets AB from all participating sites (i.e., sites S_1, S_2, and S_3). When we are doing such aggregation, all sites learn the exact support count of other sites. However, in many situations, participating sites are reluctant to disclose the exact support of itemset AB to other sites, as support counts of an itemset have statistical meanings and are a threat to privacy.

For the above-mentioned reason, we need to secure multiparty computational solutions to maintain privacy in distributed association-rule mining (Vaidya & Clifton, 2002). The goal of secure multiparty computation in distributed association-rule mining is to find the global support of all itemsets using a function where multiple parties hold their local support counts, and at the end, all parties know the global support of all itemsets and nothing more. Also, each participating site uses that global support for rules generation.

Problem Definition

Let $X = \{x_1, x_2, x_3, \ldots, x_n\}$, $Y = \{y_1, y_2, y_3, \ldots, y_n\}$, and $Z = \{z_1, z_2, z_3, \ldots, z_n\}$ be the support counts of candidate k itemsets geographically distributed over three sites such as S_1, S_2, and S_3. In order to generate global frequent k itemsets $F = \lambda (X + Y + Z)$, each site needs to send their respective support count of each candidate itemset l to the other sites. Although broadcasting of support counts does not disclose any information about individual transactions, it may disclose some valuable information about each site, such as data size, exact support of each itemset, and so forth, which may subsequently breach the privacy of each site. Hence, the challenge is to find global frequent itemset F without revealing l (i.e., support) of each site.

In addition, in a distributed association-rule-mining context, message exchange between different sites is considered one of the main tasks. This task becomes more expensive when each site broadcasts large numbers of support

counts. Indeed, the communication cost of all distributed association-rule-mining algorithms will increase when the privacy of each site is considered the primary objective. Hence, the problem can be defined as finding all association rules from various distributed sites without revealing the support counts of individual sites and without increasing the overall communication costs.

RELATED WORK

Several numbers of frameworks have been proposed for maintaining the privacy of association rules (Atallah et al., 1999; Dasseni, Verykios, Elmagarmid, & Bertino, 2001; Evfimievski et al., 2002; Kantercioglu & Clifton, 2002; Rizvi & Haritsa, 2002; Vaidya & Clifton, 2002). However, most of them dealt with sequential or centralized association mining. Mining association with secrecy konstraints (MASK; Rizvi & Haritsa) was proposed for a centralized environment to maintain privacy and accuracy of resultant rules. This approach was based on simple, probabilistic distortion of user data, employing random numbers generated from a predefined distributed function. However, the distortion process employs system resources for a long period when the dataset has a large number of transactions. Furthermore, if we use this algorithm in the context of a distributed environment, we need uniform distortion among various sites in order to generate unambiguous rules. This uniform distortion may disclose confidential inputs of an individual site and may also breach the privacy of data (i.e., exact support of itemsets), and hence it is not suitable for distributed data mining.

Evfimievski and his colleagues (2002) presented a randomization technique in order to preserve privacy of association rules. The authors analyzed this technique in an environment where a number of clients are connected to a server. Each client sends a set of items to the server where association rules are generated. During the sending process, the client modifies the itemset according to its own randomization policy, and as a result, the server is unable to find exact information about the client. However, this assumption is not suitable for distributed association-rule mining because it generates frequent itemsets by aggregating support counts of all clients (i.e., sites). If the randomization policy of each site differs from others, we will not be able to generate the exact support of an itemset. Subsequently, the resultant global frequent itemsets will be erroneous. Hence, we may not be able to discover useful rules. Furthermore, this technique individually disguises each attribute, and data quality will degrade significantly when the number of attributes in a dataset is large.

A new technique to preserve the privacy of sensitive knowledge by hiding frequent itemsets from large datasets was presented in Atallah et al. (1999). The authors apply heuristics to reduce the number of occurrences to such a degree that its support is below the user-specified support threshold. Dasseni et al. (2001) extended this work and investigated confidentiality issues of association-rule mining. Both works assume datasets are local, and hiding some itemset will not af-

fect the overall performance or mining accuracy. However, in distributed association-rule mining, each site has its own dataset and a similar kind of assumption may cause ambiguities in the resultant global rule model.

Vaidya and Clifton (2002) presented a technique to maintain the privacy of association rules in vertically partitioned, distributed data sources (across two data sources only) where each data site holds some attributes of each transaction. However, if the number of disjoint attributes among the site is high, this technique incurs huge communication costs. Furthermore, this technique worked only for two sites, hence it is not scalable.

Recently, Kantercioglu and Clifton (2002) proposed a privacy-preserving association-rule mining for horizontally partitioned data (i.e., each site shares a common schema but has different records). The authors propose two different protocols: securing the union of locally large itemsets and testing the support threshold without revealing support counts. The former protocol uses cryptography to encrypt local support counts and, therefore, it is not possible to find which itemset belongs to which site. However, it reveals the number of itemsets having a common support. The latter protocol adds a random number to each support count and finds excess supports. Finally, these excess supports are sent to the second site where it learns nothing about the first site's actual dataset size or support. The second site adds its excess support and sends the value until it reaches the last site. However, this protocol can raise a collusion problem. For example, site i and $i + 2$ in the chain can collude to find the exact excess support of site $i + 1$. Furthermore, this protocol only discovers an itemset, which is globally large, not the exact support of an itemset, and each site generates rules based on the local support counts.

In this paper, we propose an efficient technique that maintains the privacy of distributed association-rule mining according to a secure multiparty-computation definition (Goldriech, 2001). The proposed technique finds the exact support of each global frequent itemset without revealing the candidate support counts of individual sites. The proposed technique has the ability to minimize the collusion problem without increasing overall communication costs. Furthermore, it eliminates the reconstruction problem, which is raised when we distort transactions by using different randomization techniques.

PROPOSED METHOD

In this section, we propose a methodology that maintains the privacy of distributed association-rule mining. First we explain the rationale of why each site in distributed association rule mining shares its support count of all itemset with all other sites. In the context of distributed association mining, each participating site needs to know whether an itemset is globally frequent or not in order to generate candidate itemsets for the next pass. Without that piece of information, distributed association-rule-mining algorithms will not be able to generate global candidate itemsets. Second, if any site generates rules based on a partial support count of

an itemset, inconsistency problems (i.e., confidence of rules at different sites may vary) will arise. Before embarking on the details of our proposed method, let us discuss the basic notations and assumptions of this framework.

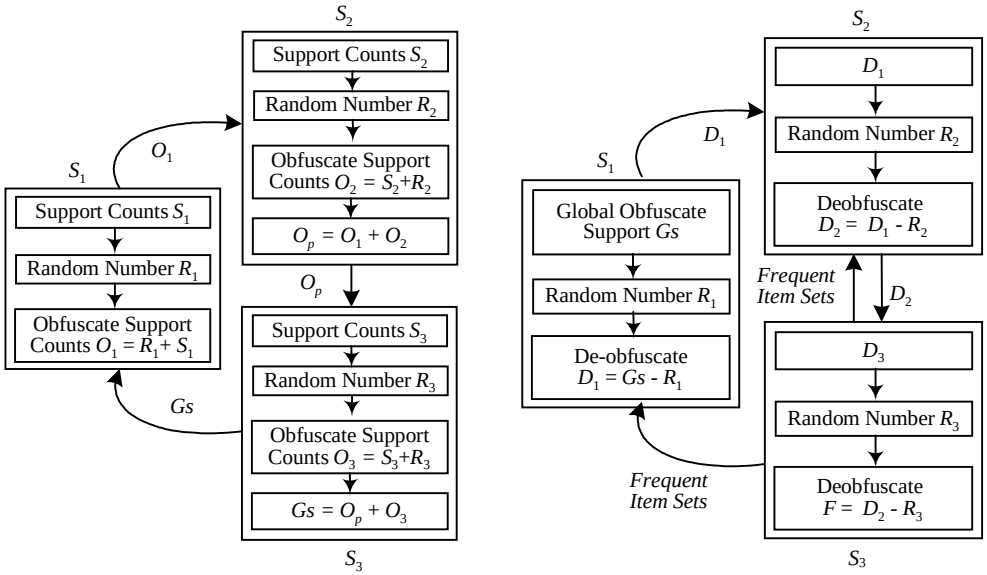
Assumptions

- **Dataset Model.** We assume there are N numbers of participating sites and each of them has its own dataset. Each transaction of the datasets has a set of items, and each item is represented by a number. The number of items in a transaction may vary. The dataset of each site may be heterogeneous; however, the items of each transaction at different sites have the same taxonomy level (i.e., hierarchy of items). Furthermore, each dataset is not required to have the same number of transactions. Only the aggregate support count will allow us to identify whether an itemset has a global support threshold or not.
- **Number of Sites.** The aim of this work is to find the exact global support of all itemsets without revealing the exact support counts of each participating site. However, when the number of participating sites is equal to two, it becomes very easy for both sites to find out the exact support counts of the other sites, no matter what kind of secure computation we enforce (Vaidya & Clifton, 2002). Hence, we assume the number of sites participating with this framework is equal to n, where $n > 2$.
  To overcome the above problem, one may think of randomization techniques. However, it is clear that if different sites distort their respective dataset using a nonuniform, randomized function or add and drop some of the items from the transaction, the resultant global rule model will be inconsistent and, subsequently, will diminish the objective of distributed association-rule mining in the first place.
- **Characteristics of Each Site.** Each site participating in distributed association-rule mining possesses a minimum level of trust. Ideally, it is easier for any distributed association-rule-mining algorithm to maintain privacy if all computations are done by a trusted third party. However, this kind of solution is not feasible due to various limitations. Due to this reason, we assume that all participating sites are *semihonest* sites. A semihonest site possesses the following characteristics:
  - follows multiparty computation protocols completely,
  - keeps record of all intermediate computation, and
  - is capable of deriving additional information using those intermediate records.

Methodology

We now discuss how we generate global association rules without revealing the exact support counts of participating sites. Our proposed method is based on the following analogy. For example, suppose we have a large real number $N \subseteq R$, which is a sum of two numbers such as N_1 and N_2, where $N_1 \subseteq R$ and $N_2 \subseteq R$, $N_1 \neq 0$ and $N_2 \neq 0$, and R is a real number. If the value of N is known, and N_1 and N_2

Figure 2: Proposed Methodology



are unknown, the value of N_1 or N_2 remains private and secure until we know the exact value of either N_1 or N_2.

The proposed method uses the aforementioned technique and considers N_1 as an exact support count of an itemset, N_2 as a random number, and N as a disguised support count. It has two distinct phases, namely (a) *obfuscation* and (b) *deobfuscation*. In the obfuscation phase as shown in Figure 2a, each support count of a candidate itemset is obfuscated (i.e., an addition of exact support count and a random number) and is sent to an adjacent site. Each adjacent site then aggregates its obfuscated support count with the receiving support count, and sends that aggregation to the next site. This sending process continues until it reaches the last site. When it reaches the last site, it aggregates its obfuscated support count with the receiving support count. We call this a global obfuscated support count.

After computing the global obfuscated support count in the last site, the method sends the global obfuscated support counts to the first site for deobfuscation as shown in Figure 2b. The first site starts the deobfuscation phase by subtracting its own generated random number from each of the global obfuscated support counts and sends to the adjacent site. Each site subtracts their respective random number and sends to the next site until it reaches the last site. When the last site subtracts its own generated random number from the global obfuscated support count, we find the exact global support of the itemset. Then, all global support counts are checked in order to prune away each nonfrequent global itemset. Finally, the global frequent support counts

are sent to all other sites and each site generates candidate itemsets for the next iteration. This process continues until there are no more candidate itemsets that can be generated from the previous frequent itemsets. To elaborate this process, let us explain using the following example.

Example: Suppose there are three sites, S_1, S_2, and S_3. After the first iteration, the global candidate 2-itemsets are equal to {*AB*, *AC*, and *BC*}. Let the local support counts of candidates of site S_1 be equal to {5, 3, and 4}, at S_2 be equal to {10, 9, and 1}, and at S_3 be equal to {5, 5, and 1}. Suppose sites S_1, S_2, and S_3 generate random number $R_1 = 100$, $R_2 = 200$, and $R_3 = 200$, and each site obfuscates its own support count of candidate itemsets by adding a corresponding random number with each support count. After performing obfuscation, consider that site S_1 sends its obfuscated support counts set (i.e., 105, 103, and 104) to site S_2. When S_2 receives these obfuscate support counts, it aggregates its own obfuscate support counts, {210, 209, and 201}, with the receiving support counts. Since each site shares the same candidate itemset, this aggregation operation can be done on the fly. Upon performing these tasks, site S_2 sends obfuscate support counts set {315, 312, and 305} to site S_3 that performs the same task as does site S_2 and finishes the obfuscation phase.

In the next phase (i.e., deobfuscation), site S_3 sends the global obfuscated support count set {520, 517, and 506} to site S_1 where it subtracts a random number R_1 from each element of those itemsets. However, this subtraction does not reveal any knowledge to site S_1 as R_1 is subtracted from the global obfuscate support count. Hence, the candidate itemset support count of other sites remains hidden to site S_1. After subtracting R_1, site S_1 sends support counts set {420, 417, and 406} to site S_2 that subtracts the random number R_2 and sends support counts to site S_3. After subtracting random number R_3, site S_3 finds the global supports counts {20, 17, and 6} and discovers global frequent itemsets. Since the global support counts is an aggregation of all local candidate support counts, it is not possible for site S_3 to discover the exact support counts of the other sites (i.e., S_1 and S_2) from that set.

One of the important outcomes of the proposed method is that it minimizes the collusion problem. This is because each site obfuscates candidate support counts with its own random number, and that random number is subtracted from global obfuscated support counts on that particular site in order to perform deobfuscation. Hence, it requires n - 1 sites of the chain to collude to find out the exact support count of any site S_i.

Algorithms

A general distributed association-mining algorithm has three distinct phases: (a) finding out the total number of overall transactions, (b) finding global support counts for various itemsets' lengths, and (c) generating global rules. Our proposed method works in the same manner, but

Figure 3: Obfuscation Procedure

```
Input:  Local support counts C,
        Obfuscate support counts C_R;
Output: Obfuscated candidate support counts C_O
R = generate_random_number();
for each element I of C {
  I_1 = I + R;
  add(I_1)
}
if(C_R == {})
  Send(C_O);
else{
  aggregate (C_O,C_R);
  send(C_O);
}
```

Figure 4: Deobfuscation Procedure

```
Input:  Obfuscate support counts C_OR, Random number R
Output: Partial candidate support counts C_p or
        Global frequent itemset Fg
for each element I of C_OR{
   I_1 = I - R;
   add(I_1)
}
if(current site is not end site)
   Send(C_p);
else{
   F_G = generate_global_frequent_itemset(C_p);
   broadcast(F_G);
}
```

obfuscates and deofuscates at every step. In the beginning, all sites generate local frequent one-itemsets based on the simple distributed mining association (DMA) algorithm (Cheung et al., 1996). After generating frequent 1-itemsets, each site generates a random number, adds to the support count, and sends it to the adjacent site (refer to Figure 3). This procedure continues for all different lengths of candidate itemsets and finishes when there are no more global candidate itemsets.

The pseudocode of the deobfuscation procedure is shown in Figure 4. After receiving the obfuscated support count, each site subtracts it by its own random number and sends the result to the adjacent site. At the end of the deobfuscation (i.e., at the last site), it will discover the exact global support of all itemsets and generates the global frequent itemsets from that set. Finally, these global frequent support counts are broadcast to all other sites.

Message Optimization

In distributed association-rule mining, exchanging messages among different sites is considered one of the main

tasks. Thus, message optimization becomes an integral part of it. However, our method as described in the previous section is not able to reduce the message-exchange size, as it accomplishes the global support counts in two rounds and in each round it exchanges messages. To reduce the message-exchange size, we propose a further modification of our method. We only follow the aforementioned method to discover a frequent itemset of length 1. After that, each site uses a function $f(x)$ to obfuscate the support counts by utilizing local support counts of those global frequent itemsets rather than a random number. We use the local support counts of global frequent candidate itemsets, since other sites knew the obfuscated support count of an itemset, but not the exact value of it. The function $f(x)$ can be calculated by using the following formula:

$$f(x) = O_S \pm C_S \ \ldots, \qquad \textbf{(2)}$$

where O_S is the exact support count of a k-itemset and C_S is the sum of n local support of a k - 1-itemset. After generating the obfuscated candidate support counts, each site sends the support counts of candidate itemsets to a single site, where the global frequent itemsets for that iteration will be obtained. We refer to the sites that send obfuscated support counts as *senders* and the sites that generate the global frequent itemsets as *receivers*. For example, if there are three sites participating in the process, two of them will broadcast the obfuscate support counts to the third site. Once the receiver site receives an obfuscated local support from different sender sites, it aggregates them using the following formula:

$$G_S = \sum_{i=1}^{n} f(x) \pm F_S \ \ldots \qquad \textbf{(3)}$$

where G_S is the global support count of a k-itemset and F_S is the sum of n global support of a k -1-itemset. The sum of local support of an itemset is equal to the global support of that itemset. So, one can easily prove that $F_S = \sum_{i=1}^{n} C_S$. As a result, when we add or subtract F_S from $\sum_{i=1}^{n} f(x)$, it gives us the exact global support of an itemset.

To illustrate the above-mentioned procedure, let us consider the example shown in Table 1. After the first iteration, it discovers {*A*, *B*, *C*} as the global frequent one-itemsets. Then each site generates candidate itemsets and their support counts. In order to obfuscate the support count of each itemset, they use Formula 1 (for each candidate k-itemset, C_S is the addition of local support counts of all k - 1-itemsets, e.g., local support of *A* and *B* for itemset *AB*) and send those obfuscated support counts to the receiver sites. Since the receiver site receives only the value of $f(x)$, this does not tell the exact support count of an itemset. It is only able to discover the global support of each candidate itemset by using Formula 3. As the result, we will be able to eliminate the deobfuscation phase of our proposed method and reduce the message-exchange size.

Table 1: Example

Global Frequent k- Item Set		Local Support at Different Sites			Candidate k+1- Item Set		$f(x)$		Global k+1- Item Set		
Name	F_S				Name	O_S	C_S	$O_S + C_S$	Name	$\Sigma f(x)$	G_S
A	150	S_1	*A*	70	*AB*	10	120	130	*AB*	335	60
			B	50	*AC*	25	105	130			
			C	35	*BC*	30	85	115			
B	125	S_2	*A*	40	*AB*	25	85	110	*AC*	370	90
			B	45	*AC*	40	75	115			
			C	35	*BC*	20	80	100			
C	130	S_3	*A*	40	*AB*	25	70	95	*BC*	315	60
			B	30	*AC*	25	100	125			
			C	60	*BC*	10	90	100			

Complexity Analysis

After the first iteration, the proposed algorithm sends support counts of candidate itemsets to a single site, where the global frequent itemsets that pass will be calculated. The sites that are sending local support counts are referred to as the senders and the site that generates the global frequent itemsets is the receiver. Since all sender sites obfuscate the support of each itemset using Formula 2 and sends those obfuscated supports to the receiver site, the receiver site is unable to generate the exact support of any itemset and can only generate the global support of all itemsets. For example, if there are three sites, two of them broadcast their local support counts of candidate itemsets to the third site. The third site is responsible for generating global frequent itemsets of that iteration. The total number of message broadcasts from a sender site to a receiver site is equal to $(1^* |C|)$. We can calculate the total message size using Formula 4.

$$T_{\text{messages}} = \sum_{i=1}^{n} (n-1) * C \quad \ldots\textbf{(4)}$$

where n is the total number of sites and C is the number of candidate itemsets.

Once the receiver site generates global frequent itemsets, it broadcasts those itemsets to all sender sites. The total number of message broadcasts from the receiver is equal to $(n - 1 * |F_G|)$. The total message-broadcasting size is the aggregate of sender and receiver sites' messages and can be calculated by using the following formula:

$$T_{\text{messages}} = (n-1) * C + (n-1) * F_G \ldots(5)$$

where n is the number of sites, C is the candidate itemsets and, F_g is the global frequent itemsets.

Example 1. Consider there are three sites, S_1, S_2, and S_3. After the first iteration, suppose the set of large one-itemsets = *{A, B, C}*. Consider in the

Table 2: Data Set Characteristics

Name	Transaction-Size Avg.	Number of Distinct Items	Number of Records
Cover Type	55	120	581012
Connect-4	43	130	67557
BMS-WEB-View-1	2	497	59602
BMS-WEB-View-2	5	3340	75512

next iteration that each site has candidate itemsets equal to {*AB*, *AC*, and *BC*}. If we consider sites S_1 and S_2 as senders and site S_3 as the receiver, then after the second iteration, the support counts of S_1 = {*AB*, *AC*, and *BC*} and S_2 = {*AB*, *AC*, and *BC*}. These will be sent to the receiver site S_3 where the global frequent itemset of this iteration will be generated. Consider the global frequent itemsets of this iteration are equal to {*AB* and *BC*}; site S_3 will then broadcast them to all sender sites. If we calculate the total message size, then we find the PPDAM only broadcasts 10 messages.

PERFORMANCE EVALUATION

We have done an extensive performance study on our proposed message-reduction techniques to confirm our analysis of its effectiveness. A client-server-based distributed environment is established in order to evaluate the message-optimization technique. Initial evaluation was carried out on four different sites. Each site has a receiving and a sending unit and listens to a specific port in order to send and receive the support counts.

We have also implemented a sequential association-mining algorithm using Java 1.4 and replicated the algorithm to four different sites to generate candidate support counts of each site. Each site generates a random number using a pseudorandom-number generator to obfuscate support counts.

Four real datasets are chosen for this evaluation study. Table 2 shows the characteristics of the datasets that are used in this evaluation. It shows the number of items, the average size of each transaction, and the number of transactions of each dataset. The Cover Type and Connect-4 datasets are taken from the University of California, Irvine, machine learning dataset repository (Blake & Merz, 1998), whereas BMS-Webview-1 and BMS-Webview-2 are real-world datasets containing several months worth of clickstream data from an e-commerce Web site and are made publicly available by Blue Martini Software (Kohavi, Broadley, Frasca, Mason, & Zheng, 2000).

To deploy he distributed association-mining algorithm, we divide all datasets into four different partitions and assign to them four different sites. In order to reduce identical transactions among different sites, each of these partitioned

Table 3: Accuracy

Data Set	Supports	No. of Sites	Total No. of Global Frequent Item Sets		Accuracy
			Traditional Approaches	PPDAM	
Cover Type	0.2%	4	166171	166171	100%
	0.4%	4	216018	216018	100%
	0.6%	4	305485	305485	100%
Connect-4	75%	4	1612127	1612127	100%
	80%	4	541591	541591	100%
	85%	4	144751	144751	100%
BMS-Webview-1	0.068%	4	752076	752076	100%
	0.074%	4	61955	61955	100%
	0.08%	4	20800	20800	100%
BMS-Webview-2	0.06%	4	86391	86391	100%
	0.067%	4	67072	67072	100%
	0.075%	4	52678	52678	100%

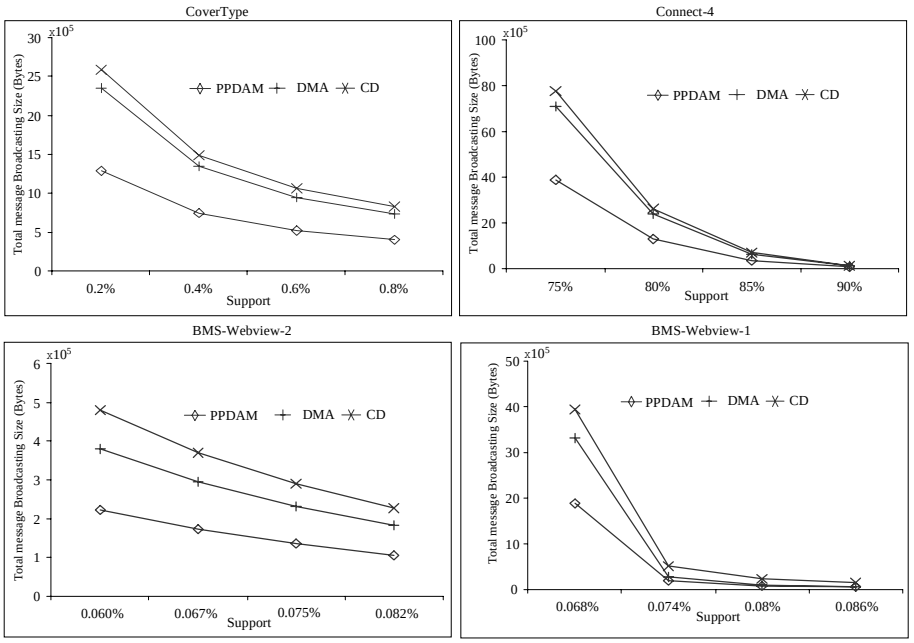
datasets was generated in such a way that each partition has 75% of the transactions of the original dataset.

Accuracy is one of the main objectives of data mining. Our proposed PPDAM method obfuscates and deobfuscates the candidate itemset support count during the mining process. Therefore, such processes create new requirements to show whether the PPDAM can generate the same number of global frequent itemsets when we generate the global frequent itemsets in the traditional distributed association-rule-mining approach. Table 3 shows the accuracy of our proposed approach in global frequent itemsets generation. From the table, it is clear that our proposed method generates the same number of global frequent itemsets as the traditional approaches. Thus, the accuracy of our method is the same as that of the traditional approach.

It is important to discuss the rationale behind the next experiments before we proceed. One of the key performance improvements sought in various distributed association-rule-mining algorithms (Agrawal & Shafer, 1996; Cheung et al., 1996) is in message optimization. As these algorithms are based on the sequential Apriori algorithm (Agrawal & Srikant, 1994) or extensions of it, the performance varies due to the message-broadcasting technique. The intention behind distributed association mining is to reduce communication cost. Hence, those improvements are a significant achievement in the context of distributed association mining.

In addition, all existing privacy preserving association rule mining algorithms (Evfimievski et al., 2002; Kantercioglu & Clifton, 2002; Vaidya & Clifton, 2002) are derived from well-known algorithms such as count distribution (CD) and DMA (Cheung et al., 1996). Those derived algorithms work in the same manner except they are involved in extra communication to generate a global rule model. Therefore, the efficiency of our proposed method over the existing algorithms can

Figure 5: Comparative Total Communication Costs



be clearly exposed if we compare it with the original algorithms rather than the derived PPDAM algorithms.

Before we compare the message-broadcasting method of our proposed PPDAM technique with CD and DMA, it is important to know in detail how these algorithms broadcast messages. For example, the CD algorithm generates support counts at each local site and broadcasts them to all other sites. All sites can then find the global frequent itemsets for that pass. Since each participating site generates support counts from the frequent itemset of previous pass, only support counts of those itemsets will be enough to generate global frequent itemsets of that pass.

The DMA algorithm (Cheung et al., 1996) introduces a new optimization technique to reduce message-broadcasting costs. For every local, large itemset, it assigns a pooling site that sends a request to check whether that itemset is globally large or not. When a pooling site receives a request, it sends a pooling request to all other remote sites except the originator site. Upon receiving the pooling request, it computes the heavy itemsets and sends them to all other sites. However, when a site sends a pooling request, it not only sends the support count of that itemset, but also sends the name of that itemset (Cheung et al.). As a result, it increases the broadcasting cost in some cases. It is worth mentioning that both CD and DMA are nonsecure parallel or distributed association-mining algorithms, meaning that none of them maintain the privacy of individual sites' inputs. To ensure privacy, two algo-

rithms that are extensions of DMA were proposed (Vaidya & Clifton, 2002). However these algorithms took *n* rounds in comparison with DMA, which took one single round in order to generate global support count. As a result, they exchange more messages than DMA.

Figure 4 depicts the total size of messages (i.e., number of bytes) transmitted by PPDAM, DMA, and CD in order to generate global frequent itemsets from different datasets. Depending on the characteristics of each dataset, we vary the support-threshold value. To generate a reasonable number of global frequent itemsets, we use a very high-support threshold for dense a dataset and a low-support threshold for a sparse dataset.

The message size was measured by assuming 4 bytes for each support count and 4 bytes for each candidate itemset name. We keep 4 bytes for each candidate itemset name as there are not only large numbers of candidate itemsets, but also each dataset generates long candidate itemsets (i.e., number of items in the itemset is large). Similar assumptions are also made in other distributed association-rule-mining algorithms.

From the above comparison between PPDAM, DMA, and CD as plotted in Figure 5, it is clear that the PPDAM algorithm exchanges fewer messages. Indeed, in all cases, PPDAM reduces communication cost by 60% to 80% compared to CD. Using the CD algorithm, each site exchanges messages with all other sites after every pass, and hence the message-exchange size increases when we increase the number of sites. In contrast, PPDAM transmits 25% to 50% less messages compared with DMA. The DMA algorithm exchanges more messages because the polling site sends and receives support counts from remote sites. Further, it sends support counts of the global frequent itemset to all sites. Consequently, it increases the message size. On the contrary, using our proposed method, each site sends its support counts to a single site and receives the global frequent support count from a single site. Hence, it reduces the number of broadcasting operations. Furthermore, the local pruning technique of DMA effectively works only when different sites have been vertically fragmented, sparse datasets. However, this will not be able to prune a significant number of candidate itemsets when each site uses horizontally fragmented datasets.

CONCLUSION & FUTURE WORK

Privacy-preserving mining of association rules has become one of the active research issues in recent years. However, maintaining privacy in distributed association-rule mining is more difficult than in the centralized approach. In this paper, we propose a methodology — PPDAM that efficiently generates distributed association rules without revealing the support counts of each site. The proposed method generates rules based on the exact global support of an itemset. The resultant rule model achieved by this method is the same as if one generates it using some of the well-known distributed or parallel algorithms. The proposed method does not distort the original dataset and for this reason, it does not require any further com-

putational costs. Nevertheless, the performance evaluation shows that the overall communication cost incurred by our proposed method is less than those of the CD and DMA algorithms.

Distributed association-rule mining has the ability to find out rules beyond the organization boundaries. In the future, we plan to investigate how the association-mining algorithm can be employed on different datasets that do not share any common feature. We also plan to investigate how we can maintain privacy in such a scenario.

REFERENCES

Agrawal, R., & Shafer, J. C. (1996). Parallel mining of association rules. *IEEE Transactions on Knowledge and Data Engineering, 8*(6), 962-969.

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large database. *Proceedings of the 20th International Conference on Very Large Databases*, 407-419.

Ashrafi, M.Z., Taniar, D., & Smith, K.A. (2004a). ODAM: An optimized distributed association rule mining algorithm. *IEEE Distributed Systems Online*.

Ashrafi, M.Z., Taniar, D., & Smith, K.A. (2004b). Reducing communication cost in privacy preserving distributed association rule mining. *Proceedings of Database Systems for Advanced Applications*.

Atallah, M., Bertino, E., Elmagarmid, A., Ibrahim, M., & Verykios, V. (1999). Disclosure limitation of sensitive rules. *Proceedings of the IEEE Knowledge and Data Engineering Exchange Workshop KDEX*.

Blake, C. L., & Merz, C. J. (1998). *UCI repository of machine learning databases*. University of California, Irvine, Department of Information and Computer Science. Retrieved: www.ics.uci.edu/~mlearn/MLRepository.html

Cheung, D. W., Ng, V. T., Fu, A. W., & Fu, Y. (1996). Efficient mining of association rules in distributed databases. *IEEE Transactions on Knowledge and Data Engineering, 8*(6), 911-922.

Dasseni, E., Verykios, V. S., Elmagarmid, A. K., & Bertino, E. (2001). Hiding association rules by using confidence and support. *Proceedings of the International Information Hiding Workshop (IHW)*.

Evfimievski, A., Srikant, R., Agrawal, R., & Gehrke, J. (2002). Privacy preserving mining association rules. *Proceedings of the SIGKDDD*, 217-228.

Fayyed, U., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advance in Knowledge Discovery and Data Mining*. The AAAI Press.

Goldriech, O. (2001). *Secure multipart computation* [Working draft, version 1.3].

Han, E., Karypis, G., & Kumar, V. (1997). Scalable parallel data mining for association rules. *Proceedings of ACM SIGMOD*, 277-288.

Kantercioglu, M., & Clifton, C. (2002). Privacy preserving distributed mining of association rules on horizontal partitioned data. *Proceedings of ACM SIGMOD Workshop of Research Issues in Data Mining and Knowledge Discovery DMKD*.

Kohavi, R., Broadley, C., Frasca, B., Ma-

son, L., & Zheng, Z. (2000). KDD-Cup 2000 organizers report: Peeling the onion. *SIGKDD Explorations, 2*(2), 86-98.

Mueller, A. (1995) *Fast sequential and parallel association rule mining algorithms: Comparison* (Tech. Rep. No. CS-TR 3515). College Park: University of Maryland.

Naor, M., & Nissim, K. (2001). Communication preserving protocols for secure function evaluation. *Proceedings of the ACM Symposium on Theory of Computing,* 590-599.

Rizvi, S. J., & Haritsa, J. R. (2002). Maintaining data privacy in association rule mining. *Proceedings of 20th International Conference on Very Large Databases,* 682-693.

Vaidya, J., & Clifton, C. (2002). Privacy preserving association rule mining in vertically partitioned data. *Proceedings of ACM SIGKDD,* 639-644.

Wirth, R., Borth, M., & Hipp, J. (2001). When distribution is part of the semantics: A new problem class for distributed knowledge discovery. *Proceedings of Joint 12th European Conference on Machine Learning and Fifth European Conference on Principles and Practice of Knowledge Discovery in Databases*.

Zaki, M. J. (1999). Parallel and distributed association mining: A survey. *IEEE Concurrency, 7*(4), 14-25.

Zaki, M. J., Ogihara, M., Parthasarathy, S., & Li, W. (1996*). Parallel data mining for association rules on shared-memory multiprocessors* (Tech Rep. No. TR 618). University of Rochester, Computer Science Department.

Mafruz Zaman Ashrafi is a PhD student at Monash University. His research interests include data mining, distributed computing, e-commerce, and security. He received his MS in computer science from RMIT University.

David Taniar holds bachelor's (Honours), master's, and PhD degrees — all in computer science, with a particular specialty in databases. His research areas now expand to Data Mining and Warehousing. He has published more than 100 research articles in international journals and conference proceedings. He has published five books, including two edited books in the area of Web Technology, covering web databases and web mining. Dr. Taniar is now a senior lecturer at the School of Business Systems, Faculty of Information Technology, Monash University, Australia. He is on the editorial board of several international journals. He is elected as a fellow of the Institute for Management Information Systems (FIMIS).

Kate Smith is an associate professor, deputy head and director of research in the School of Business Systems at Monash University, Australia. She holds a BSc (Honours) in mathematics and a PhD in electrical engineering, both from the University of Melbourne, Australia. She is also co-director of the Monash Data

Mining Centre. Dr. Smith has published more than 130 refereed journal and international conference papers in the areas of neural networks, combinatorial optimization, and data mining. She has received around $1.5 million in competitive research funding. She serves on the editorial board of several journals, and is a member of the organizing committee for many international data mining and neural network conferences. She regularly acts as a consultant to industry in the area of data mining.

LIBRARY RECOMMENDATION FORM for IGI Journals

Please forward to your institution's librarian

Dear _____
(librarian's name)

I would like to recommend that our library subscribe to the following Idea Group Inc. publications. (I have indicated the justification for the subscription(s) below.)

JOURNALS

- ❑ **Journal of Cases on Information Technology /** ISSN 1548-7717/Organizational rate US$245
- ❑ **Information Resources Management Journal /** ISSN 1040-1628/Organizational rate US$315
- ❑ **Journal of Organizational and End User Computing /** ISSN 1546-2234/Organizational rate US$315
- ❑ **Journal of Global Information Management /** ISSN1062-7375 /Organizational rate US$315
- ❑ **Journal of Database Management /** ISSN 1063-8016/Organizational rate US$315
- ❑ **Journal of Electronic Commerce in Organizations /** ISSN 1539-2937/Organizational rate US$225
- ❑ **Intl Journal of Distance Education Technologies /** ISSN 1539-3100/Organizational rate US$225
- ❑ **Intl Journal of IT Standards & Standardization Research /** ISSN 15398-3062/Organizational rate US$205
- ❑ **Intl Journal of Web Services Research /** ISSN 15398-3062/Organizational rate US$225
- ❑ **Intl Journal of Infor. and Communication Technology Education** / ISSN: 1550-1876/Organizational rate US$195
- ❑ **International Journal of Knowledge Management** / ISSN: 1548-0666/rganizational rate US$195
- ❑ **International Journal of Intelligent Information Technologies** / ISSN: 1548-3657/Organizational rate US$195
- ❑ **International Journal of Technology and Human Interaction** / ISSN: 1548-3908/Organizational rate US$195
- ❑ **International Journal of Data Warehousing and Mining** / ISSN: 1548-3924/Organizational rate US$195
- ❑ **International Journal of E-Collaboration** / ISSN: 1548-3673/Organizational rate US$195
- ❑ **International Journal of Enterprise Information Systems** / ISSN: 1548-1115/Organizational rate US$195
- ❑ **Intl Journal of Business Data Communications & Networking** / ISSN: 1548-0631/Organizational rate US$195
- ❑ **International Journal of E-Business Research** / ISSN: 1548-1131/Organizational rate US$195
- ❑ **International Journal of Electronic Government Research** / ISSN: 1548-3908/Organizational rate US$195
- ❑ **Journal of Cases on Electronic Commerce** / ISSN: 1548-0623/Organizational rate US$195
- ❑ **International Journal of Semantic Web and Information Systems** / Organizational rate US$195
- ❑ **Information Management /** ISSN 1080-286x/Organizational rate US$70
- ❑ **Information Technology Newsletter /** ISSN 1057-7939 (US$35.00 /year)

_____ _____
(Recommended by) (Department or Division)

THE BENEFIT OF THE JOURNAL(S) FOR OUR LIBRARY IS:

I have indicated the appropriate response (1=highest benefit; 2=moderate benefit; 3=little benefit).

1	2	3	**REFERENCE:** For research articles in the field of information technology management.
1	2	3	**STUDENT READINGS:** I plan to recommend articles from the above to my students.
1	2	3	**PUBLICATION SOURCE:** This journal is a suitable source for my current research results and findings.
1	2	3	**PEER EVALUATION:** This journal is highly regarded by my peers and is cited often in other's research work.

Note to the librarian: Idea Group Inc. (IGI) journals can be ordered directly or through your designated subscription agency.

IGI journals are available both in print and electronic form for one price.

Questions or orders may be forwarded to:

Idea Group Inc.
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661
E-mail: cust@idea-group.com; http://www.idea-group.com

CALL FOR MANUSCRIPTS

International Journal of Intelligent Information Technologies

An official publication of the Information Resources Management Association!

Prospective authors are invited to submit manuscripts for consideration for publication in the International Journal of Intelligent Information Technologies. IJIIT publishes original material concerned with all aspects of intelligent information technologies. The journal invites both conceptual and empirical manuscripts of high quality not currently under review by another publication.

ISSN 1548-3657
eISSN 1548-3365
Published quarterly

MISSION:

The advent of the WWW has sparked renewed interest in the area of intelligent information technologies. There is a growing interest in developing intelligent technologies that enable users to accomplish comples tasks in Web-centric environments with relative ease, utilizing such technologies as intelligent agents, distributed computing in heterogeneous environments and computer supported collaborative work. The mission is to bring together researchers in related fields such as information systems, distributed AI, intelligent agents; and collaborative work, to explore and discuss various aspects of design and development of intelligent technologies.

COVERAGE/MAJOR TOPICS:

- manufacturing information systems
- privacy, security and trust issues
- Semantic Web, Web services
- intelligent decision support systems
- and more! Please visit www.idea-group.com/ijiit

Full submission guidelines available at: http://www.idea-group.com

All submissions should be emailed to: sugumara@oakland.edu

Now when your institution's library subscribes to any IGP journal, it receives the print version as well as the electronic version for one inclusive price. For information contact Corrina Chandler, Electronic Content Coordinator, 717/533-8845, ext. 14

ORDER FORM

- ❑ I wish to join the INFORMATION RESOURCES MANAGEMENT ASSOCIATION (Membership entitles you to JGIM and the semi-annual Information Management). $65 Individual*; $195 Institutional* $ _____
- ❑ I wish only to subscribe to the Journal of Global Information Management.* Individual $95.00 Institutional $315.00 $ _____

*Outside USA, add $15

IDEA GROUP PUBLISHING
701 E. Chocolate Ave., Suite 200
Hershey, PA 17033-1240, USA
Tel: 717/533-8845
Fax 717/533-8661

Name _____

Organization _____

Address _____

City/State/ZIP _____

Telephone _____

METHOD OF PAYMENT

___ Check or money order payable to Idea Group Publishing
___ Please bill my company
___ Please charge my ____ MasterCard ___ VISA ___ Am. Express

Card No. _____ Expiration Date _____

Signature _____

PLEASE RECOMMEND THIS JOURNAL TO YOUR LIBRARY!

Join the international organization serving information providers and users!

INFORMATION RESOURCES MANAGEMENT ASSOCIATION

Membership Application

Name: _____

Affiliation: _____

Address: _____

City/State/ZIP _____

Country: _____

Telephone: _____

Fax: _____

E-mail: _____

Join online at:
http://wwww.irma-international.org/
***or* return with payment to:**
IRMA
701 E. Chocolate Ave., Suite 200
Hershey PA 17033-1240 USA
or Fax to: 717/533-8661

Federal ID# 25-1896021

Please check one or more of the following selections.

❒ **IRMA Basic Membership+**
(Includes subscription to Information Management and one journal.)

Individual -One Year US$65/ Two Years US$120*	**US $_____**
Organizational-One Year US$195/Two Years US$380*	**US$_____**
Student One Year only US$35™ *	**US $_____**

Please check your preferred journal.

___Information Resources Management Journal (4 issues)
___Journal of Global Information Management (4 issues)
___Journal of Cases on Information Technology (4 issues)
___ Intl. Journal of Distance Education Technologies (4 issues)
___ Intl. Journal of Web Services Research (4 issues)
___ Intl Journal of Info Technology & Comm Education (4 issues)
___ Intl Journal of Intelligent Info Technologies (4 issues)
___ Intl Journal of Data Warehousing & Mining (4 issues)
___ Intl Journal of Enterprise Info Systems (4 issues)
___ Intl Journal of E-business Research (4 issues)
___Journal of Organizational & End User Computing (4 issues)
___ Journal of Database Management (4 issues)
___ Intl. Journal of IT Standards & Standardization Research (2 issues)
___ Journal of Electronic Commerce in Organizations (4 issues
___ Intl. Journal of Cases on Electronic Commerce (4 issues)
___ Intl. Journal of Knowledge Management (4 issues)
___ Intl Journal of Tech & Human Interaction (4 issues)
___ Intl Journal of E-Collaboration (4 issues)
___ Intl Journal of Business Data Commun. & Networking (4 issues)
___ Intl Journal of Electronic Govt Research (4 issues)

❒ **Subscription to each additional journal(s)**-*Please check above*

Individual -One Year US$50/Two Years US$90*	**US$_____**
Organizational-One Year US$80/Two Years US$150*	**US$ _____**
Student -One Year only US$20 ™ *	**US$ _____**
**** Outside U.S., add $15 per year for each journal ordered.***	**US$ _____**
TOTAL	**US$ _____**

™ For student membership eligibility, you must enclosed a signed statement by your academic advisor verifying your student status.
+ Lifetime Basic Membership can be obtained for $800 with subscription to additional journal(s) available for an additional $750 each.

METHOD OF PAYMENT

❒ Check or money order payable to IRMA (and listing a U.S. corresponding bank)
❒ Credit Card ___ MasterCard ___ VISA ___ American Express

Amount that you are authorizing to be charged US $ _____

Credit Card #:_____ Expiration Date: _____

Signature _____

3-digit security code (MC and Visa) on signature panel or 4-digit code (AmEx) on card front: _____

Some of the benefits of IRMA membership include:

- Journals crucial to your professional enrichment.
- Discounts on your IRMA International Conference registration and association publications.
- Interaction with colleagues around the world.
- Participation on professional leadership committees of IRMA.

ISSN 1548-3657