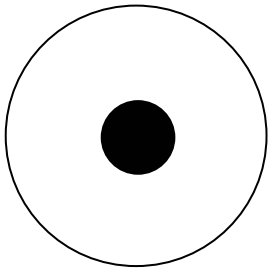
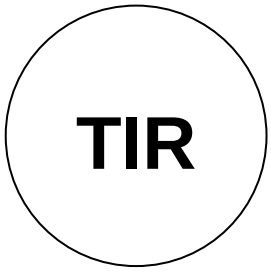




The IPSI BgD Transactions on Internet Research

Multi-, Inter-, and Trans-disciplinary Issues in Computer Science and Engineering

A publication of
IPSI Bgd Internet Research Society
New York, Frankfurt, Tokyo, Belgrade
January 2006 Volume 2 Number 1 (ISSN 1820-4503)

Table of Contents:

The IPSI BgD Internet Research Society

The Internet Research Society is an association of people with professional interest in the field of the Internet. All members will receive this TRANSACTIONS upon payment of the annual Society membership fee of €100 plus an annual subscription fee of €1000 (air mail printed matters delivery).

Member copies of Transactions are for personal use only

IPSI BGD TRANSACTIONS ON INTERNET RESEARCH

www.internetjournals.net

Secured Object Signatures for Web Objects and XML Objects

Chang, H., Peter

Abstract— ***This paper presents object-oriented security models for a secured object signature of web and XML objects. The models address web and XML securities in general. They address in particular security wrappers for web pages, security nodes for network machines, and mid-tier securities for middle-tier machines. This paper discusses in a high level implementations for security nodes and mid-tier securities. It discusses an implementation for security wrappers in detail. The implementation encrypts web and XML objects in a company's Intranet web site and enhances the protections of them against unauthorized accesses by an intruder. The encryption method uses the combination of a randomly generated session key, a session id that consists of remote host id and port number, and an access sequence number that is maintained by the web server. Although the encryption method addresses the security wrappers, its approach is equally applicable to security nodes and mid-tier securities as well.***

Index Terms—*object-oriented, security, Web, XML, XSL*

1. *Introduction*

As Internet e-commerce activities grow more and more every year, Internet security is also becoming more important in securing the capitals invested in these activities. Oppliger [8] stated in a workshop held by the Internet Architecture Board (IAB) back in 1994 that scaling and security were considered to be the two most important problem areas for the Internet as a whole. Chang [4] uses an object-oriented analysis and modeling approach to define the architecture of a proposed method for securing web objects. It is known that object has both attributes and behaviors. Blaha et al. [2] covers in detail the object-oriented terminology and technology. Network machines, web servers, web documents (XHTML web pages), XML documents, XSL documents, database servers, and mid-tier machines are examples of web or XML objects.

For web documents, examples of attributes are URL web address, title, and hyperlinks. Examples of behaviors are navigation to other web pages through hyperlinks and the dynamic operations implemented by JavaScript functions defined in the web pages.

This paper covers the security model for web and XML documents, network machines, and mid-tier machines. For brevity, it concentrates on web and XML documents and covers the network machines and mid-tier machines in high level. The implementation approach is applicable to the latter two. Web pages and XML documents are stored or generated by programs on a web server. They may be protected from unauthorized people with the addition of a security mechanism in the server programs. Static web pages in XHTML files are in general accessed through hyperlinks in other web pages. Static web pages can also be represented in XML and XSL files. XHTML files, XML/XSL files, and XML files in general are more vulnerable to unauthorized accesses by an outside intruder. This paper concentrates the discussion on the intranet web site as it is used more for B2B activities.

Common encryption techniques involve with encrypting the whole document. Kalakota et al. [6] contains general discussions on encryption and web transaction security. W3C [12] provides general guidelines on securing XML documents for Web services. W3C [13] provides guidelines for XML digital signatures. Encryption processes that encrypt whole documents generally incur performance overheads. Sometimes, it is not the whole content of the documents that need to be encrypted. Rather it is how the access to these documents, for example, a hyperlink for a web document, needs to be protected through encryption.

This paper presents a selective encryption approach in the sense that instead of encrypting a whole document it encrypts a selected element of the document, using the hyperlink element as an example. In other words, it uses an encryption method that protects the access to XHTML documents or their XML/XSL equivalences through hyperlinks. This paper names the encrypted result for a hyperlink a *secured object signature* of the hyperlink. The session key, which is used as the encryption key, is first generated when a user logs into the web site. This paper introduces a concept of

Manuscript received October 12, 2005. P. H. Chang is with the College of Management, Lawrence Technological University, Southfield, Michigan, U.S.A. (e-mail: chang@ltu.edu).

including placeholders and references to templates to the hyperlinks. It introduces the access sequence number that tracks a user's access, which is incremented each time the user of the current session accesses a XHTML document or its XML/XAL equivalence through a hyperlink. This paper also introduces the algorithm for a generic Common Gateway Interface (CGI) script, implemented in the Perl language, that handles the events of the access through hyperlinks. This paper uses a database managed by a Relational Database Management System (RDBMS) to track and update the access sequence number. Hoffer et al. [7] covers the RDBMS technology. This paper presents one concrete example for the implementation that an XHTML document or its XML/XSL equivalence for a company's phone directory is listed through a hyperlink on the company's intranet home page. In this paper, after the encryption method is applied to the hyperlink reference, two scenarios will be discussed. The first scenario is that a valid user accesses the phone directory through the hyperlink and gets through. The second scenario is that an outside intruder attempts to access the phone directory and is denied the access.

2. SECURITY ARCHITECTURE FOR WEB OBJECTS

Figure 1 describes an object model for a client machine that hosts a web browser application, which accesses web or XML documents on web servers in a company's intranet or the Internet. The rest of this paper considers a web page or an XML document as an object instance of the WebDocument object class. For an ease of discussion, this paper concentrates the discussion on web pages up to Section 5. Discussions on XML documents will follow on Section 6.

When the browser requests a web page from a web server, the web server downloads the web page to the client machine to be rendered by the browser and displayed to the user.

In object-oriented terminology, a relationship among object classes is called an association. Binary association, the association between two classes, is the most common ones. In some cases, an object class may be associated with itself due to that it has two roles, for example, a person class that models a company's workers may have two roles: manager and employee. Such an association is called the unary association.

Figure 1 shows examples of binary associations. One of them is the association between the WebServer class and the WebDocument class. Figure 1 also shows examples of unary associations. One of them is the association between web pages that contain instances of the XHTML anchor tag for hyperlink: <a href="url">, referred as the link source, and the web pages identified by the web address' url, referred as the link target. For each web page as an instance of the WebDocument class, there may be zero or more hyperlinks contained therein. Each web page may also be referred to as link reference by zero or more web pages. Figure 1 shows the multiplicity reflecting the many-to-many association. For a many-to-many association, it is necessary to create an association class to correctly model the association because there may be attributes that belong more appropriately to the association rather than to the WebDocument class. The association class is depicted as SecurityWrapper in Figure 1. This association class may model security methods such as encryption and database management system (DBMS) facilities for the web pages. Blaha et al. [2] discusses in detail the association class.

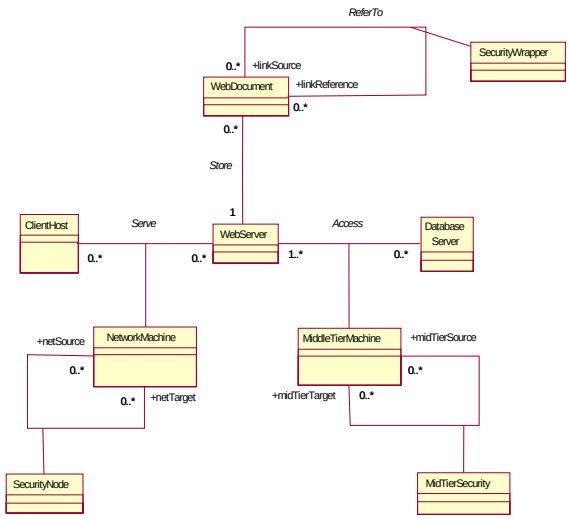


Figure 1. Object Model for the Security Architecture

The communication protocol between the client machine and the web server is HTTP. Data packets are transmitted between the client and server through network machines, which are modeled as the object class NetworkMachine in Figure 1. Examples of the network machines are hubs, switches, routers, domain name servers, and so on. A network machine may transmit data packets to many other network machines. Conversely a network machine may also receive data packets transmitted from many other network machines. So this association is a many to many unary association. From the same reason stated above, it is necessary to create an association class for this many-to-many association. This association class may model a security node that enforces security for data transmission depicted as SecurityNode in Figure

1. Examples of security nodes are firewalls and proxy servers.

Web servers may also need to access database servers through intermediate machines that host middleware software, which are modeled as the object class MiddleTierMachine in Figure 1. Similarly there is a many-to-many unary association on the MiddleTierMachine object class because a mid-tier machine may transmit (receive) data packets to (from) many other mid-tier machines. It is also necessary to create an association class for this association. This association class may model the security among mid-tier machines depicted as MidTierSecurity in Figure 1.

The rest of this paper generally refers the three association classes: SecurityWrapper, SecurityNode , and MidTierSecurity as the security association class. Chang [4] presented an earlier version of the object model that addresses the association class SecurityWrapper, but does not address the association classes SecurityNode and MidTierSecurity.

3. IMPLEMENTATION OF SECURITY ASSOCIATION CLASS

A security standard such as ISO X. 700 generally addresses security in areas that contain access, authentication, authorization, and auditing. Additionally areas like data confidentiality, data integrity, and non-repudiation services to communicating peers are also emphasized. In addition, some object-oriented approaches for security modeling addresses the issues of security in each of the seven layers of ISO's OSI model.

Many security implementations nowadays use configuration parameters stored in files instead of an operational databases because the number of these parameters is generally smaller than the number of transactional records that require a database and it is generally simpler to add, update, and delete the parameters values in files. Redundancy, however, leads to inconsistency. For example, suppose a security parameter value is stored in multiple files. In the event that this value needs to be changed, it should be changed in all the files that contain it. If it is not changed in one file, then the security measure related to this parameter is compromised. The object model depicted in Figure 1 provides a solid foundation for implementation to databases. The benefit of database such as a relational database is that the redundancy of data is minimized. The performances of DBMS' and the server software and hardware platforms that support them are getting better. In addition, the DBMS itself is a security layer that adds extra level of security that the file system is lack of. For example, most relational database management systems (RDBMS') provide record and field level security, while file system provides only security at the file level.

Implementations of the *SecurityNode* association class include firewalls, screen routers, and proxy servers, which use the approach of packet filtering to examine and control the transmission of data packets on the exterior network entering the intranet. Oppliger [8] covers firewalls in more details. Oppiger's paper, however, does not address how are the security configuration data such as values of security parameters stored. The object model in Figure 1 may be expanded as a basis to design databases that store them.

Implementations of the *MidTierSecurity* association class include establishing and enforcing security policies for applications that span multiple administrative domains. The policy addresses among others security events and the associated services. A commonly adopted paradigm for event services is the publisher-subscriber model. Middleware software resides on the mid-tier machines. Tripathi [10] covers the middleware security. Tripathi's paper also does not address how are the security configuration data such as values of security parameters stored. The object model in Figure 1 may also be expanded as a basis to design databases that store them.

As for an implementation of the *SecurityWrapper* association class this paper proposes a method of using encryption and database managed by a DBMS. Common Gateway Interface (CGI) script is one protocol that a web server uses to respond to clients' requests sent by web browsers through web pages. Other protocols include Microsoft's Active Server Page (ASP) and Sun's Java Server Page (JSP) and servlets. These other means are aligned with web services; for example, ASP is with Microsoft ASP.NET (Tabor [9]), and JSP/Servlet is with Sun Sunone (Watson [11].) Web services are based on XML, which will be discussed in Section 6.

For brevity this paper assumes that the web servers execute CGI programs responding to clients' requests. CGI is usually implemented with programming languages such as Perl or C++/C. These languages require relatively less supporting libraries than those languages used for J2EE or .NET framework. The CGI examples presented in this paper were developed in the Perl language (Bunce et. al. [1]). The same method is applicable to other protocols and programming languages such as Java or C#.

Such protocols and tools are parts of the implementations of the middleware concept. Chang [3] discusses similarities among implementations of the middleware.

The association class *SecurityWrapper* may model any security feature that secures the implementation of the association between the group of instances of WebDocuments with link source role and the group of instances of WebDocuments with link target role. This paper proposes the addition of an encryption wrapper to the URL address in the XHTML anchor tag as the security feature. The encryption is implemented in CGI programs written in the Perl language. The key for the encryption is generated randomly. After its generation, its value is maintained in a database table. The remaining sections of this paper describe this implementation.

4. SESSION ID

There are in general two types of access to web documents. The first type is a viewing of static web page; for example, reading a newspaper article on one of the web site of a newspaper company. In this case, a reader simply browses over the static web pages. The web server does not need to know the identification of the reader and hence does not need to maintain a session with the reader. The second type is transaction oriented, that is, a user interacts with the web page by providing input data for a period of time; for example, doing online shopping through an implementation of the shopping cart design. In this case, a user needs to provide input to multiple web pages. The web server needs to know an identification of the user and hence needs to maintain a session with the user.

A web server maintains two environment variables associated with the client: REMOTE_HOST and REMOTE_PORT. The former contains the IP address of the client machine and the latter contains the port number used by the client machine to communicate with the web server.

Assume that a user uses a desktop or laptop personal computer to access web documents on a web server. The combination of REMOTE_HOST, REMOTE_PORT, and the web server system time at the first instance of access uniquely identifies the session for the user. From now on this paper considers a session id to be of the format: ipAddress-portNumber-timestamp, where ipAddress is the value of REMOTE_HOST, portNumber is the value of REMOTE_PORT, and timestamp is the system time on the web server. This paper addresses an approach that when a user logs on to a web page to start a session, the login CGI script generates a session id as was just described and a session key to be used as the encryption key.

5. PLACEHOLDERS AND ENCRYPTION

In many cases, intranet users attempt to log into their intranet web sites from remote locations outside their working facilities through the Internet. In responding to a client request for a user's login, after authenticating the user, a CGI script generates a home web page containing welcoming information, which may possibly include hyperlinks.

A scenario is that the home web page named, say, index.html in a company's intranet web site includes a hyperlink to the employee phone directory as shown in Figure 2.

```
<a href="phoneDirectory.html">Phone Directory</a>
```

Figure 2. Hyperlink to the web page for Phone Directory

The phone directory may be a static XHTML file stored on the web server. Thus, if a user successfully logs on the company's intranet web site, the user can freely access the phone directory through the hyperlink. Inductively, the user can freely access any static XHTML file that is referenced by a hyperlink on the web page that the user is currently on. As was mentioned in Section 4, the IP address of the client machine and the port number used by the client machine are contained in the two environment variables: REMOTE_HOST and REMOTE_PORT, respectively. If the values for these two variables, which form the session id, are intercepted on the Internet by an intruder, the intruder can conceivably at least get onto the company's intranet web site and access all static XHTML files that are referenced by hyperlinks in the intranet home page.

```
<a href="getcontent.pl?template=phoneDirTemplate.html
&sessionId=ZZZZZZS&sessionKey=ZZZZZK
&accessSeqNumber=ZZZZZA">Phone Directory</a>
```

Figure 3. Hyperlink in index.html with Placeholders

This paper presents an encryption algorithm that adds an extra layer of security that protects an important Web object such as the hyperlink in Figure 2. An implementation code for this algorithm is available from the author upon request. From this code, the decryption algorithm can be derived. This paper's goal is using an algorithm that is effective, but not very complex and thus does not add too much overhead to the processing time.

This paper first proposes to replace the

reference to the name of the static web page in the home web page index.html by the name getcontent.pl of a CGI script with a template file name and placeholders as its arguments as shown in Figure 3.

The template file phoneDirTemplate.html is a template file for the static web page phoneDirectory.html discussed previously. It differs from the latter only in that it contains placeholders as replacements for static links in the latter. These static links are inserted by the owner of the file phoneDirectory.html that allow a user to navigate to some common web pages such as the home page. Figures 4 shows a hyperlink in phoneDirectory.html to home web page. Figure 5 shows the revised hyperlink in the template file phoneDirTemplate.html.

```
<a href="index.html">Home</a>
```

Figure 4. Hyperlink in *phoneDirectory.html* to the web page for home

```
<a href="getcontent.pl?template=../htdocs/index.html
&sessionId=ZZZZZZS&sessionKey=ZZZZZK
&accessSeqNumber=ZZZZZA">Home</a>
```

Figure 5. Hyperlink in *phoneDirTemplate.html* with Placeholders

Assume that the login web page contains the following line:

```
<form method="post" action="login.pl">
```

where login.pl is the name of the CGI script that handles the authentication of users. The login.pl script sets up a session key, a two digits number less than 62, as the key for the encryption. The reason that the session id is not used as the key is that it tends to be a long string, which is not suitable for the encryption algorithm that is propose in this paper. The key steps of the login.pl script involve with setting up encryption environment and session id, authenticating the user who logs in, updating database table with 0 as the initial access sequence number, encrypting session id and the access sequence number, and replacing the placeholders ZZZZZZS, ZZZZZK, and ZZZZZA in Figure 3 by the encrypted data for session id, the session key, and the encrypted data for access number, respectively. The resulting hyperlink acts like a secured object signature for the hyperlink in Figure 2.

To show that the encryption secures the static hyperlinks, this paper considers two scenarios. The first scenario is that a valid user logs on to the intranet home page and at some point of the session accesses the phone directory through its hyperlink. The second scenario is that an intruder intercepts a copy of a transmitted XHTML file and manages to get on the intranet home page. The intruder then attempts to access the phone directory through the hyperlink.

For the first scenario, a CGI script named getContent.pl first validates that the access sequence number matches the one in the database. If it does not, then it logs an error message and terminates the CGI program. Otherwise it updates the access sequence number in the database and displays the content of the template of the phone directory with the substitutions of encrypted session id, session key, and encrypted access sequence number for the placeholders in the hyperlinks, if they exist.

The key steps of the getContent.pl script involve with decryption and checking whether the access sequence number is in sequence. That is, whether it matches the access sequence number in the database. If they do not match, then there is a possible intrusion. In this case, reject the access. If they match, then increment the access sequence number in the database by 1, re-encrypt the this access sequence number along with session id, and repeat the steps of replacing placeholders as those for the login.pl script.

For the second scenario, the intruder intercepts a copy of the transmitted XHTML file. When the intruder gets on the company's intranet web site, the access sequence number is out of sequence, because the session for the valid user has already updated it in the UserAcess table in the database. If the intruder wants to access the phone directory through the hyperlink, the getContent.pl script will discover that the access sequence number with the session id and session key on the intercepted copy does not match the one in the UserAccess table in the database. The getContent.pl script then denies the intruder from accessing the company's phone directory.

6. *Encryption in XML/XSL and Selective Encryption*

XML is used in many areas nowadays such as Web services technology and databases. In particular, a web page in XHTML format can be represented by an XML file and an XSL file. For example, the XHTML element in Figure 2:

```
<a href="phoneDirectory.html">Phone Directory</a>
```

can be represented by the XML element:

```
<PhoneDirectory>
  <name>Phone Directory</name>
  <link>phoneDirectory.html</link>
</PhoneDirectory>
```

and the corresponding XSL element:

```
<xsl:variable name ="linkname">
    <xsl:value-of select = "PhoneDirectory/link"/>
</xsl:variable>
 <a href="{$linkname}"><xsl:value-of select =
 "phoneDirectory/name"/></a>
```

W3C [14] defines the XSL syntax. The XHTML element in Figure 3

```
<a href="getcontent.pl?template=phoneDirTemplate.html
   &sessionId=ZZZZZZS&sessionKey=ZZZZZK
   &accessSeqNumber=ZZZZZA">Phone Directory</a>
```

can also be represented by the following XML elements:

```
<PhoneDirectory>
  <name>Phone Directory</name>
  <link>getcontent.pl</link>
  <template>phoneDirTemplate.html</template>
  <sessionId>ZZZZZZS</sessionId>
  <sessionKey>ZZZZZZK</sessionKey>
  <seqNumber >ZZZZZA</ seqNumber >
</PhoneDirectory>
```

and the corresponding XSL elements:

```
<xsl:variable name ="linkname">
  <xsl:value-of select = "PhoneDirectory/link"/>
</xsl:variable>
<xsl:variable name ="template">
  <xsl:value-of select = "PhoneDirectory/ template"/>
</xsl:variable>
<xsl:variable name ="sessionId">
  <xsl:value-of select = "PhoneDirectory/sessionId"/>
</xsl:variable>
<xsl:variable name ="sessionKey">
  <xsl:value-of select = "PhoneDirectory/sessionKey"/>
</xsl:variable>
<xsl:variable name ="seqNumber">
  <xsl:value-of select = "PhoneDirectory/seqNumber"/>
</xsl:variable>
<a
href="{$linkname}"?template={$template}&sessionId={$s
essionId}&sessionKey={$sessionKey}&accessSeqNumber=
{$seqNumer}><xsl:value-of select =
"phoneDirectory/name"/></a>
```

XML is used in many other areas as well such as second-generation Web services and Service-Oriented Architecture (SOA). The role of a computer program as an XML parser is that it parses the structures of an XML data packet or an XML file. In the last example, the CGI script getContent.pl and the web browser are the XML parsers. The approach of using a template file with placeholders and applying an encryption method to convert the lines containing the placeholders to secured object signatures for certain XML elements can be applied to XML in general.

Geer [5] points out a research done by Ron Schmelzer of the market research firm ZapThink that XML will rise from 3 percent of global network traffic in 2003 to at least 40 percent by 2008. Some of the XML traffic may include overheads of full data encryption. The selective encryption approach proposed in this paper may be used with the binary XML approach discussed in Geer [5] to lessen the XML traffic volumes.

7. CONCLUSION

Ensuring web security is a complex problem that involves almost every aspect of the enterprise of an organization. Object-oriented analysis and modeling technique is especially adaptable to analyze complex problems and model solutions for them. This paper covered the secured object signatures based on a security model for web documents and XML documents, network machines, and mid-tier machines by using the object-oriented technique. The security model is based on three security association classes: *SecurityWrapper*, *SecurityNode*, and *MidTierSecurity* as described in Section 3. Although this paper concentrated on the security association class *SecurityWrapper* by presenting the approach that uses encryption methods and databases, the approach is applicable to the two other security association classes, which span the whole spectrum of the network infrastructure for web and XML applications.

REFERENCES

[1] Bunce, T., Descartes, A., "Programming the Perl DBI," *O'Reilly and Associates, Inc*., 2000.

[2] Blaha, M., W. Premerlani, W., "Object-Oriented Modeling and Design for Database Applications," *Prentice Hall*, 1998.

[3] Chang, P., "A Platform Independent Middleware Architecture for Accessing Databases on a Database Server on the Web," *IEEE*, IEEE R4 Conference, Indianapolis, Indiana, June 2003.

[4] Chang, P., "On An Architecture for Securing Web Objects," ASEE/NCS 2004 Conference, Western Michigan University, Kalamazoo, Michigan, April 2004

[5] Geer, D, "Will Binary XML Speed Network Traffic?," *IEEE*, IEEE Computer, April 2005, 16-18.

[6] Kalakota, R., Whinston, A., "Electronic Commerce, A Manager's Guide," *Addison Wesley Longman, Inc*., 1997.

[7] Hoffer, J., Prescott, M., McFadden, F., "Modern Database Management," 6th ed., *Prentice Hall*, 2002.

[8] Oppliger, R., "Internet Security: Fire Walls and Beyond.," *ACM*, Communications of the ACM, Volume 40 Issue 5, May 1997. 92-102.

[9] Tabor, R., "Microsoft .NET XML Web Services," *SAMS*, 2002.

[10] Tripathi, A., "Challenges Designing Next-Generation Middleware Systems," *ACM*, Communications of the ACM, Volume 45 Issue 6, June 2002. 39-42.

[11] Watson, M., "Sun ONE Services," 1st ed., *John Wiley & Sons*, 2002.

[12] W3C, "Web Service Architecture Requirements," http://www.w3.org/TR/2002/WD-wsa-reqs-20021114.

[13] W3C, "XML Signature WG," http://www.w3.org/Signature/

[14] W3C, "XSL Transformations," http://www.w3.org/TR/xslt, (XSLT), Version 1.0, 1999.

Toward Next Generation Distributed Business Information Systems: Five Inherent Capabilities of Service-Oriented Computing

Chung, Sam and Davalos, Sergio

Abstract— The research conducted examines how the emerging Service Oriented Computing (SOC) paradigm enables the development of Distributed Business Information Systems (DBIS) and thus influences the acceptance of DBIS. For this purpose, a Service Oriented Architecture (SOA) that can be implemented by currently available SOC technologies is used. The SOA is applied to the development of DBISs for retail Supply Chain Management (SCM) and stock currency conversion. To analyze the use of the SOC paradigm on the development of DBISs, the following five inherent capabilities of SOC are proposed: interoperability, reusability, adaptability, interdependability, and agility. We investigate how the proposed capabilities are naturally provided by the SOC paradigm in the development of the DBISs. These capabilities enable clearer communication between software and business professionals by allowing 'software as a service' and 'business process as a service.'

Index Terms—Adaptability, Agility, Business-to-Business Application Integration, Distributed Information System, Enterprise Application Integration, Interdependability, Interoperability, Reusability, Service Oriented Architecture, Service Oriented Computing, Supply Chain Management

1. INTRODUCTION

The purpose of this research is to examine how the Service-Oriented Computing (SOC) paradigm enables the development of Distributed Business Information Systems (DBIS and thus influences the acceptance of DBIS through its five natural capabilities - interoperability, reusability, adaptability, interdependability, and agility. Web services, a promisingly better technology for integrating heterogeneous applications, has recently emerged [1, 2, 3].

A web service uses standard interface definitions and general interaction protocols based upon eXtensible Markup Language (XML).

Manuscript received September 26, 2005. This work was supported in part by the University of Washington Tacoma under Grant Founder's Endowment.

S. Chung is with the Computing & Software Systems program, Institute of Technology, University of Washington Tacoma, USA (e-mail: chungsa@u.washington.edu).

S. Davalos is with the Milgard School of Business, University of Washington Tacoma, USA (e-mail: sergiod@u.washington.edu).

For the integration of applications, software developers have adopted the SOC paradigm in the belief that the SOC paradigm can overcome implementation complexity involved with the current distributed object computing paradigms [1, 2].

In this research, two models of Service Oriented Architecture (SOA) are described that are based on currently available SOC technologies: 'publish-discovery-binding' and 'publish-compose -discovery-binding' models. The SOA models are then applied in the development of DBISs for retail Business Information System (BIS), Supply Chain Management (SCM) [4], and stock currency conversion.

For the retail BIS, two heterogeneous BISs are developed to support the operations of a retailer and a supplier. The business logic layer for each BIS is implemented with web services. Web services are published to public service registries. The two BISs are then integrated with the published web services for Business-to-Business (B2B) transactions between the retailer and the supplier.

The primary purpose of the second application, a Stock Currency Converter (SCC) system, is to examine the use of a composite web service on a business process execution engine and utilizing Service-Oriented Programming (SOP). A composite web service is built by composing two existing web services.

The use of SOC on the development of DBISs is examined based upon the proposed five natural capabilities of the SOC paradigm. In terms of the five capabilities, naturally provided by the SOC, the case study results show that the SOC paradigm is very capable for developing Distributed BISs requiring Enterprise Application Integration (EAI) and B2B Application Integration (B2BAI). Both software and business professionals share the common conceptual building block – a service: software as a service and business process as a service.

This paper consists of six sections. In the next section, two different types of application integration based on the SOA and its application to EAI and/or B2BAI are discussed. In Section 3,

we propose and define five inherent capabilities of SOC based upon the SOA that can be used for evaluating the development of DBIS. In section 4, two case studies are presented that require successful use of the five capabilities in a DBIS environment. One case study involves B2B transactions in retail SCM, which need EAI and B2BAI. The second case study continues the examination of a DBIS where Stock Currency Converter (SCC) system transactions are implemented by using .NET web services and SOP. In Section 5, we discuss the results of the application of the SOC paradigm to the development of distributed BISs in terms of the five capabilities. In Section 6, we conclude that the five inherent capabilities show why the SOC computing paradigm is very effective for EAI as well as B2BAI. Also, we show how it brings about a significant programming paradigm shift from object orientation to service orientation so that the gap between software and business professionals can be significantly reduced.

2. BUSINESS APPLICATION INTEGRATION AND SOA

In this section, two different types of application integration are discussed: EAI and B2BAI. Then, we investigate the SOA that can be supported by current commercial web service platforms. We also introduce examples of web services to real business applications that need both EAI and B2BAI, such as retail SCM.

Depending on where integration happens, there are two different types of application integration: EAI and B2BAI [1]. Although applications in an enterprise generally tend to use the same language and platform, it is quite possible to have to contend with heterogeneous platforms and languages for various applications for EAI. Since applications belong to the same organization, explicit application interfaces and the semantics of business transactions between applications can be implicitly assumed. However, in B2BAI, this assumption does not hold. Standardized interfaces and general protocols for application and system interactions need to be explicitly specified and agreed upon by the different stakeholders.

Web services, an emerging technology, can be used for both EAI and B2BAI since the technology supports the development and use of standard interfaces for integration and general protocols for interactions among heterogeneous applications. Web services are self-describing, open components that support rapid, low-cost composition of distributed applications. Web services can communicate with other web services through a variety of Internet protocols and XML formats [5, 6, 7]. Web services are based on open standards: Simple Object Access Protocol (SOAP), Web Service Definition Language (WSDL), and Universal Description, Discovery and Integration (UDDI).

SOA is a conceptual architecture for organizing and understanding the development and use of web services. Based on the roles of participants, there are two types of SOA: 'publish-discovery-binding' and 'publish-compose-discovery-binding' model. Figure 1 shows an SOA using the "publish-discovery-binding" model. A web service provider implements web services for addressing business logics. WSDL is used to describe standard interfaces of the available services. The web services are "published" to a web service registry – a UDDI. A web service registry contains information such as the service location, the provider, etc. A service broker (SB) uses UDDI specifications to define the registry service for web services. A service requestor (SR) or consumer (SC) can "discover" web services from the registry. Based upon the discovered service information, the SC's application can be "bound" to the services at execution time through SOAP. SOAP is used to transfer the message between heterogeneous applications. SOAP is a lightweight XML-based messaging protocol that is used for requesting/responding to messages over a network using existing transport protocols such as HyperText Transfer Protocol (HTTP), Simple Mail Transfer Protocol (SMTP), or Multipurpose Internet Mail Extensions (MIME). The HTTP protocol is most popular as the transport protocol of SOAP.

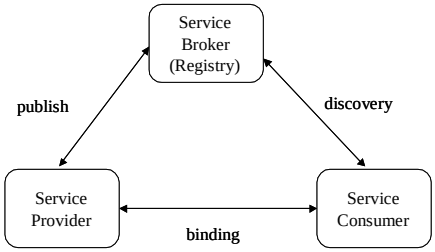


Figure 1. A SOA using the 'publish-discovery-binding' model

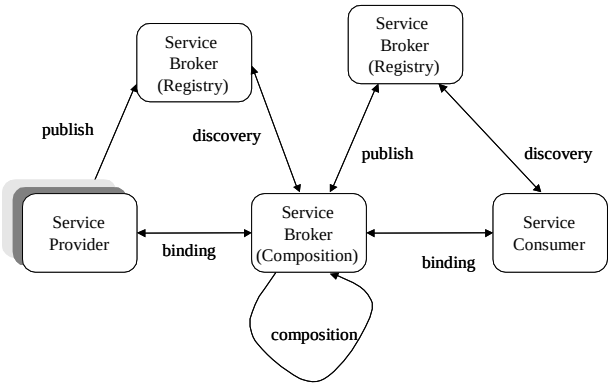


Figure 2. A SOA using the 'publish- compose-discovery-binding' model

A set of composite services called a business process can be represented in Business Process Execution Language (BPEL). A composite

service in BPEL is executed on a business process execution engine such as Oracle BPEL engine [8]. Figure 2 shows a SOA using the 'publish-compose-discovery-binding' model. A service consumer discovers a required service and invokes the discovered service. A service provider develops a service in a programming language and publishes its service specification, UDDI specification, to a service registry. A service broker composes a set of services as a new service and administers the registry. Since protocols and specifications of the current SOA are still evolving, new SOAs may be proposed in the near future.

3. FIVE NATURAL CAPABILITIES OF SOC

The inherent capabilities of the SOC paradigm are useful for analyzing the effects of the SOC paradigm on the development of Distributed BISs. Due to the heterogeneity of platforms, networks, and languages, the necessary capability is interoperability. Also, in terms of software engineering, the reusability capability has been emphasized for reduction in the cost of software maintenance and evolution. In addition, there are adaptability, interdependability, and agility. Although the current SOC related technologies need to be enhanced, we argue that the following five capabilities are naturally provided for distributed information systems by the SOC paradigm using the SOA: interoperability, reusability, adaptability, interdependability, and agility.

Interoperability is "the ability of information systems to operate in conjunction with each other encompassing communication protocols, hardware software, application, and data compatibility layers [9]." Interoperability is the key property of any distributed computing paradigm where there is heterogeneity between the implemented distributed systems. Since software systems are typically developed using different platforms and computer languages, a distributed computing paradigm that can integrate applications of heterogeneous software systems is needed. The SOC paradigm, using the SOA, has this capability by providing a standard interaction protocol SOAP, a standard interface description WSDL, and a standard XML based external data representation.

Reusability is defined as "the ability to design software modules in a manner so that they can be used again and again in different systems without significant modification [10]." Reusability enables software developers to reduce the development time by reusing existing code. In developing distributed information systems, software developers have even attempted to reuse remote methods. Reusability is inherent to the SOC paradigm using the SOA since the SOC paradigm is an evolution of a distributed computing paradigm that supports reusability through object orientation concepts. In addition, more reusability is gained through collaboration and/or composition of existing web services.

Adaptability is "the ability to change or be changed to fit changed circumstances [11]." Adaptability enables software developers to design new distributed systems with minimal effort when circumstances change. Adaptability is inherent to the SOC paradigm using the SOA since the SOC paradigm provides service related architectural and design patterns and service registries for the developers. By understanding the service patterns before developing a new service based distributed system, the developers can adjust their architecture and designs. Also, since specification of web services can be published to a UDDI registry and software developers can access the registry, software developers can adjust the selection of web services for their new projects based on current demands and requirements.

We define interdependability as the capability to minimize coupling and maximize cohesion. "Coupling is the level of interdependency between a method and the environment, (other methods, objects and classes), while cohesion is the level of uniformity of the method goal [12]." The level of interdependability is very useful for team collaboration since it is important to know who will take which roles among the service requestor, broker, and provider. A web service requestor implements only the interface and web service invocation. If a user interface needs a business function, the business function is invoked as a method in the form of a web service. A web service provider implements the service component. The web service is a (object-oriented) component with an interface. A web service broker handles the advertisement of the web service. Since role specific views for service requestor, broker, and provider are clearly defined in the development of a service-oriented distributed system, interdependability is inherent to the SOC paradigm using the basic SOA.

Agility is "the ability to respond quickly to short-term changes in demand or supply and handle external disruptions smoothly [13]." Agility enables an organization to respond quickly to demands. Agility is inherent to the SOC paradigm using the "compose-publish-discovery-binding" model since the SOC paradigm enables service composition for the developers. Instead of developing new software components, software developers can compose existing web services and invoke the composite web service to respond the demands. A web service composition can be model-driven with lightweight coding, which we call Service-Oriented Programming (SOP). SOP can enable software developers to deliver product releases in much shorter periods of time compared to the traditional Object-Oriented Programming (OOP).

4. *CASE STUDIES*

4.1 SERVICE-ORIENTED RETAIL SCM SYSTEM

A Retail SCM was developed as a set of business processes that coordinates and integrates the flow of goods from supplier to retailer to customer in order to reduce inventory and expedite prompt and precise services [4]. However, a BIS in a retail store is usually an isolated application and/or in most cases tightly coupled with its main office's information systems. Moreover, the information system cannot easily exchange data with information systems of other suppliers.

There is considerable interest in the development of web-based SCM applications. The Web Services Interoperability (WS-I) organization used SCM as a B2B example [14, 15]. WS-I has links on SCM architecture and use case models. Alonso et al. used SCM examples to explain web services coordination protocol and composition [1]. In addition, there are examples of applications of web services to real business problems. Schmerken describes the application of SOA to B2B application integration in Finance [16]. Maresca et al [17] applied SOA to the design of the Cargo Community System for the transport operators located in the Port of Genoa. Goepfert and Whalen describe three case studies of web service application in real business such as access management module, provisioning module, sales force automation module, chat rooms, message boards, instant messenger, web polls, etc [18].

In order to analyze how the emerging SOC paradigm will influence the development of DBISs, the transactions between business information systems of retailers and suppliers are first considered. For purpose of simplification, only one retailer and one supplier are considered in this example. The number of participants can be increased if necessary. The transactions are described with the following four use cases:

- Use Case 1 - Retailer's purchase order entry**:** The retailer's purchasing department enters the purchasing order that will be stored into the retailer's database. The retailer sends the purchasing order form to the supplier's sales department.
- Use Case 2 - Supplier's invoice entry: The supplier's sales department issues an invoice statement based upon the retailer's purchase order and its shipping information. That invoice statement is then sent to the retailer's inventory department.
- Use Case 3 - Supplier's purchasing order retrieval from retailer: The supplier's sales department looks over the purchase order and processes the order. Then the supplier's shipping department fulfills the order and updates the purchasing order to show the number of items shipped.
- Use Case 4 - Retailer's invoice retrieval from supplier: The retailer looks over the invoice to confirm the quantity ordered, quantity shipped and the correct prices. Then, the invoice can be closed. If not, then contact the supplier's with the discrepancies.

The SOA is used for implementing the use cases. The main advantage of using the SOA is the fact that the current commercial web services platforms such as Microsoft .NET or IBM Websphere support it. In the SOA, a synchronous interaction protocol (SOAP/HTTP) is used between a service requestor and a service provider. For B2BAI, web services are published onto an external public web service registry that can be accessed by any organizations that participate in business transactions. To exchange data, a domain specific XML format is employed. (In this paper, IxRetail is used for data interchange [19].) Since we can assume that a service consumer knows where web services are located for EAI, web services for EAI are not published onto a service registry, contrary to the web services for B2BAI.

Based on the four use cases, the following web services are implemented:

- The RetailerWS class provided by the retailer for EAI: The retailer's web client can communicate with the retailer's database and load a list of suppliers from the retailer's database.
- The SupplierWS class provided by the supplier for EAI: The supplier can retrieve the list of Retailers and purchase order from the supplier's database. The supplier checks available new purchase order through a web interface. The web interface will upload a list of retailers from the supplier's database using a private web service called SupplierWS.
- The RetailerDB2IxRetailWS class provided by the retailer for B2BAI and published to a public service registry: The supplier can select the retailer and retrieve a list of purchase order dates from the retailer's database through a method from the public web service published by the retailer called RetailerDB2IxRetailWS.
- The SupplierDB2IxRetailWS class provided by the supplier for B2BAI and published to a public service registry: The retailer can pull the available invoice dates from the supplier's database.

Table 1 shows the use case employed in this case study, web services implemented, and the service requestors or consumers. In the use case 1, a user interface is created for the data entry in order for a retailer to create purchase orders. The purchase order system was created with Microsoft Visual Studio .NET as a Windows client application using C# as the programming language due to the heavy interaction between

the user and the application.

Table 1. The Information of Web Services

Use Case	Web Service Class Name	Web Service Requestor
1	RetailerWS	Retailer
2	SupplierWS	Supplier
3	SupplierWS RetailerDB2IxRetail WS	Supplier
4	RetailerWS SupplierDB2IxRetail WS.	Retailer

Since the purchase order application is an internal process, it uses a private web service called RetailerWS to communicate with the retailer's database. The RetailerWS is a private web service because the web service is published onto a public UDDI registry. The RetailerWS class contains methods that will add a new supplier, retrieve all supplier IDs, delete a supplier, retrieve a supplier's information, and save the purchase order information.

In the use case 2, once the shipment has been fulfilled, the supplier's sales department has to enter the quantity shipped information into the Order Fulfillment Form. This part of the supplier process uses private web services. It uses the SupplierWS class to retrieve the list of retailers and a purchase order from the supplier's database. It selects the purchase order that has already been inserted into the supplier's database. The shipped quantity information is entered into the supplier's database through the Order Fulfillment Form. The web client application retrieves the purchase order from the database and the Order Fulfillment form to update the units shipped.

The use cases 3 and 4 will discuss that the BISs interact with each other to process B2B transactions. In order to use a web service of the other BIS, the client application has to search for it from a known public web registry. For this project, the web services are published onto the Microsoft's UDDI web service registry.

For example, in the use case 4, a retailer retrieves the invoice from the supplier's database. The retailer's web application client can upload a list of suppliers from the retailer's database through the RetailerWS class. Then it pulls the available invoice dates from the supplier's database through the public web service called SupplierDB2IxRetailWS. Once the retailer selects the invoice, it will retrieve the invoice from the supplier's database and convert it into IxRetail format by a method in the SupplierDB2IxRetail WS web service. At the completion of that function, an XML file will be returned to the retailer. The process will call another web service IxRetail2RetailerDBConverterWS to convert that XML file into the retailer's database structure.

4.2 STOCK CURRENCY CONVERTER

In the Stock Currency Converter (SCC), there are investors from foreign countries who get the stock quote of a company within the United States. The SCC web application receives inputs for both a country name and a stock symbol to be quoted, and displays the quoted stock prices in the selected country's monetary unit.

Service providers publish software components to a service registry as Web Services, exposing their interface so that a service consumer can access services exposed by the service provider. A service registry, in most cases, will be present to aid the service consumer in discovering services published by the service providers. The interactions between SC, SB, and SP are modeled by using a Domain Specific Language (DSL) [20, 21, 22] or Unified Modeling Language (UML) that can be converted into codes.

The service consumer is first interested in the type of user interface. Based upon the user requirement, a web application is selected for 24*7 service regardless of location. Also, the service consumer is interested in discovering and invoking necessary services from a service broker(s). The discovery can be done manually or programmatically, with or without intelligence, or implicitly/explicitly. Since the discovery is done manually and implicitly in this case, there is no interaction between service consumer and service broker for the discovery. The assumption is that the service consumer knows what services are available and where they are located at in advance.

An executable model, compared to expressive visual models in UML, can represent the diagram by using Microsoft Visual Studio 2005: The web application 'SCCWebApplication' invokes a web service called 'CurrencyStock'. Its diagram is described in Microsoft Domain Specific Language (DSL) and generates an executable framework in C# and ASP .NET.

The service broker is interested in service discovery, matchmaking, and composition. By using a BPEL composition method [23], the service broker composes a new composite service by using two services 'getRate' and 'getQuote'. The service broker discovered two web services 'Currency Exchange Rate' and 'Delayed Stock Quote' through the 'xmethods.com' service registry. This composition is model-driven, which is shown in Figure 3. By using Oracle's BPEL designer and BPEL as the domain specific language for service composition modeling, two services are composed into a new service which is published as a new web service.

Service-Oriented Programming is applied here because we are using the concept of SOA to integrate software components in order to create a new software component. In the stock quote

currency conversion example, there are service providers called 'xmethods.com' that first publish services to a services directory. Then, we discover services that are useful to us and call those services directly, and the results are combined to form a new result useful to a user. This new application using BPEL becomes another service that others might invoke directly or be composed with another service(s).

The interactions between all the participating partners are described using the syntax of BPEL, and then deployed and executed by the BEPL engine. The Oracle BPEL Designer and Process Manager are tools that provide users with a development environment to create, deploy, and execute applications using BPEL. The BPEL Designer provides a visual modeling approach for developers to create new applications. This approach is a form of Model Driven Architecture, where a system is designed with visual techniques and a code will be generated according to the visual diagrams. Here the visual diagrams are converted to BPEL syntax. A deployment tool is also provided through the BPEL Designer. The end product will then be executed by the BPEL engine, which will serve BPEL services. All of the tools and techniques to help build Service Oriented Programs will be part of the software factory.

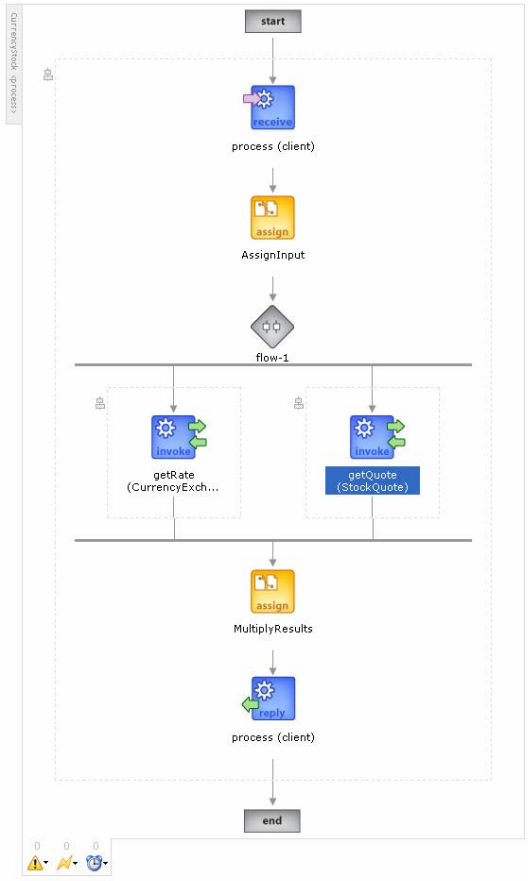


Figure 3. A Model-Driven Service Composition using Oracle Domain Specific Language for a Service Broker

5. *ANALYZING THE FIVE INHERENT CAPABILITIES*

Based upon the case study results, we examine how the SOC paradigm provides the five inherent capabilities: interoperability, reusability, adaptability, interdependability, and agility, which are necessary for successful development of distributed, enterprise-level BISs.

First of all, this Case Study 1 demonstrates that the SOC paradigm provides interoperability for integrating heterogeneous applications. For example, the use case 3 in Table 1 employs two different types of web service classes: SupplierWS and RetailerDB2IxRetailWS. The SupplierWS class was provided by the supplier and used for internal enterprise integration. The RetailerDB2IxRetailWS class was provided by the retailer and used for integrating the supplier's and the retailer's applications. Therefore, the interoperability allows the software engineers to attempt to integrate heterogeneous systems for either EAI or B2BAI.

Secondly, this Case Study 1 shows that the SOC paradigm provides reusability for software developers. In Table 1, the use case 3 employs the SupplierWS class that was implemented by the supplier for the use case 2. The use case 4 uses the RetailerWS class that was implemented by the retailer for the use case 1. Therefore, the reusability supports the software developers to reuse existing web services.

Thirdly, Case Study 1 shows that the SOC paradigm provides (minimal) adaptability for software developers. Since the architectural pattern SOA in Figure 1 is also used for implementing BISs of both the retailer and the supplier, a developer can adjust the SOA for the supplier's BIS if the developer understands the architecture of the retailer's BIS. In addition, since specifications of web services for B2BAI are published to a UDDI registry in Figure 2, software developers can access the registry and select web services for their new projects easily. The use of web service registries encourages software developers to adapt their distributed systems to the changed circumstances by allowing the discovery of existing web services. Therefore, adaptability encourages the software developers to use the architectural pattern and existing services for developing a new BIS seamlessly. If the semantics of services are provided and automatic discovery of services on demand is fully provided in the near future, the adaptability can be significantly improved.

Fourthly, Case Study 1 shows that the SOC paradigm provides interdependability for software developers. There are two different types of user interface: windows client application and web client application. Both of them use the same web service class RetailerWS. If the retailer requests a web service of the supplier, the retailer does not have to know the

implementation of the service. The retailer can only focus on his or her own role as a service requestor. If there are any changes to the implementation of the business logic and its service description remains constant, the service broker is independent upon the changes in terms of providing the services of web service discovery. Therefore, the interdependability of service requestors, brokers, and providers brings more clear roles to a software development team. It also brings a more precise boundary between presentation logic through user interfaces and business logic through web services.

Lastly, Case Study 2, shown in Figure 3, depicts that the SOC paradigm provides agility for software developers. Since the existing services can be easily composed according to the given demand and its model-driven service-oriented programming enables the service broker to develop his or her service with minimal programming efforts, the service broker is able to respond quickly to short-term changes in demand. If the automatic composition of services at execution time is fully provided in the near future, the agility can be significantly improved.

6. *CONCLUSIONS*

Although the current SOC related technologies need to be enhanced and are being evolved, we argue that the current SOC paradigm using the SOA is suitable for developing distributed business information systems because the following five capabilities are naturally provided: interoperability, reusability, adaptability, interdependability, and agility. Since this SOC paradigm provides these inherent capabilities for software developers, the SOC paradigm is very effective for not only EAI within an organization but also for B2BAI between organizations. Although other distributed object-oriented computing paradigms such as CORBA or RMI may support interoperability and reusability, they cannot easily obtain the other three capabilities such as adaptability, interdependability, and agility. The employed case studies demonstrate that those features can be supported by the premature SOC technologies currently used. The common conceptual building block with the five inherent capabilities, service, will enhance the communication between software and business professionals and proliferate the concepts of software as a service and business process as a service amongst both professional communities.

REFERENCES

[1] Alonso, G., Casati, F., Kuno, H., and Machiraju, V. Web Services Concepts, Architectures and Applications. *Springer Verlag*. 2004.

[2] Zimmermann, O., Tomlinson, M., and Peuser, S. Perspectives on Web Services: Applying SOAP, WSDL, and UDDI to Real-World Projects. *Springer Verlag*. 2004.

[3] Ferris, C. and Farrell, J. What are Web services? *Communications of the ACM*. Vol. 46, No. 6. pp. 31. June 2003.

[4] Kumar, K. Technology for Supporting Supply Chain Management. *Communications of the ACM*. Vol. 44, No. 6. pp. 58-61. June 2001

[5] Turner, M., Budgen, D., and Brereton, P. Turning Software into a Service. *IEEE Computer*. Vol. 36, No. 10. pp. 38-44. October 2003.

[6] Chung, J., Lin, K., and Mathieu, R. Web Services Computing: Advancing Software Interoperability. *IEEE Computer*. Vol. 36, No. 10. pp. 35-37. October 2003.

[7] Papazoglou, M., and Georgakopoulos, D. Service-Oriented Computing. *Communication of the ACM*. Vol. 46, No. 10. pp. 25-28. October 2003.

[8] Oracle (2005). BPEL Process Manager. [Online] Available: http://www.oracle.com/technology/products/ias/bpel/index.html

[9] Definitions of Interoperability (2005) [Online] Available: http://www.ichnet.org/glossary.htm

[10] Definition of Reusability. (2005) [Online] Available: http://myphliputil.pearsoncmg.com/student/bp_hoffer_modernsad_3/glossary.html

[11] Definition of Adaptability (2005) [Online] Available: http://www.cogsci.princeton.edu/cgi-bin/webwn

[12] Definition of Coupling and Cohesion (2005) [Online] Available: http://www.ugolandini.net/AccoppiamentoCoesioneE.html

[13] Agility. From Wikipedia (2005) [Online] Available: http://en.wikipedia.org/wiki/Agility

[14] Sample Application Supply Chain Management Architecture. {2005) [Online] Available: http://www.wsi.org/SampleApplications/SupplyChainManagement/2002-11/SCMArchitecture-0-11-WGD.pdf

[15] Supply Chain Management Use Case Model (2005) [Online] Available: http://www.wsi.org/SampleApplications/SupplyChainManagement/2002-11/SCMUseCases-0.18-WGD.pdf

[16] Schmerken, I. (2005, September) Can Web Services Live Up to the Hype? *Wall Street & Technology Online*. [Online] Available: http://www.wallstreetandtech.com/story/mag/showArticle.jhtml?articleID=14702913&pgno=1

[17] Maresca, M., Baglietto, P., and Zingirian, N. (2002) Deployment Experience of Service Oriented Architecture in the Business Community of the Port of Genoa: Lessons Learned. *XML Europe 2002 – Conference*. Barcelona, Spain. [Online] Available: http://www.idealliance.org/papers/xmle02/dx_xmle02/html/abstract/02-03-03.html

[18] Goepfert, J. and Whalen, M (2002, August) An Evolutionary View of Software as a Service. *IDC*. [Online] Available: http://itpapers.zdnet.com/abstract.aspx?&tag=tu.web.ont.dir1&scid=90&docid=46783

[19] *The Association for Retail Technology Standards (ARTS) of the National Retail Federation*. [Online] Available: http://www.nrf-arts.org

[20] Greenfield J. and Short, K. Software Factories: Assembling Applications with Patterns, Models, Frameworks and Tools. *Object Oriented Programming Systems Languages and Applications (OOPSLA) 2003*. Anaheim, CA. pp. 16-27. October 26-30, 2003

[21] Greenfield, J., Short, K., Cook, S., and Kent, S. Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools. *Wiley Publishing, Inc*. Indianapolis, IN. 2004.

[22] Domain-specific programming language from Wikipedia. (2005). [Online] Available: http://en.wikipedia.org/wiki/Domain-specific_programming_language.

[23] Milanovic, N., and Malek, M. Current Solutions for Web Service Composition. *IEEE Internet Computing*, Vol. 16. pp. 51-59. November/December 2004.

What is the Effect of Product Attributes on Public-Key Infrastructure adoption?

Mattila, Anssi; and Mattila, Minna

Abstract — ***the arguments for adopting Public-Key Infrastructure (PKI) are strong and yet PKI products have suffered from relatively low adoption rates. The results presented in this paper help managers to better understand the effect of product attributes on PKI adoption. In this paper product attributes are defined as features, quality and design. When implementing the technological developments, the knowledge on security products, organizational IT assets and organizational culture becomes of an outmost importance. Therefore, this paper further makes a significant methodological contribution when combining marketing instruments with product development processes to better capture the dynamic nature of the innovation adoption factors.***

Index Terms — innovation adoption, product attributes, technology, PKI

1. Introduction

For the time being companies are more and more concerned about the security of their information technology assets courtesy of the connection to the Internet. They invest money in various security products, and hope for the technology to provide the security. The repertoire of security products is and has been wide including both software and hardware. Firewalls, monitoring devices/software, security management tools etc. can easily comprise a myriad of non-compatible systems that require money, attention and endless patience. This has been realized and systems that can provide security for more and more complex network environments have been under way for some time. Public-Key Infrastructures (PKI) belong to this category of products and its prospects are interesting. Whether they will be fulfilled will be seen in the future.

So far PKI has been mostly applied to e-commerce and secure communications. E.g. the Bank of Canada Public-Key Infrastructure has been up and running since April 1999, and there are hundreds of banks around the world running their own PKIs [1]. Some businesses have been merging their PKI with Virtual Private Network (VPN) in order to safely open their IT resources to the outside world [6]. PKI has been claimed to be complex, difficult to install/implement, inadequately standardized and so forth. It seems that all the implemented PKIs are more or less pilot projects, so there is no long-term research and experience on the subject, except trying to achieve stable and sound standards.

The four most influencing factors affecting the success of a new product development process are high-quality new product process, a defined new product strategy for a business unit, adequate resources of people and money, and R&D spending for new product development [12].

When implementing the technological developments, the knowledge on security products, organizational IT assets and organizational culture becomes of an outmost importance. Therefore, this paper further makes a significant methodological contribution when combining marketing instruments with product development processes to better capture the dynamic nature of the innovation adoption factors.

2. PKI Product Attributes

Quality:

There are several kinds of quality: quality of inputs, processes and outputs. There is no simple uni-dimensional measure of quality. In our research the quality of PKI systems should be evaluated. A PKI system can be understood as an output, so what means do we have to evaluate the quality of a PKI product. One way to evaluate quality, is to measure product characteristics, which can be set an exact value, e.g. height, weight, volume, time etc. In other words variables can be formed from these characteristics to be able to make measurements and give exact values, which can be evaluated afterwards. The benefit of this is to be able to notice how much the product misses quality specifications if it doesn't pass the check. However, this requires special skills from the employees, special equipment, time and effort.

Features:

The main goal cannot be to include as many as possible features into software. If features are meaningless to users it is the same if there are n or n+10 features. Therefore the number of features is not the most significant point. From the vendor, or from the producer of software point of view including features, which are not in the competitors' products, can be a competitive edge in case buyers value those features. Additional features cannot still increase the price

of the product unless the increase is well founded and understood by the customers.

Design:

Another way to add customer value is through product design. Design is a broader concept than style. Design goes to the very heart of the product. A sensational style may grab attention and produce pleasing aesthetics, but does not necessarily make the product perform better. Good design contributes to a product's usefulness as well as to its looks. For example, a properly designed user interface in the case of a software product offers a significant tool for positioning. That investment in design pays off has been recognized by global companies, which have embraced design. Good design can attract attention, improve product performance, cut production costs and give the product a strong competitive advantage in the target market.

3. NEW TECHNOLOGY ADOPTION REASONS

Some kind of competitive advantage should be gained by adopting a new technology. Shapira et al. 1996 introduces several ways to ease and accelerate technology adoption process. Some of the proposals are hard to implement for small companies due to the lack of needed personnel or funds [8].

An increase in the value of a product or a reduction in the costs of bringing the product to market means competitive advantage. Using automated equipment and computer control, which affects the quality of a product, can minimize human errors. [4]

Information security technology, like a PKI, can be seen as an increase in the value of certain products. Product support, e.g. of paper machine, can utilize a PKI system in a way, that customers data remains secret all the time, as well as any communication with the customer. In inter-organizational communication it is not uncommon, that organizations make demands on information security level of peer organization, and in here a PKI system is of great assistance, though security could be achieved by other means or by other technology too.

Au et al. (2000) conducted a study on attitudes towards adoption of new technology. The results indicated that the adoption process is related in several beliefs such as perceived difficulty, adoptive experiences, suppliers' commitment to the firm, perceived benefits, compatibility and enhanced value [9].

Deployment of a PKI system could be seen as improvement in quality of the product in the eyes of a customer. Customer can realize that its needs are better taken care of. Today it should be minimum requirement, that all customer data, and communication, is well taken care of by deploying encryption, authentication etc., and this is transparent when PKI technology is used. However, is it reasonable to make purchase decision of a PKI system based on estimation of a possible increase in the quality of a product, e.g. paper machine?

PKI technology doesn't belong to production or manufacturing technologies, instead it is a technology that can create more value to a product or it supports these aforementioned technologies. Therefore the downsides (possible financial risks, jeopardizing the market position etc. [4]) of deploying a technology among the first won't be much of a threat if PKI fails: it quite unlikely affects the manufacturing process.

In software business it is difficult to make masses buy. IT decision makers in companies tend to buy front-runner products. Once users learn how to use a certain product, become very familiar with it and so forth, then they are not that eager to switch to another product. Being a first-mover and trying to "set a standard" e.g. by marketing extensively, is a key element when trying to succeed in software business [11].

Evaluation of a new technology to be adopted should begin with financially analyzing all quantifiable factors that can be set a monetary value. In addition to this all qualitative and intangible factors are analyzed. Then results of these two analyses are summed up to facilitate the evaluation of all involved risks related to uncertain costs and revenues. [4]

4. A COGNITIVE MODEL OF THE PKI ADOPTION

Based on theoretical literature review the research framework (Figure 1) was formed, in which the relations of corporation's macro- and microenvironment, and product development are presented. In software business high-speed changes in technology are the normal course of business. Macro factors are presented to accomplish the overview of software ventures' environment.

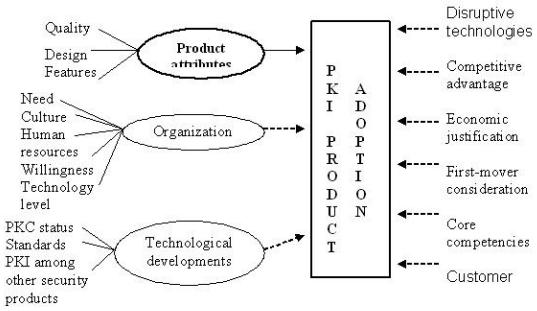


Figure 1: A cognitive model of the PKI adoption

Altogether the factors presented in figure 1, contribute toward success (or failure) to retain business customers in terms of creating added value. Some of the factors have more direct effect on the PKI adoption than others as can be seen from figure 1. In this study, however, the main focus lies within the two inner boxes; in other words the relation between the product attributes and PKI product adoption.

5. Data Collection / Methodology

The preliminary study was conducted to become familiar with empirical study's research field. Five test interviews were made with the key personnel of the case firms. In these preliminary open interviews the list of important issues was not used, and the interviews were recorded. The interviewed persons wanted to have the possibility to communicate freely about all issues they felt relevant to this research. During these interviews, the organizational views of future directions of technology adoption within an organization were clarified.

A qualitative, interpretive study approach was used to evaluate product attributes in the technology adoption context. The qualitative study approach is considered appropriate when little is known about the phenomenon under investigation, the concepts are immature due to the lack of theory and previous research and a need exists to explore and describe the phenomena [10].

Two organizations and three PKI products were included in the study. Metso Corporation is a globally leading supplier of processes, machinery and systems for the pulp and paper industry and a foremost expert in the key technologies of this sector. Sonera Ltd. is the leading mobile communications operator in Finland. Keon product family is a common product of RSA Data Security and Security Dynamics. Keon provides encryption of data, encryption of telecommunication, end-user authentication and encryption of application server data. Entrust/PKI is security application, which is intended for enterprise use. Its main goals are to enhance e-commerce security, encryption of files and e-mail and being able to digitally sign them. Certicom is a US located company specialized in security of wireless environments. Founders of the company are also related to invention of elliptic curve cryptography (ECC), which has partly made it possible to extend PKI to wireless environments.

The actual research was conducted during 1999-2001. In the two-year project we studied Public-Key Infrastructures and their applicability to industry and companies in general. This stage of research provided concrete data about the product attributes and their functions. After defining our requirements for the test PKI system we checked the supply and ordered the best fitting one. Our requirements, Single Sign-On -feature (SSO) and smart card authentication, made the supply scarce. We wanted to use SSO in an automation system, not only in logging to various web pages. Due to complexity and extent of PKI systems we required product support to be nearby.

Our first PKI system to be tested was Keon. Developing an agent for an automation system helped testing its SSO capabilities and smart cards features. SSO password saving problem (in clear) is avoided in Keon PKI by using special short lifetime certificates signed by security server. After Keon testing, we checked PKI-markets again and we chose Entrust PKI system. Entrust had a good reputation, possibility to use cross-certificates and generally seemed to be a versatile product. This constituted a good point of comparison for Keon.

At the time we finished testing Entrust, PKI wireless PKI product producers had tried to establish a firm foothold on the PKI market. Therefore we chose to explore wireless possibilities of PKI, like extending PKI to PDAs. Several companies used PDAs in different ways, though the applications were still not versatile and beneficial so that PDAs could be used in an efficient way.

The group discussions were used to further understand the observations, and especially actual product attributes' effect on the perceived willingness to adopt a PKI product within an organization. The first persons to participate in the group discussion were selected by using a purposive sampling to interview the personnel involved in business development. These experts further named some people working in the same area of operation, which could provide information from the field. This snowball method, in which the key person(s) name the next persons participating in the research, is suitable in situations, in which the other important people for the research are otherwise hard to identify [5]. Altogether 15 in-depth interviews were conducted among academics, managers from both seller and buyer side, and software development personal. The discussion themes covered topics from the theoretical framework presented in figure 1 including experiences from PKI product testing, development, and usage, organizational abilities to benefit from PKI products, and different kinds of indirect effects.

In qualitative research the aim is often to concentrate on small amount of cases and analyse them thoroughly; criteria of scientific research, when using qualitative methods, is not the quantity of the data but the quality of it. The amount of data or the number of persons interviewed does not have generally significant importance when evaluating the success of the study – the aim is not to make statistical generalizations but describe a phenomenon or understanding of some function [2].

6. Results

Quality:

Fast operation can be understood as part of the quality of a PKI system (Figure 2). It is important that system operates rightly, but it has to operate in due time. In software products the speed of operations of software greatly depends on hardware on which it is running, but high

quality products can use underlying hardware better, thus being faster. Slow operation of software products might hinder the usage, and people might try to avoid the usage as they can.

The lack of support material and bad product support hampers e.g. ease of operation and repair. During the installation and piloting phase support material, as well as product support, is of great importance. These are the phases when people try to get the PKI environment function correctly and it is not uncommon that difficulties appear. As seen from the grid support material is highly valued, but not seen extensive and profound enough. Product support is seen better carried out.

Compatibility of software products is something to expect. Among the PKI products compatibility problems have risen and software houses are doing something to fix the problem. Large software products require compatibility with several hardware products and software products as well. Incompatibility of software products is the origin of several difficulties e.g. in installation, management, ease of operation and repair. According to definition ease of operation is part of quality, and compatibility is something that really facilitates ease of operation.

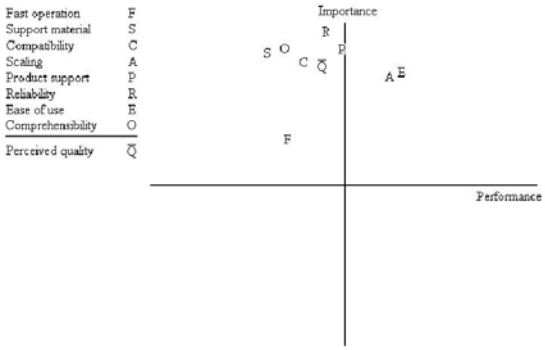


Figure 2: Perceived quality of the PKI products

Large number of users of PKI system makes scaling of PKIs highly significant. This comes up especially during the installation of PKI and PKI clients. The clients are installed on end users' PCs, which can be troublesome if scaling has not been taken properly into consideration (ease of operation).

Reliability and ease of use comes straight from the definition in the beginning. Comprehensibility has a link with reliability and ease of use. When you understand what you do (comprehensibility) and this requires only slight effort, and still everything (PKI, PKI services, etc.) works fine, then it feels that you can rely on PKI and using its services is no burden. Although some of the attributes mentioned above can be measured objectively, quality should be measured in terms of buyers' perceptions [7 kotler].

Features:

When looking at the performance-importance grid (Figure 3) elements in PKI, which could be a competitive edge as marketing PKI products, should be the ones that are in the leftmost upper corner. However, a PKI system is something that customers are rarely familiar with until they buy it. Therefore it is difficult for the customers to know which things in a PKI should be valued. Are we looking for a PKI system, which comes with very good support material, everything in it is easy to comprehend, it is very compatible and even highly reliable? As features of a PKI system was thought things with which users are in touch, and/or they can quite easily form an opinion.

Speed of operations of a PKI system and its clients is very tangible for end users. When a user logs in a system, the time it takes for the system to respond is easily noticed, and the longer it takes, the more it irritates. During log in periods the calculations different servers in PKIs have to go through are quite similar regardless of PKI product used. Therefore the time it takes to log in a system is quite the same.

Compatibility of different software products becomes evident at least during installation. However, this affects end users considerably less that people managing PKI. Similarly management (M) of the PKI system and GUIs don't concern end-users as much. Management and GUIs are features, which are noticeable for management people. Well implemented GUIs and management assist greatly in complicated system like a PKI.

Smart cards and tokens are mostly used during log in sessions. Smart cards and tokens are not an essential part of all PKIs. When they are deployed in a PKI both end-users and management people come into contact with them.

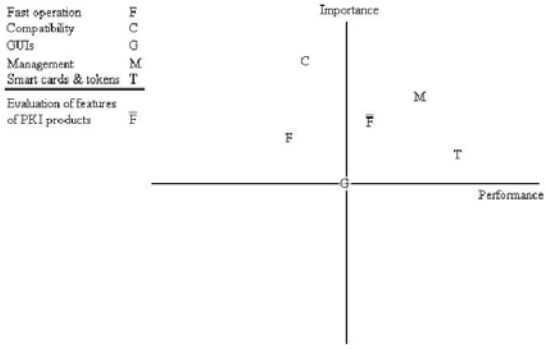


Figure 3: Features of a PKI system

Design:

Good design contributes to a product's usefulness as well as to its looks. For example, a properly designed user interface in the case of a software product offers a significant tool for positioning. Good design can attract attention, improve product performance, cut production costs and give the product a strong competitive advantage in the target market. Figure 4 represents the quality factor of a PKI.

Fast operation of software products depends greatly on hardware on which it is running. Errors

in design of software can affect the speed of operations. However formal methods and libraries among others are deployed nowadays, which minimizes the chance of faults in design, and enhances the reliability of software as well.

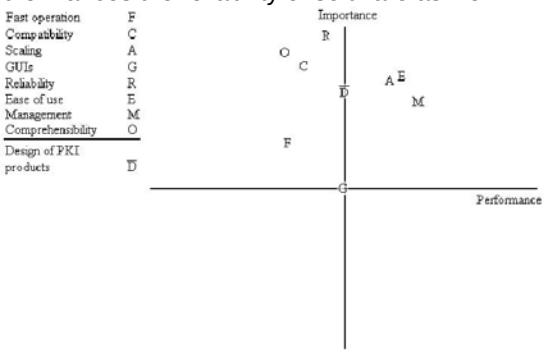


Figure 4: Design of the PKI products

Scaling of PKIs has to be considered carefully in the design. Large user base, client software on end users' PCs and so forth can make management personnel's work difficult unless scaling is well designed and implemented.

Different GUIs are a tool for the management personnel of a PKI. The GUIs (which has something to do with PKI and PKI services) that end users use don't require so much from the design. However, it is obvious that good design of the GUIs, which are used by the management personnel, facilitates ease of use and will assist in management of the PKI in general.

Comprehensibility is important for security. Especially the end users, who might not be very familiar with PKIs and e.g. certificates, should be taken into consideration. As designing GUIs and education (be it a guide book or education in general) for end users it might enhance the security if the end users would understand e.g. what all the messages are about.

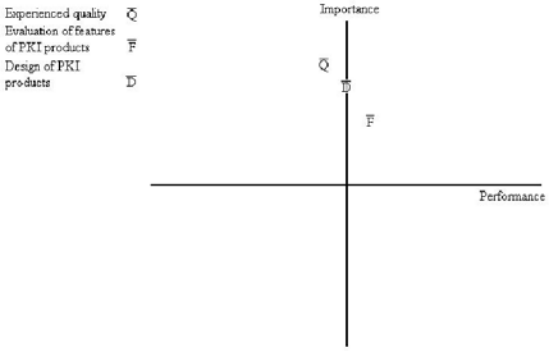


Figure 5: Summary of the product attributes

SWOT analysis of corporation PKI adoption:

Strengths of PKI adoption are closely related to asymmetric encryption and benefits created by it. A PKI is strongly based on security services, which are implemented using asymmetric cryptography. Usage of smart cards to carry personal certificates emphasizes strong authentication. However, the certainty of authentication process is comparable to identification procedure of CA issuing personal certificates. If you can't trust identification of "subjects" done by a unknown CA, how much can you trust the bondage between the Public-Key on a certificate and the real person behind the "subject-field" contained in a certificate? In here it is not that clear where to put strong authentication: it is obviously strength, but it can turn out to be a weakness.

Strong authentication can be applied to remote users as well. However, strong authentication of remote users is not possible only in PKI environment, but it comes in the package. Strong encryption possibility during remote connections is clearly a big plus for companies. Along the way from Brazil to Finland there are several places to eavesdrop telecommunication traffic. The best way to guarantee safety of your data is to take care of it by yourself with tools that can provide enough protection.

If a corporation really can trust the CA, which issues certificates, a PKI can be very useful in restricting access to corporate IT assets. Unfortunately these assets can too often include only http-servers, but there is always hope that situation will become better. Some PKI products offer authentication and encryption features to other application servers, combined with SSO too, but the implementation is not always effortless.

Using PKI to limit employees' access to resources, which they really need in their work, can be very beneficial. It is not uncommon that workers have access to places they should not have. According to statistics the threat quite frequently comes from inside the company. If workers have access to "forbidden" places, they might be teased to try to do something bad. By creating a well limited working area a corporation sends a clear message and makes the abuse of its resources more difficult (Figure 6).

Strenghts -strong encryption and authentication -remote access with strong authentication & encryption -limitation of access to corporations' IT assets -increase in security level	Weaknesses -so far no global benefit -strong authentication and encryption not that easy to implement to other that http-servers -compatibility -certificate problems (field usage) -expenses -implementation a big effort -effect on end-user
Opportunities -better service for customers -connections with strong authentication and encryption abroad -CA implementation	Threats -foreign/unknown CAs -"globalization" of PKI won't come true

Figure 6: SWOT analysis of corporation PKI adoption

SWOT-analysis describes PKI's benefits and disadvantages. Weaknesses of PKI include e.g.

the lack of confidence in certificates signed by unknown CA limits PKI usage. There is no guarantee that ”John Smith” is really the one you know is living in New York. It is impossible to now how the CA identified him when it issued him a certificate, which binds his name to this offered Public-Key. Like mentioned in the Strengths-chapter extending strong encryption to other than http-servers is not unambiguous. Some PKI products offer this possibility, but then come problems with e.g. firewalls when trying to access server from outside.

Compatibility of different PKI products is improving. Big software houses have noticed the problem and initiatives have been taken. Compatibility problem concerns also the usage of different certificate fields. Certificate standards or drafts, e.g. rfc 2459 [3], can be very extensive works covering lots of pages. Based on the open interviews, the researcher got an impression that the usage of different certificate fields might have been misunderstood by software producers, or they might have been thinking their own advantage in determining the meaning of a given field.

Implementation of a PKI affects end user at least in logging process. According to our experience the time logging takes is still quite long. Hopefully it is not required frequently during working days. The time that encryption and authentication processes takes from the processor will lose gradually importance in the future, because processing power will increase faster that complexity of mathematics in encryption/decryption (Figure 6).

Opportunities of PKI product adoption are to do with mainly with the privacy dimension they offer. Many companies have customers and business partners with whom it exchanges information, which should have better privacy. While implementing a PKI environment it is quite easy to take important people from customers and business partners into user base. So, this is a change to offer clients better service by enhancing security of their business interactions. And it is possible to offer this ”service” to foreign customers too.

Threats of PKI adoption are not yet so visible. One thing worth mentioning is of course the difficulty to trust certificates issued by unknown CA. This might be one of the weakest links in PKIs and has potential to create problems. If more trustful CAs won’t appear it might be difficult to make good use of possibilities of PKI (Figure 6).

7. CONCLUSION

PKI products are expensive but powerful tools in limiting access to own IT assets. Strong authentication, strong encryption, smart cards and SSO are details that really strengthen security when deployed correctly. However, PKI alone cannot make any company’s network secure without e.g. a firewall. Firewalls itself can be rather expensive investment due license expenses, software updates and so on. PKI is not a necessity, because encryption, authentication among others can be implemented without it. On the other hand, to implement strong authentication, strong encryption, and strong security in general, PKI can be very important part of it.

The effect of product attributes on adoption of PKI is not an unambiguous matter. According to performance-importance grid the emphasis in PKI products has been so far on features, and transition to more customer-centric product development might be desirable. Quality and design of the products are seen as more important factor than features, and maybe more intensive participation of customers in the product development would make the customers and product developers understand better each other’s, and thereby customers could be offered better, more suitable products. And maybe this would lower the adoption threshold of the customers in the future. For companies it is difficult to see a PKI system as a necessity, even if it is based on the latest security technology.

REFERENCES

[1] Clarke, J., "Internet Security: PKI in pilot mode but expectations run sky high," *Medias Transcontinental Inc*, Computing Canada, 2000, pp. 67-82.

[2] Eskola, J. and Suoranta, J., “Johdatus laadulliseen tutkimukseen,” *Vastapaino*, Tampere, Finland, 1999, pp. 23-31.

[3] Housley, R., Ford, W., Polk, W. and Solo, D., “RFC 2459: Internet X.509 Public-Key Infrastructure Certificate and CRL Profile,” http:// www.ietf.org, 24.1.1999.

[4] Krajewski, J. and Ritzman, L., “Operations Management: Strategy and Analysis,” 5th edition, *Addison-Wesley Publishing Company*, London, UK, 1999, pp. 213-332.

[5] Malhotra, N. and Birks, D., ”Marketing Research: An Applied Approach”, *Prentice Hall,* Essex, UK, 2000, pp. 342-367.

[6] Salamone, S., “TimeStep Merges PKI With VPNs,” *United Business Media*, Internetweek, 1999, pp. 22-25.

[7] Kotler, P., Armstrong, G., Saunders, J., Wong, V., “Principles of Marketing,” 3rd European Edition, *Pearson Education Limited,* Essex, UK, 2001, pp. 453-489.

[8] Shapira, P., Rosenfeld, S., “An Overview of Technology Diffusion Policies and Programs to Enhance the Technological Absorptive Capabilities of Small and Medium Enterprises,” background paper prepared for the Organization for Economic Cooperation and Development Directorate for Science, Technology and Industry, August 1996.

[9] Au, A., Enderwick, P., “A cognitive model on attitude towards technology adoption, *Emerald Group Publishing Limited,* Journal of Managerial Psychology, 1996, pp. 266-282.

[10] Creswell, J., “Research design: Qualitative and quantitative approaches,” *Sage Publications*, Thousand Oaks, UK, 1994, pp. 146-147.

[11] Hoch, D., Roeding, C., Purkert, G., Lindner, S., Müller, R., ”Secrets of software success: Management insightsw from 100 software firms around the world,” *Harvard Business School Press*, Boston, US, 1999, pp. 122-124.

[12] Cooper, R., Kleinschmidt, E., “Winning businesses in product development: The critical success factors,” *Industrial Research Institute*, New York, US, 1996, pp. 18-29.

Building a European database of mathematical e-learning modules

Bringslid, Odd; and Demlova, Marie

Abstract— The EU project dMath will build a European database of mathematical e-learning modules. The modules will be organized according to standard textbook levels and a module is divided into independent objects for downloading by a publisher to create an online course in mathematics. This database may be used as an online dictionary as well and is evolutionary in the sense that developers may create own objects for the database (reviewing by the editorial board of the database).

Existing solutions on a professional level in web-based e-learning in mathematics have important drawbacks. The provider of the course often does not have any control with design and content. This will make it difficult to make changes and adapting to local conditions. The lecturer then has to adjust his curriculum to the e-learning system and not vice versa. The database will give the publisher the opportunity to download from a wide range of reusable learning objects (RLO) and perfectly adapt the course to her needs. The system also offers a sophisticated online calculator which is unique of its kind focusing on pedagogical value.

Index Terms— Mathematical e-learning, Online Calculations, Java Server Pages, Pedagogical Value, Interdisciplinary Mathematics, Markup Languages, XML/MathML Database Client, Content Management Systems (CMS), Reusable Learning Objects (RLO), Shareable Content Objects Reference Model (SCORM), Learning Management System (LMS), Evolutionary User Controlled Development (EUCD).

1. *Introduction*

The project dMath [1] [2] in the EU Leonardo da Vinci Programme [3] is a follow up from the EU Minerva project Xmath [4]. The intention is to commercialize Xmath by using the framework and the technologies of the Xmath project, mainly connected to the web technologies XML/MathML [5] and Mathematica/webMathematica [6]. dMath will build a European database of mathematical e-learning modules using this innovative technology. A publishing system with the working title SciLAS (Scientific Learning Authoring Suite) is being developed including a XML/MathML editor as a fat client to the database repository. The repository itself will have the working title dMathArchive. The repository objects, called Reusable Learning Objects (RLO) and being a part of a module, may have a link to the Xmath Calculator [7] for online step-by-step calculations and are free to download for the publisher. The publisher may then combine different objects to create an online mathematical course without having skills in advanced web technologies.

The main idea is to get a high degree of independence compared to a ready-to-use course developed by someone else where the provider of the course may not interfere either with content nor layout. This will provide support for cooperative development (evolutionary system), reuse (RLO), exchange of courseware and creating a possibility to adapt courseware to the requirements of different authors, educators and students.

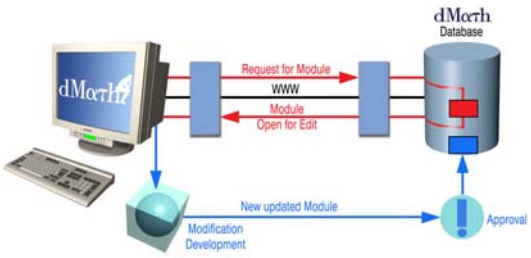


Figure 1: The dMath Database System

The partners of the dMath project are Buskerud University College (promoter) [8], Czech Technical University [9], Soft4Science [10], Pontificial University Comillas of Madrid [11], Savonia Polytechnics [12], Sogn og Fjordane University College [13], Industrial Documentation [14] and Slovak Technical University [15].

Manuscript received July 09, 2005. This work was supported in part by the EU Leonardo Programme under Grant NO/03/B/PP-165.011. First author, assoc.prof. Odd Bringslid is with the Institute of technology, Buskerud University College, Norway (odd.bringslid@hibu.no) and the co-author professor Marie Demlova is with the Czech Technical University in Prague (demlova@math.feld.cvut.cz).

2. THE DIFFERENT PARTS OF THE SYSTEM

The dMath system consists of three main parts.

A. The software framework Scientific e-learning Authoring Suite (SciLAS)

1. XML/MathML Authoring tool SciWriter with a palette of mathematical symbols, a fat client to the system [16]
2. Learning content management system (LCMS), CMS for learning content [17]
3. SOAP, Simple Object Access Protocol, published by W3C [18] (authors/editors)
4. SCORM package, Shareable Content Object Reference Model, a set of specifications published by ADL, Advanced Distributed Learning [19] (publishers, local LMS).

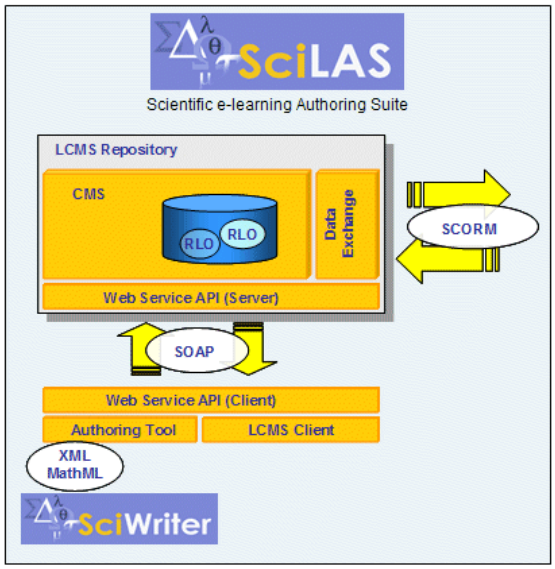


Figure 2 Architecture of the software framework

B. dMathArchive (repository of RLO's)

Modules are divided into the type object (RLO) which may be downloaded by a publisher. Characteristics of RLO's will be

1. **Modularity**, RLOs are self-contained , each object can be taken independently
2. **Reusability,** a single RLO may be used in multiple contexts for multiple purposes
3. **Interoperability**, RLOs should be independent of both the delivery media and knowledge management systems
4. **Accessibility,** a RLO should be tagged with metadata so that it can be stored and referenced in a database
5. **Granular structure,** RLOs can be combined to form a hierarchy such as Module, Course or Curiculum. The lowest level of granularity is called an asset, some media like sound, animations and JPEGs cannot define a RLO alone (cannot be tracked by a local LMS using SCORM runtime environment)

C. Xmath Calculator

Step-by-step online calculator for training and simulation. Available as an independent resource on the web [7]

SciWriter [16] is an editor for authoring scientific articles and mathematical or technical e-learning content. It is a single source multiple publishing channels system (LaTEX, XML/MathML, HTML, SCORM package, e-mail). MathML is important in the project rendering mathematics like text and not graphics. For Microsoft explorer you then will need a plug-in (free download). An option is to use HTML and publish equations as images on the web.

By using the XSLT stylesheet for MathML (universal MathML stylesheet) provided by the W3C Math Working Group [5], SciWriter documents containing MathML nodes can be viewed by the following web browsers:

1. Microsoft explorer using plug in MathPlayer [20]
2. Microsoft explorer using plug in TechExplorer [21]
3. Mozilla [22]
4. Netscape [23]

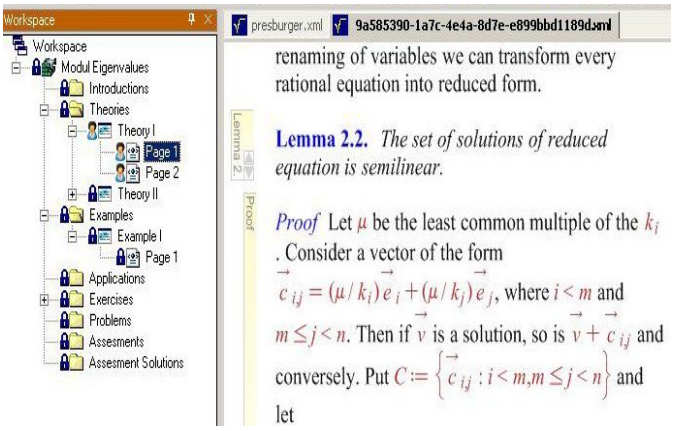


Figure 3 SciWriter, Module Eigenvalues in client workspace

The RLO's of the archive are then developed by an author/editor using the SciWriter client (SOAP). A publisher may download RLO's from the repository (SCORM) to a local learning management system (LMS) and then being quite independent in how to set up a web.-based course in mathematics and without being an expert into web technologies. The publisher may be the provider of the course for students and engineers in industrial companies needing further

education in mathematics.

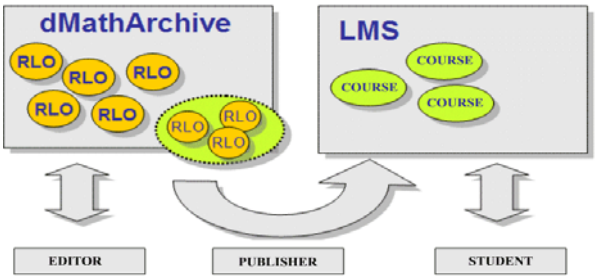


Figure 4 Using the Archive

The Xmath calculator is further developed. A sophisticated online calculator has been missing in connection with mathematical and scientific e-learning and will increase the performance/price rate of the system.

3. *Modules and objects*

A module is a collection of objects- a menu card of objects- which are grouped under a mathematical topic. Metadata assigned to the object may suggest to the publisher the order in which these objects can be organized. As an example the main topic level could be Calculus, next level Functions and then the module Continuity divided into different objects for download.

A module may contain an introduction object, theory objects, examples objects, exercises objects, assessments objects, applications objects, graphics objects, animations objects and student projects objects. Not all modules will necessary include all the different kinds of objects. The objects in a module may be downloaded independently of each other but the publisher will have access to information about the connection between the objects and how to organize them. However if a publisher only needs the animation object, say, he will then download only this object.

In the project modules in Number Theory, Calculus, Geometry, Differential Equations, Linear Algebra, Statistics and Probability, Numerical Methods, Interdisciplinary Mathematics and Fourier Series will be developed.

4. Interdisciplinary Mathematics

Some of the modules require a special attention. The project is highly Multidisciplinary and Interdisciplinary and will develop modules in interdisciplinary mathematics [24]. Our aim is not to define what interdisciplinary mathematics is (similarly as for cybernetics). Our modules called "Discovering Mathematics" do not belong to any specific field of mathematics; they are built primarily from the "Problems point View". Mathematical problems, mostly drawn from calculus, are presented in a form which enables to create, in the process of their solution, an interconnected network of new problems.

To help the user to develop ability in solving problems the network then helps to find easier problems which she can solve if she is not able find a solution of the first problem or more complicated ones of the same character.

5. *Mathematical dictionary*

The interface to the repository of modules (dMathArchive) is intended to be easy to use and a selected mathematical item may be search for to find the definition and an introduction to the item. This will work like a mathematical dictionary with a wide range of items being present. Searching for "Eigenvalue" the definition and an introduction to the topic "Eigenvalue" will show up.

6. *Evolutionary system*

In the design of the dMathArchive the concept of "Evolutionary user controlled development" (EUCD) will be important. The dMath project will implement this concept by letting users/publishers develop their own objects for upload to the database repository or modify existing modules/objects. This will do the archive "continuously growing" and give it an "agile development " with focus on changes. Of course the modules and objects have to be reviewed by the editorial board of the archive.

The same philosophy may be used for developing software packages for the Xmath Calculator.

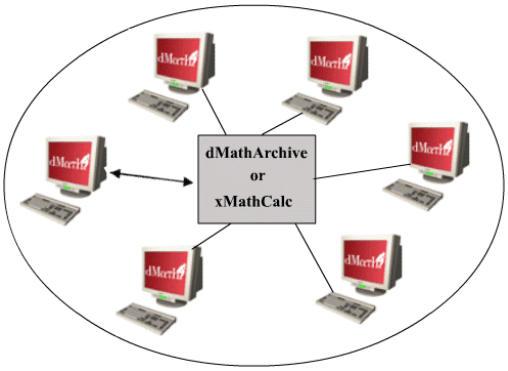


Figure 5 Evolutionary User Controlled Development

.

7. XMATH CALCULATOR

The Xmath Calculator is an online tool for mathematical simulation and training with focus on pedagogical value. The calculator gives intermediate steps for arbitrary input from the user in a wide range of applications.

An expression is broken down and analyzed by the developed software packages. The rules familiar to hand calculation are given and so are the intermediate results. By scrolling the user may view the first step, second step and so on. The user may then choose to go further by hand after the first or second step. In the same way an expression may be broken down into different levels so are the calculating steps broken down accordingly. This gives a structured output as would a professor do it on the blackboard stating the rules on each level.

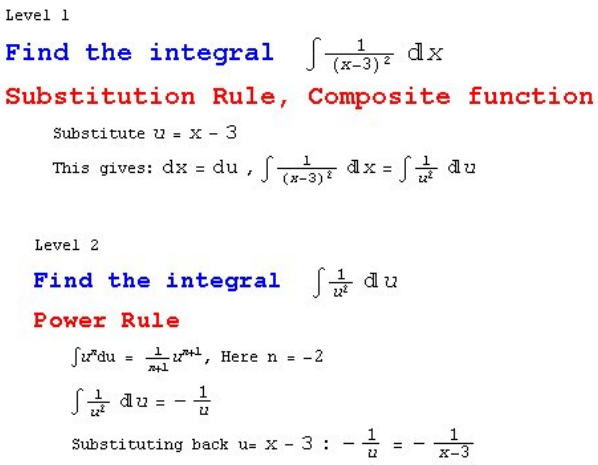


Figure 6 First two steps of integration task

The calculator is based on webMathematica technology [5]. The output may be chosen as MathMLForm giving a non-image format to be rendered in Microsoft Explorer by using proper stylesheets and a free plug-in. The input files are XML enhanced by special tags saved as *.jsp (Java Server Pages) calling the Mathematica software packages by using Mathematica commands.

webMathematica is based on two standard Java technologies: Java Servlets and JavaServerPages (jsp). Servlets are special Java programs that run in a Java-enabled web server typically called a “Servlet Container”. Both the markup languages MathML and SVG (vector graphics) may be used in a *.jsp file.

```
<math xmlns="http://www.w3.org/1998/Math/MathML">
 <mrow>
  <mi>d</mi>
  <mi>x</mi>
  <mo stretchy='false'>]</mo>
  <mo>=</mo>
 </mrow>
 <mrow>
  <mfrac>
   <mi>d</mi>
   <mrow>
    <mi>d</mi>
    <mi>x</mi>
   </mrow>
  </mfrac>
  <mo stretchy='false'>[</mo>
 </mrow>
</math>

<msp:evaluate>
If[ test===1,  MSPFormat[result1, MathMLForm],
                    MSPFormat[ Integrate[function,x], MathMLForm] ]
</msp:evaluate>
```

Figure 7 MathML and webMathematica tags in *.jsp files

8. FURTHER DEVELOPMENTS

The dMath database is now being uploaded with the modules and will be ready for use in September 2006. The RLO’s will then be tagged with semantic metadata in different languages the database then being searchable in a multilingual environment. This kind of enrichment is important to make the system easy to use in European countries.

The testing period on the full system is important. A transnational network will be established developing an innovative e-learning pedagogy in mathematics [25,26] and improving the access and willingness of teachers and trainers to use mathematical e-Tools. It will be important to improve mathematical teaching and strengthen the recruit to mathematical education.

The network will arrange seminars and conferences for European trainers, teachers and others being interested in using the dMath products: the XML/MathML word-processor, uploading to the “Orange” content management system, downloading to a local LMS using the SCORM package [19] and then tailoring a course without being an expert into advanced web technologies, and using the Xmath Calculator

9. CONCLUSION

The dMath project will build a database of Reusable Learning Objects (RLO) focusing on cooperative development, reuse, exchange of courseware and the possibility to adapt to local conditions. This will enable the teacher to build up his own professional designed course without having skills in advanced web technologies. This gives a high rate of performance/price because one do not pay for the download, there will not be too much of course design working and the

teacher may adapt the course perfectly to her needs.

The dMath project focuses on pedagogical value. This is important because one of the most frequently asked questions connected to e-learning is about its pedagogical value compared to traditional methods. The teachers are somewhat afraid of the necessary changes in their behaviour using e-learning resources in the ordinary classroom and not for saying in the virtual classroom. The role of the teacher may be more like an adviser than a lecturer.

Using e-learning resources will develop a new kind of pedagogic using innovative methods. The establishment of thematic networks aiming at collecting, analysing and disseminating innovative learning, best practices and new learning tools is then necessary. Also the exchange and dissemination of innovative methods that improve access, readiness and willingness of teachers and trainers in general to engage in further training is important.

ACKNOWLEDGMENT

Acknowledgements are expressed to the EU project dMath Partners [1] for their contributions to the project.

REFERENCES

[1] http://dmath.hibu.no
[2] Bringslid, Odd, Building a European database of mathematical e-learning modules, Proceedings of the IPSI-2005 FRANCE April 23-26 2005, ISBN 86-7466-117-3
[3] http://europa.eu.int/comm/education/programmes/socrates/minerva/index_en.html
[4] http://dmath.hibu.no/xmath/
[5] http://www.w3.org/Math/
[6] http://www.wolfram.com/products/webmathematica/index.html
[7] http://xmath.hibu.no:8080/webMathematica/Xmath/Xcalc.msp
[8] http://www.hibu.no/
[9] http://www.cvut.cz/en/
[10] http://www.soft4science.com/index.html
[11] http://www.upco.es/webcorporativo/default.asp
[12] http://www.pspt.fi/eng/
[13] http://www.hisf.no/english/index.ssi
[14] http://www.indok.no/index.asp?section=&language=1
[15] http://www.sjf.stuba.sk/
[16] http://www.soft4science.com/products/sciwriter/s4s_SciWriter.html
[17] http://www.gen21.com/enterprise.htm
[18] http://www.w3.org/TR/soap/
[19] http://www.rhassociates.com/scorm.htm
[20] http://www.dessci.com/en/products/mathplayer/
[21] http://www.integretechpub.com/techexplorer/
[22] http://www.mozilla.org/
[23] http://browser.netscape.com/ns8/
[24] Demlova, Marie , Discovering Mathematics, Proceedings of the 2nd European Workshop on MathML and Scientific e-contents, 16th-18th September 2004 ISBN 952-9533-96-9, download [1], dMath Publications
[25] Bringslid, Odd, Multimedia Books in the Mathematical education of engineers, European Journal of Engineering Education VOL 24 , No2, 1999, 181-199
[26] Bringslid, Odd, Mathematical e-learning using interactive mathematics on the Web, European Journal of Engineering Education VOL 27 , No3, 2002, 249-255

Odd Bringslid
Assoc. Professor at The Institute of Technology, Buskerud University College, Norway. Project coordinator of EU Leonardo project dMath [1], EU Minerva project Xmath [4] and a former third party program designer for HP calculators. Officially appointed as Project leader of Educational Government Projects. Research into mathematical e-learning and Computer Algebra. **Marie Demlova**, Head of the Departmenet of Mathematics at Faculty of Electrical Engineering, Czech Technical University in Prague. Professor of Applied Mathematics. Research into semigroup theory, automata theory and theoretical computer science.

An Online Strategy for Evaluating Educational and Social Intervention Programs

Hansen, William B.; Reese, Eric; Wyrick, Cheryl H.; Jackson-Newsom, Julia; Bryant, Kelvin S.; and Dyreng, Douglas I.

Abstract— *Evaluation Lizard is a web application designed to provide decision making support and operational capacity for completing evaluations of educational and social interventions. Users complete essential evaluation design tasks online. These include defining treatment and comparison groups, defining treatment and survey events, and creating surveys. The application allows users to administer surveys online or to prepare specialized scannable paper-and-pencil surveys. The application allows users to manage an evaluation team (research assistants and data collectors) and the client organization (staff administering the intervention and participants). The application allows users to download their data and retrieve basic analysis reports.*

Index Terms—*evaluation, online tools, survey*

1. *Introduction*

THIS paper addresses the development of a web-based application for supporting the completing the evaluation of social interventions. The web application described, Evaluation Lizard, was developed specifically for use in the evaluation of drug abuse prevention programs, but can be used to assist in the evaluation of a broad array of social interventions.

2. *Problem Statement*

There has been a discernable shift in public policy throughout the United States and Europe toward the evaluation of programming designed to affect the health and well-being of their citizens. For example, in Europe, the European Monitoring Centre for Drugs and Drug Addiction [1] has set forth principles that promote the adoption and evaluation of research-based drug abuse prevention efforts in schools. As noted in a handbook for evaluation produced by EMCDDA [2],

"In the last few years, a growing number of prevention activities have been carried out in all the member states of the European Union (EU). Most of the projects, however, haven't been effectively evaluated, and so there is an urgent need for improving the knowledge about the process of 'prevention evaluation' and exchanging experiences and results."

Similarly, the US Department of Education Safe and Drug-free Schools and Communities program now includes four principles of effectiveness [3]. These principles state that a grant recipient shall:

(1) base its program on a thorough assessment of the objective data about the drug and violence problems in the schools and communities served; (2) establish a set of measurable goals and objectives and design its activities to meet those goals and objectives; (3) design and implement its activities based on research or evaluation that provides evidence that the strategies used prevent or reduce drug use, violence, and disruptive behavior; and (4) evaluate its program periodically to assess its progress toward achieving its goals and objectives and use its evaluation results to refine, improve, and strengthen its program and to refine its goals and objectives as appropriate.

The emphasis that emerges from these principles is one that reinforces a value on research-based programs and evaluation. Two of the four principles call for data collection – needs assessment and local evaluation.

3. *The Limits of Existing Approaches to Evaluation*

Requirements to complete local evaluations place a new burden on schools and communities who are not experienced with either needs assessment or evaluation methods. This is especially true when non-academic programs, such as drug or violence prevention programs,

Manuscript received October 7, 2005. This work was supported in part by the U.S. Department of Health and Human Services under Contract Number N44 DA-0-5065.

William B. Hansen is with Tanglewood Research of Greensboro, NC, USA (e-mail: bill@tanglewood.net) and is the contact author.

Eric Reese is with Tanglewood Research of Greensboro, NC, USA (e-mail: eric@tanglewood.net).

Cheryl H. Wyrick is with Tanglewood Research of Greensboro, NC, USA (e-mail: cheryl@tanglewood.net).

Julia Jackson-Newsom is with Tanglewood Research of Greensboro, NC, USA (e-mail: julia@tanglewood.net).

Kelvin S. Bryant is with Tanglewood Research of Greensboro, NC, USA (e-mail: kelvin@tanglewood.net).

Douglas I. Dyreng is with Tanglewood Research of Greensboro, NC, USA (e-mail: doug@tanglewood.net).

are introduced. Many school districts and community groups feel overwhelmed with these new requirements. In short, they are being required to complete an activity for which they are unprepared and for which they have few options or tools. Hiring an evaluation specialist to design and complete a needs assessment or an evaluation is often costly, typically beyond the budget of most small projects. Even when hired, the standardization of measures and methods have not been established. Such conditions often lead to sub-standard evaluation projects. These conditions make cross-project comparisons nearly impossible to complete.

4. THE SOLUTION PROVIDED BY EVALUATION LIZARD

Evaluation Lizard was initiated to be a low-cost, easy-to-use alternative to hiring a professional evaluator. The rationale for this approach rests of several research achievements in the field of prevention. These include:

(1) The development of standardized measures. As with any field of science, methods for assessing substance use behaviors and correlates of behaviors (e.g., attitudes and beliefs) were initially idiosyncratic and rarely shared by individuals and research groups investigating drug use. However, the field has increasingly moved towards standardized measures. In the United States, the Core Measures Initiative of the Center for Substance Abuse Prevention (CSAP) [4] has resulted in the compilation and publication of many relevant measures. Similarly, in the European Union, the EMCDDA has created an Evaluation Instruments Bank [5] that includes survey items in many languages.

(2) The development of a list of recognized programs. The ultimate result of this funding on prevention research in the United States has been the creation of a National Registry of Effective Programs and Practices (NREPP) [6]. The goal of NREPP is to identify drug abuse prevention and treatment programs and mental health promotion programs with proven effectiveness that can be disseminated broadly to schools and community service organizations. During the past several years, over 900 programs have been nominated for inclusion in the list of proven programs. As of January 1, 2004, 153 had been judged to have evidence of effectiveness and were included in the registry. North American program developers continue to search for effective strategies with the goal of reducing drug use.

(3) A by-product of research has been the systematization of social science evaluation designs [7]. Social scientists pride themselves on their ability to create evaluation designs that allow studies to interpret findings with validity – that is, to be able to demonstrate reductions in drug use that can scientifically be attributed to the effect of the program and not the effect of extraneous factors. The field has generally progressed to the point where there are now a handful of recognized evaluation designs. Thus, the structure of evaluations has also become simplified.

The cumulative effect of this progress allows evaluation designs to be specify criteria that can be used to evaluate known programs using standardized measures in local settings. The goal of Evaluation Lizard was to create an online system that allowed the simplified evaluation of local implementations of research-based drug abuse prevention programs.

This system was designed for use by local evaluators who are expected to continue to conduct evaluations according to acceptable practice, including protecting the rights of subjects as well as interpreting findings. The best use of this application currently is to assist evaluators conduct local evaluations of previously proven programs.

5. THE SOLUTION PROVIDED BY EVALUATION LIZARD

Designed to support schools and community groups who wish to complete outcome evaluations of drug prevention programs, Evaluation Lizard [8] facilitates the the following evaluation tasks: (1) create an evaluation project, (2) describe and define an intervention that will be evaluated, (3) define an evaluation design consisting of evaluation conditions, measurement events, and treatment events, (4) select outcomes measures including demographics, measures of drug use, and measures of mediating variables targeted by the intervention, (5) specify informed consent procedures, (6) collect data either online or using paper-and-pencil surveys, (7) create a database of survey data, and (8) complete basic descriptive analysis of the data. Our vision of this web application is that it will not only serve the short-term needs for evaluation, but that it will also a searchable library of measures, a repository of data from research and evaluation projects, and an exceptionally large database that can be mined for information.

Organized around primary functions associated with evaluating social interventions, the website's core competencies include: (1) managing projects, (2) managing people associated with projects, and (3) completing analysis and getting reports. The library of measures is a feature associated with the Evaluation Lizard that allows researchers and evaluators to find survey research measures from a large bank of measures. A help feature is also included to guide users through the website. Each of these functions is described in the section that follows.

5.1. Managing Projects

The first step in Evaluation Lizard is giving a new project an appropriate name and providing descriptive information useful to the Project Coordinator and others who are participating in the project. . Crucial to subsequent definitions is for the Project Coordinator to specify the type of organization in which the evaluation will be conducted. Evaluation Lizard selects educational organizations as the default organization type. Other organization types currently available include ecclesiastical and community.

To define a project, three elements must be specified: (1) the program being evaluated, (2) an evaluation design to be applied to the project, and (3) the measures that will be used in surveys that are administered to participants in the project.

5.2. Prevention Programs

A meta-analysis of NREPP prevention programs was recently completed by members of our team [9] that has allowed us to systematically identify important characteristics of research-based prevention programs. This meta-analysis evaluated the program content of 48 programs that specifically addressed substance abuse prevention and that had published program implementation manuals. Program content refers to the specific kinds of information presented in the program. Most often the program content of educational approaches focus on change personal management skills (such as decision making and goal setting), social skills (such as assertiveness and resisting peer pressure), or disposition (such as attitudes towards drugs, commitment to avoid drug use, and developing positive social norms that promote the non-use of drugs). Thus, the specific program content of each program is profiled. Once users select a program for inclusion in their project, the profile of that program is identified.

5.3. Measurement Library

We have assembled a library of measures that currently includes measures from four domains, demographics (gender, age, and ethnicity), behaviors (use of alcohol, tobacco, marijuana, inhalants, cocaine, heroin, etc.), mediators (measures that assess psychological changes targeted by program content such as skills, attitudes, and beliefs), and moderators (measures that assess participation in a program). Measures come from a compendium of survey research items developed by researchers in the field of drug prevention. Items are grouped into scales that are hierarchically categorized for ease of use by users. Thus, within each domain, items are grouped into classes, categories, and scales. For example, a measure that assesses students' beliefs about the consequences of drinking alcohol will be categorized with similar items that allow the evaluator to find items of interest.

5.4. Survey Design

Evaluation Lizard uses program content to define the appropriate components of surveys. Thus, we use NREPP programs' profiles to define what items should be included in surveys. Surveys for programs that teach students about decision making include items to assess their decision making ability and ability to control impulsivity. Surveys for programs that promote the development of personal commitments to avoid drug use assess students' intentions to use substances. Programs that focus on alcohol use include survey items for students to self-report about their alcohol consumption and drunkenness. We have created standardized surveys for each of the substance abuse prevention programs on the NREPP list. Thus, for any of the 48 programs for which we have been able to document program content [9], each has a content-appropriate, ready-to-use survey available in Evaluation Lizard.

5.5. Evaluation Design

Evaluation Lizard includes an application that allows evaluators great flexibility in terms of specifying evaluation designs. However, to simplify its use, a set of prescripted templates are available for the evaluator to choose from that automatically define basic treatment group and measurement options. These prescripted designs include a needs assessment, a pretest-posttest

treatment group only design, a posttest only treatment and control group design, a pretest-posttest treatment and control group design, a pretest-posttest multiple group design, and a lagged-cohort treatment and control group design.

Design Template

Template	Design
○ Assessment	Group 1 O_1 [O_2 O_3 O_4 O_5 …]
○ Pretest-Posttest Single Group	Group 1 O_1 X_1 O_2 [O_3 O_4 O_5 …]
○ Posttest Only Control Group	Group 1 X_1 O_1 [O_2 O_3 O_4 …] Group 2 O_1 [O_2 O_3 O_4 …]
◉ Pretest-Posttest Control Group	Group 1 O_1 X_1 O_2 [O_3 O_4 O_5 …] Group 2 O_1 O_2 [O_3 O_4 O_5 …]
○ Pretest-Posttest Multiple Comparison Group	Group 1 O_1 X_1 O_2 [O_3 O_4 O_5 …] Group 2 O_1 X_2 O_2 [O_3 O_4 O_5 …] Group 3 O_1 X_3 O_2 [O_3 O_4 O_5 …]
○ Lagged Cohort Comparison	Group 1 O_1 X_1 O_2 [O_3 O_4 O_5 …] Group 2 O_2 O_3 [O_4 O_5 O_6 …]

Users can modify the number of elements in their design. They can modify the number of treatment and control conditions as well as the number of testing events. They can also order treatment and testing events. For instance, Evaluation Lizard will allow the user to create a design in which there was a pretest, a treatment intervention, a posttest, a second treatment intervention, and additional posttests. Evaluation Lizard tracks all events in a design.

Back | Continue | Modify Design... | Reset Design

Design Review

Miscellaneous Project Parameters	
Condition Assignment Level	Section
Condition Randomization	Manual
Subjects' Informed Consent Method	None

Condition	Event 1	Event 2	Event 3
Group 1	Pretest 1	Treatment 1	Posttest 1
Group 2	Pretest 1	{No Treatment}	Posttest 1

When specifying the design, users also define three important features of the design. First, users define the level at which condition is assigned. Users define this using the organization type specified. Users define whether and how units will be randomized. There may either be no randomization, randomization controlled and reported by the user, or randomization automated by Evaluation Lizard. Finally, users define informed consent procedures to be followed. Options for the latter include none (e.g., no informed consent required), passive (e.g., subjects are assumed to be included unless specifically excluded) or active (e.g., subjects are excluded until permission to participate has been granted).

As a final feature, the time frame for completion of each event can be specified. Time frame is useful for limiting access to online surveys should the evaluator wish to specify this. Time frame is also important for defining when the database will be made publicly available outside the project for analysis. (We plan to embargo data for five years following the completion of a project. After this date, elements of the database stripped of identifiers will become available for public use.)

5.6. Managing People

Evaluations involve people. Evaluation Lizard facilitates the management of people from three different groups. The first group is the project staff who comprise the evaluation team.

The Project Coordinator for each project can give privileges to other members of the evaluation team. Privileges include: modifying personal information and privileges of other evaluation team members, modifying information about the client organization, modifying the project (e.g., selecting programs to be evaluated, modifying evaluation designs, selecting measures to be included in surveys), assigning individuals or groups to experimental conditions, requesting data analysis, having access to raw data files, adding and modifying participant information, viewing data collection status reports, and initiating the printing of paper and pencil surveys.

The second group is the client organization – the organization where the evaluation is taking place, such as a school system. The client organization is described in hierarchical terms that include a geo-coding of people's roles. Thus, for an evaluation conducted with clients who are in a public school educational setting, people are assigned roles for the nation, region, state, county, city, district, school, teacher, and section.

The third group of people is the evaluation participants – subjects who complete surveys and receive interventions. One of the features added to the application that is important in conducting evaluations in schools is the ability to associate participants with classes. In longitudinal evaluations, students may change which classes and which schools they are enrolled in. Evaluation Lizard provides a tool for associating students (or other types of

participants) with the group in which the survey will be administered.

Evaluation Lizard includes online tools for managing all three groups. Evaluation team members and members of the client organization can be assigned privileges crucial to completing evaluation tasks. Tools for importing or entering participant information and assigning participants to conditions are provided.

5.6. Survey Administration

Evaluation Lizard has two strategies for administering surveys. The original intent was to administer surveys online. Online surveys are not new [10] and Evaluation Lizard follows commonly employed strategies. Students use a web browser to navigate to a URL provided to survey administrators. Students login with a login id and password, both of which are provided to the administrator Students then complete the survey. Data are tied to student login information.

The second strategy for administering surveys was developed in response to requests from schools. Scannable surveys are printed. Initial surveys are created using a word processing system and then printed to Adobe Acrobat® PDF files. These surveys are further processed so that coded information is individualized for every page of every survey, allowing its source to be traced. Thus, survey pages have bar coded identifiers (but no human readable text) printed at the bottom of each page to identify the survey once it is scanned. Printed surveys have tear off cover sheets that have participant information to ensure it gets to the correct person. The local team is responsible for survey administration and interpretation of findings.

5.7. Data Analysis and Reports

Evaluation Lizard allows evaluators to complete basic analysis tasks common to evaluation. Three tasks may be completed by evaluators: (1) generate survey status reports, (2) complete basic analyses, and (3) download raw data for local analysis.

Evaluation Lizard tracks survey participants across events and can provide evaluators with dynamic information about survey completion. This report allows administrators to identify missing subjects, which is particularly valuable for projects that wish to be diligent in getting as much participation as they planned.

Evaluation Lizard automatically generates descriptive analyses for all items included on the survey. The application generates frequency tables. All events completed to date are broken down by event (e.g., pretest, posttest, etc.) and condition (e.g., treatment versus control). Tables include frequencies for each response as well as percentages of individuals within-condition who responded. For behavior items and for mediating variable scales, Evaluation Lizard also produces bar charts that compare conditions at each testing event as in the sample below. Reports are printed as Adobe Acrobat® PDF files.

Evaluators can download the raw data from their projects for further analysis. For example, evaluators will likely want to complete inferential statistical analyses on their own. When evaluators download data from Evaluation Lizard, data are stripped of positive identifiers. Thus, the data cannot be traced back to the individuals who completed the surveys. Subjects are instead linked by an ID number generated by Evaluation Lizard. This allows confidentiality to be maintained.

6. Analysis of the Capability of the Evaluation Lizard Application

6.1. Functionality of the Application

Evaluation Lizard is a complex relational database-driven application, consisting of numerous condition-driven relationships. Originally, it was hoped that the application would be usable by untrained evaluators. However, because of its complexity, the use of many elements of the application continues to be reserved for use by trained evaluators. During the initial phases of its formal launch, a quality control specialist documented non-functioning and error-producing elements of the application. These errors have gradually been cleared and, as of this writing, few known programming errors exist. The greater need seen with the application in its current state is increased simplification of tasks and the adoption of additional pre-coding wherever possible. For example, the geo-coding of states has been pre-loaded so that the user can select in which states in the US the evaluation will occur. Region and nation then follow automatically. Additional pre-coding for other areas of the world, and pre-coding of counties and cities would significantly aid the speed with which client organizations can be entered. The system has not yet been tested to verify compliance with Section 508 of 36 CFR Part 1194.

6.2. Accumulation of Survey Items

As of January, 2005, there were 2,481 survey items in the measurement library. Most items do not stand alone, but are part of a scale, of which there were 436. The library contains 814 questions that assess behavior (such as alcohol consumption), 64 demographic items, and 1,603 mediating variable items which form 95 different scales. The measurement library is now continually expanding with new measures being

added as they are developed, discovered, or required by users.

6.3. Use of the Application

During the first year of its operation as a commercial service in 2004, 7,540 surveys were administered using Evaluation Lizard. In 2004, 25 organizations ordered surveys. In the first nine months of 2005, 10,803 surveys have been administered through Evaluation Lizard, roughly doubling the rate of use from the first year. To date in 2005, 55 agencies have purchased surveys. Every agency that purchased surveys during 2004 returned to purchase surveys in 2005, indicating a satisfaction with the service. The system has been used primarily to evaluate a select number of SAMHSA drug prevention programs: e.g., All Stars, Project Northland, and Creating Lasting Family Connections.

6.4. Cost

Surveys are billed at a rate of $2 per printed survey and $.50 for online surveys for standard surveys. This includes all costs except shipping and specialized report writing. Thus, survey design, evaluation design, printing or the creation of online surveys, data reduction (scanning for printed surveys), database creation, and the creation of basic analyses and reports are all provided within this fee. This is a significant savings compared to the costs of a typical evaluation consultant who will often have costs exceeding $10 per administered survey.

To date, few customers have opted for online surveys. Availability of computers in schools, bandwidth, and control when administering surveys are thought to account for customers' preference for paper-and-pencil versions. This situation is expected to be improved in the future.

7. CONCLUSION

Evaluation Lizard is a web application that can make the local evaluation of drug prevention programs affordable and easy to complete. The system takes advantage of the state of the art in prevention evaluation practice, eliminating many barriers schools and community groups might otherwise encounter. Novice evaluators with little experience can produce effective evaluations using Evaluation Lizard. In addition, there is enough flexibility for seasoned researchers to build custom evaluations.

Evaluation Lizard has been a collaborative effort between social scientists – who have extensive training and practical experience in evaluation design – and a team with expertise in information technology. This marriage has been unusual and revealing in that complex tasks typically done by hand by social scientists have been translated into algorithm- and database-driven procedures design by computer programmers. The experience of producing this application has revealed the degree to which human-based decisions are often complex and idiosyncratic. Finding strategies to simplify and systematize decisions about how to conduct evaluation designs required extensive effort and a willingness to adapt approaches by both teams.

Evaluation Lizard, while overcoming many procedural problems, is not a simple guaranteed solution for the difficult and complex evaluation process. In the future this tool (or others like it) need to be developed to allow users to do such things as warehouse data collected using other methods of data collection and to mine data in various ways. For example, most local evaluation as actually implemented by teams often lack the rigor called for in order to have high degrees of internal and statistical validity [7]. New methods for creating synthetic control groups and for combining multiple evaluations through meta-analytic or other techniques need to be developed. In many respects, it is only after a large repository of data exists that such developments will be possible. Thus, even though Evaluation Lizard cannot yet address many of the challenges that accompany evaluating social interventions, this web application provides an opportunity to make progress towards addressing these issues.

REFERENCES

[1] G. Burkhart, M. Nilson, and D. Ballotta. "Drug prevention in EU schools: Information and reporting systems are crucial," Drugs in Focus. September-October 2002, http://www.emcdda.eu.int/index.cfm?fuseaction=public.AttachmentDownload&nNodeID=526.

[2] C. Kröger, H. Winter, and R. Shaw. "Guidelines for the Evaluation of Drug Prevention Intervention: A Manual for Programme Planners and Evaluators," Institut fur Therapieforschung: Munich, Germany. 2000, http://pl-www.emcdda.eu.int/?fuseaction=public.AttachmentDownload&nNodeID=1753.

[3] US Department of Education. "Safe and Drug-Free Schools Program; Notice." *Federal Register*. 63(104), pp. 29902-29906, June 1, 1998.

[4] CSAP, "Core Measures Initiative Phase I Recommendations," January, 2003, http://alt.samhsa.gov/grants/2004/downloads/CSAP_Core_Measures.pdf.

[5] EMCDDA, "Evaluation Instruments Bank," 2005, http://eib.emcdda.eu.int/.

[6] http://modelprograms.samhsa.gov.

[7] W. R. Shadish, L. C. Leviton, and T. Cook. *Foundations of Program Evaluation.* Thousand Oaks, CA: Sage Publications Inc, 1990.

[8] http://lizard.tanglewood.net/default.asp.

[9] W. B. Hansen, J. H. Derzon, L. Dusenbury, D. Bishop, K. Campbell, and A. Alford, "Analysis of Magnitude of Effects of Substance Abuse Prevention Programs Included in the National Registry of Effective Programs Through 2003: a Core Components Analysis," Report submitted to the National Center for the Advancement of Prevention. Tanglewood Research, Greensboro, NC, USA, June 2004.

[10] S. McCoy, and P. V. Marks. "Using electronic surveys to collect data: Experiences from the field," 2001. Seventh Americas Conference on Information Systems. 2001, pp. 1502-1505.

The Successive Principle – a scientific crystal ball for management

Lichtenberg, Steen

Abstract—*The management tool known as the Successive Principle*[1] is an example of successful interdisciplinary R&D involving management, Cost Engineering*, Systems Economy*, statistical theory and some psychology.*

Existing methods of project cost estimating, planning and business analyses too often lead to overruns, delays, etc. when used as the basis for decision-making.

The solution discussed in this paper has achieved a strongly enhanced realism in terms of its estimating procedures, and has largely eliminated unplanned overruns, delays, etc.

The paper outlines the basic principles behind this solution as well as some examples of experiences and applications.

In conclusion, management and other decision makers can now restrict overruns etc. to cases of severe force majeure events.

Index Terms— *Interdisciplinary Research, MBA & Systems Economy, Project & Risk Management, Statistical Theory, Group Psychology, Budget Overruns & Delays.*

1. INTRODUCTION

This publication exemplifies successful interdisciplinary research and development work. It exemplifies the synergic value of integrating different scientific areas. The result is a multipurpose management tool called the Successive Principle.

Like glasses which give the short-sighted driver a sharper long-distance view, this tool gives the management user a sharper and more realistic long-distance view of the future. A realistic quantitative result of large, complex ventures can now be predicted with substantially augmented realism.

Users have even called it a scientific crystal ball. Furthermore, it identifies in ranked order the most interesting factors of the venture in question.

Revised manuscript received Nov. 20., 2005. This work was supported in part by the Norwegian National Science Fund, and Danish university funding, which are greatly appreciated.

The author is Emeritus Professor in Project Management and Systems Engineering with the Technical University of Denmark, Construction Management Department. At present he is a management consultant with Lichtenberg & Partners, Copenhagen (e-mail: steen@lichtenberg.org).

[1] An * refers to a list of terms.

Its primary professional areas of application are Management & Project Management, Systems Engineering, Risk Analysis & Cost Engineering. It is also of interest in R & D work, when a realistic, non-biased quantitative result is required from an input which includes uncertain, fuzzy elements.

2. THE CURRENT PROBLEM

Severe overruns in cost and time frequently bedevil large programmes, projects, strategic ventures etc., both private and public. These cases sometimes are disasters [1].

The Sydney Opera House, the Channel Tunnel and some of the Olympic Games are the most well known but are only the tip of the iceberg.

The Sydney Opera House

Several research projects have shown that among large IT projects only a small minority came out on budget, while the average overrun was considerable [2].

Recent research by Professor B. Flyvbjerg into large infrastructure projects has shown a similar result [3].

This always causes severe problems. Where do we find supplementary funding? Might we end up with an unfinished shell, or at best with a substandard facility? Do we go bankrupt or at of the company's future activities? Cost

The Channel Tunnel

Olympic Games

engineers, planners and other professionals do complex, extensive and skilled work preparing a detailed basis for budget and schedules, so why do we suffer these problems?

3. THE BASIC REASON SEEMS TO BE THE OLD NEWTONIAN SCIENTIFIC PARADIGM

A 300-year old paradigm requires us to focus upon matters that can be documented, and to avoid congestion and other subjective and fuzzy matters. It is still a firm basis in the higher education of engineers and economists. This is of course valuable in many cases; however, it can be disadvantageous in some circumstances.

Working with plans for large projects and other ventures, the planners and estimators have to work with incomplete project material, specifications, etc. when preparing the basis for a budget. The cost of the documented material is carefully detailed and skilfully calculated on the basis of historical data and other experience. In addition, they tend to assume that implementation will be relatively controlled, and unhampered by major problems. Finally, they apply a traditional dispensation whereby a somewhat arbitrary 10% is added for contingencies without any documentation.

4. THIS INEVITABLY CAUSES OVERRUNS

A still larger part of the project is not documented even at the point when the crucial decisions have to be taken. Add to this that a conventional budget estimate takes no relevant account of factors such as future added facilities, complications, requirements, unforeseen influence exerted by authorities, the owner, the users, local NGOs, nature's caprices, human failures, etc., etc.: all typically – but not always – representing much larger amounts than the 10% contingency figure. No wonder we often experience large overruns.

Another consideration is that the many parties who have a stake in getting the project approved/authorised naturally wholeheartedly accept the aforementioned conventional and wholly "legal" budgets.

5. RISK ANALYSIS PROCEDURES AS A SOLUTION

Various risk analysis procedures has been developed attempting to remedy uncontrolled overruns and delays. They all include identification of sources of uncertainty, an evaluation phase, and typically a Monte Carlo calculation of the result, consisting of a mean value and a standard deviation, or the so-called S-curve, linking probability to keep to budget versus budget size. A primary method is the Range Method, developed by Michael W. Curran already around 1970 [4].

The existing methods generally suffer from sufficiently covering identification of sources of uncertainty, from biased quantitative evaluation, and from lack of efficient handling the statistical correlations among various uncertain factors. A huge amount of uncertain factors do also hinder a thorough treatment[2].

These problems, since the early 1970´s has been in focus with a successful result, as further developed below.

6. A NEW GUARANTEE AGAINST UNPLANNED OVERRUNS

However, a somewhat untraditional procedure has been developed which brings you very close to a guarantee against overruns, except in the case of major catastrophes, obviously. It is known as the Successive Principle*. It was developed during the 1970s and later by the author and colleagues.

One example is the complex high-tech, multipurpose 10,000-seat arena*, Oslo Spectrum* in Norway. The original budget was $45 million. The use of the Successive Principle three years later, before the project was due to start, identified $125 million as a realistic cost. The project was then rationalised, after which an analysis process generated $80 million as a mean value* +/- approx. $10 million as the standard deviation*. The project organisation was allotted the $80 million as a budget, while the official building committee was given the $10 million as a reserve. However this reserve was never used.

The official project account after the successful erection deviated by less than 1% from the calculated mean value.

Oslo Spectrum in Norway

Another example is the *Lillehammer Olympic Games.* The initial budget rose from $230 million to $385 million over the summer, more than four years before the games. A risk analysis showed an expected final total cost of $1230 million. This was of course politically unacceptable. The investment plans were then reorganised –

The Lillehammer Olympic Games.

[2] The range method, however have solved this particular problem.

in part supported by the analysis result – and followed by several updating analyses. The expected investment figure was eventually reduced to $800 million as a mean value. This became the working budget, while the official committee was allocated a reserve of approx. $90 million. However, the final official accounts equalled the analysis mean value of $800 million, so the reserve fund was saved and was used to operate the facilities after the games.

The *telecommunications company Ericsson's first mobile or cellular phone* is an example of the use of the Successive Principle as a support for making the right decisions. It was originally allocated relatively low priority among a set of new ideas in an R&D department at Ericsson.

A Successive Principle analysis then revealed it to be a highly promising idea. Accordingly, it was developed and became *Ericsson's* greatest commercial success ever.

Ericsson's first mobile or cellular phone.

It might be said that the above three cases merely capitalised on coincidence or good luck. But over the course of more than 25 years, so far no negative feedback has been received from the sub-set of 250 – 300 cases which were analysed under controlled conditions. But of course not all of the cases ended up quite so close to the exact mean value as the above.

7. DEVELOPMENT OF THE "SCIENTIFIC CRYSTAL BALL"

A. A brief history

This research is based on the new scientific paradigm which accepts and deals with uncertain and fuzzy issues. These ideas was supported by scientists, like Tverski, and Kahneman [5], Spetzler, and Stäel von Holstein [6], Apostoiakis [7], and others.

The aim of the research into the Successive Principle has been to allow professionals to calculate the projected total actual cost or duration of new projects or ventures in a more realistic and controlled manner – more focused on the relevant future uncertainties. It was initiated at the Technical University of Denmark by the author in the beginning of the 1970s [8]. An international research network was formed soon afterwards. It included Stanford University and MIT in U.S.A. as well as universities in Loughborough, UK, Gothenburg, Sweden, and not least the Technical University of Norway in Trondheim [9] - [14]. From the 1980s onwards the practical procedure known as the Successive Principle was in use internationally in most public and private business areas in order to safeguard them against overruns, delays, etc. [15], and [16].

B. The Principle entails the following basic requirements:

- Focus on the future and its uncertainty.
- Inclusion of all internal and external matters that might affect the total result – including fuzzy ones.
- Non-biased subjective evaluation of their impact on the result made by a relevant group.
- Scrupulously following the natural laws of uncertainty during the calculation of the result.

C. The solution tools

- Basic Systems Economy* and Cost Engineering* tools, such as the Net Present Value concept*, Work Breakdown Structures* etc.
- The Bayesian statistical theory* [7].
- The use of group synergy in a balanced and broad-based analysis group of competent people [12].
- A top-down principle involving successive steps detailing the most critical items or factors – and only these – in each step.
- Ensuring sufficient statistical independence* among the uncertain items and factors.
- Using an evaluation procedure which takes the many pitfalls into account.
- A top ten list of the most critical items or factors, updated and used during the successive process and finally used as a key result.

8. THE SUCCESSIVE PRINCIPLE – A PROCESS OUTLINE

The Successive Principle was originally a tool for early and fast cost estimating and scheduling in the construction industry and was known by some users as "intelligent cost estimating". It has developed into an integrated management instrument, as illustrated in section 11, and 12 below, and further developed in a recent handbook [17].

A primary principle is to let a balanced group of key persons perform a few analysis sessions together, during which they identify and then organise all possible sources of uncertainty. They then operate top-down, systematically detailing and evaluating the most important issues in successive steps.

This allows us to keep an overview, to focus on the really important aspects and to avoid wasting resources on the many issues of little or no importance. Another important facet is the arranging of all uncertainties into discrete statistically independent elements and then working with the conditional uncertainty* of each of the elements. This allows simple yet sufficiently accurate statistical calculations.

9. THE PROCEDURE OUTLINED

The procedure is only briefly outlined below due to limitations of space. A more elaborate description and discussion is found in [17].

The procedure is organised into the following eight steps.

Step A. Establish a suitable analysis group.
Step B. Clarify the goals and objectives, as well as any firm preconditions.
Step C. Identify all issues of potential importance.
Step D. Organise the issues into discrete groups, and define for each group a base case assumption and how it could change for better or for worse.
Step E. Quantify all uncertain elements using triple estimates* and good evaluation techniques.
Step F. Calculate a provisional overall result and draw up a top ten list of the most critical items or activities.
Step G. Specify the most critical elements in successive steps, guided by the current top ten list.
Step H. Once a satisfactory result has been arrived at, complete the analysis work with a suggested action plan.

10. IMPORTANT DETAILS

Step A. Establish a suitable analysis group.

An appropriate analysis group is appointed according to the specific purpose of the analysis. In addition to a number of experts representing the major key areas, the analysis group should include individuals who can provide the vital elements of creativity, flair and breadth. The analysis group should ideally include both young and mature individuals, both generalists and specialists and should represent “both halves of the brain". You will usually also need an individual who can play the role of "devil's advocate"; this is especially important in the case of a project whose project team generally wants a successful result, and whose judgement may therefore be over-optimistic.

It is also important to select an appropriate and agreeable location where the analysis group feels comfortable, and in particular will not be disturbed in its day-to-day work; and of course to find rooms where the necessary facilities/ equipment are available and meet professional standards.

The subsequent steps are performed in group sessions, while using modern group psychology inspired by Robert B. Gillis [12].

Step B. Clarify the goals and objectives, as well as any firm preconditions.

The analysis management team will have prepared a draft. However, it is important to discuss it properly in the group and to make adjustments until full understanding and consensus have been reached.

Step C. Identify all issues of potential importance.

The identification of sources of uncertainty (possibilities or risks) is typically achieved by means of a brainstorming process. This usually identifies 50 – 100 issues.

It is specifically verified that the analysis group
has examined a sufficiently broad variety of issues, and e.g. not only the technical ones.

Step D. Organise the many issues.

The identified key words are grouped together into 8-12 statistically independent groups. A clear and simple base case assumption is defined for each group, as well as how it could change for better or for worse. The normal volume of this description is four to seven pages.

Step E. Quantify, using triple estimates and good evaluation techniques.*

A master schedule network or a master calculation structure is chosen. Each main activity, main cost item or factor is quantified from the highest level using the triple estimating technique*. In order to avoid evaluation bias, a specific ‘Group Triple Estimating technique’* has been developed by N. Lange [18] and other, much inspired by C.S. Spetzler and Stäel von Holstein [6]. Shortage of space prevents further mention here of the psychology involved [17, section 5.2].

The activities and related main items are evaluated under the relatively firm base case assumptions. This ensures a sufficient degree of statistical independence. For each group of overall influences a correction figure is evaluated, also using the Group Triple Estimating technique*. It may be in absolute units or as a percentage evaluation.

Step F. Calculate the resulting total and a top ten list of the most critical items or activities.

Statistical independence* is thus largely achieved. To reduce any remaining dependencies further, the analysis group operates with the concept of conditional uncertainties*. This allows a simple yet sufficiently accurate statistical calculation to be made.

The result of the above evaluations is calculated. The calculation follows the natural laws of uncertainty, in this case the Bayesian statistical theory*. In addition to the total mean value* and its uncertainty, a top ten list emerges, showing the most important and critical local sources of uncertainty.

Step G. Specify the most critical elements.

This preliminary estimate or schedule is now detailed in successive steps, with the most critical elements being specified at every step. The

guidance in this "intelligent" detailing process is provided by the aforementioned top ten list. It actually leads to an optimal breakdown and evaluation of only those elements which warrant the attention.

Step H. After the final result has been achieved, the analysis work is completed with an action plan.

After a number of such cycles, the elements with inevitable uncertainty will increasingly dominate: after 8 - 12 cycles they usually account for 80 to 90% of the total uncertainty. Consequently, we are close to the minimum uncertainty of the grand total and similarly close to a successful conclusion of the analysis. At this stage, the degree of detailing usually involves fewer than a hundred items of which a considerable number are correction items.

The analysis group will usually be prompted by the top ten list to draw up a suggested action plan by way of a conclusion to the entire analysis process. The aim is to identify actions which may either exploit opportunities, protect the task against risks, or simply reduce uncertainty. A brainstorming process at this point is a highly appropriate means of identifying such ideas.

11. ON EXPERIENCES AND RESULTS

Experiments have verified the psychology behind the subjective evaluations and the use of the Group Triple Estimate procedure [17, section 5.2], and [18]. The statistical calculation procedure has been verified already during the 1970s and 1980s by professors I. Thygesen and P. Thyregod [19].

Practical experiences are primarily drawn from the 250-300 full-scale tasks performed during the last 25 years where it is verified that the procedure has followed “the rules of the game”. They cover most business areas and all sizes up to the “mega” size.

The primary result is a most realistic mean value of the actual future total result, whether in terms of cost, time, profitability, resource or consumption. This result is given in statistical mode, with a mean value* and a standard deviation*, or alternatively as the so-called S-curve, indicating the probability vs. the total value.

The top ten list of the resulting most uncertain aspects is also much appreciated by users. Improved team-building amongst the parties involved is also highly appreciated as an important side-effect.

A full statistical breakdown unfortunately is hindered by confidentiality. An exemption is meeting the deadline of large projects [20]. However no negative response has been met so far, only frequent surprise to see a small deviation between the prognosis and the final result. Users frequently implement the Successive principle after having tried it. Even the Norwegian Ministry of Finance has been inspired by the positive results of using the Successive principle to require for all major National projects to use an uncertainty analysis at the front end [21].

12. PRINCIPAL FIELDS OF APPLICATION

The Successive Principle is an integrated approach for:

* Quality assurance of budgets, schedules, profitability analyses and other financial analyses.
* Risk and opportunity analyses.
* Suggested action plans for improvements.
* Ranking of alternative solutions.
* Team building and consensus support.

Sorted by area, it is used as follows.

A. Senior Management, Quality and Risk Management

- Practical elimination of unpleasant surprises (e.g., overruns or delays).
- Risk-assessed corporate budgeting & planning.
- Greater certainties that key issues are being identified and actioned.
- Support for corporate contingency and risk management.
- Loss-making projects may be cancelled in due time.

B. Sales and Marketing

- Sales budgeting and planning
- Consideration of opportunities as well as risks
- in competitive situations
- Bid preparation and development. User companies have proved more frequently to be successful in competitions.

C. Project Management

- Project start-ups are significantly improved.
- Development of realistic plans and budgets.
- Reductions of costs or project duration during project implementation.
- Creative problem solving is supported.
- Team building is supported.

13. SOME LIMITATIONS

Only the overall result is reliable, not each sub-item or activity. Catastrophes and other major 'either-or' or force majeure events require supplementary procedures. The approach is limited to organisations with a modern, open management policy and acceptance of group work.

It supplements rather than replaces planning. It requires trained facilitators who know the “rules of the game”. Subjective uncertainty must be accepted. The implementation process requires effort and the support of senior management.

Finally, it must be admitted that the untraditional nature of the Successive Principle often

hinders its proper use in more conservative environments.

14. SUMMARY

The Successive Principle is cited in this paper as an example of the benefit of interdisciplinary research. It has used newer scientific paradigms which accept that fuzzy issues must be seriously dealt with. The end result is an integrated management and decision support methodology. It relates well to contemporary post-industrial management principles and attitudes. The strengths and benefits of the approach include

- Enhanced grasp of an uncertain future.
- Consideration and handling of the turbulence and uncertainty of business in a systematic and scientifically sound way.
- Integration of objective and subjective aspects.
- Identification of and focus on the most important uncertainties (risks and opportunities).

ACKNOWLEDGEMENTS

The author gratefully acknowledges the decade-long cooperation with fellow researchers as well as with Futura International, the international network of consultants dedicated to the use of the Successive Principle [15].

REFERENCES

[1] Kharbanda, O.P. & Stallworthy, E.A.: "How to learn from Project Disasters*, Gower (UK),* 1983.

[2] Molokken-Ostvold, K: "A Survey on Software Estimation in the Norwegian Industry", *10. International Symposium on Software Metrics,* sep. 2004, Chicago, USA, pp.208-219.

[3] Flyvbjerg, Bent: "Procedures for Dealing with Optimism Bias in Transport Planning", *The British Department for Transport,* Guidance Document, June 2004.

[4] Curran, Michael W.: "Range Estimating", *Cost Engineering (AACE), vol.31/no.3,* March 1989, p.18.

[5] Tversky, A. & Kahneman, D.: "Judgment under Uncertainty: Heuristics & Biases", *Science, Vol.185/27,* sep.1974, p.1124-1131.

[6] Spetzler, C. S., and Stäel von Holstein, C-A. S.: "Probability Encoding in Decision Analysis", *Management Science*, *vol. 22, no. 3,* Nov. 1975, pp. 340-358.

[7] Apostolakis, Georg: "The Concept of Probability in Safety Assessments of Technological Systems", *Science, vol. 250,* 7. Dec. 1990, pp. 1359-1364.

[8] Lichtenberg, Steen: "The Successive Principle", *proceedings, International PMI* Symposium*, Washington, DC, 1974, pp. 570-578.

[9] Klakegg, Ole J.: "The Stepwise process", research report, *Norwegian Technical University, Department of Construction Planning,* 1990, vol. 1 and 2 (in Norwegian).

[10] Klakegg, Ole J.: "Scheduling", Dr.Thesis, *NTH, Department of Construction Planning, Trondheim, Norway,* 1990 (in Norwegian).

[11] Moller, L.B.: "Scheduling under uncertainty", Ms.Thesis, *Technical University of Denmark, DTU,* 1977 (in Danish).

[12] Gillis, Robert B.: "IMPACT - Interactive Processes & Communication Techniques", publ.: *R. B. Gillis & Assoc. Inc.,* Vancouver, Canada, 1988.

[13] Lichtenberg, Steen: "New PM Principles for the Conception Stage - Outline of a new Generation", *International Journal of Project Management,* vol. 7/1, Feb. 1989, pp. 46-51 (also presented as a keynote paper at the *IPMA World Congress, Birmingham, 1988*).

[14] Lichtenberg, Steen: "The Successive Principle - A New Decision Tool for the Conception Phase", *Proceedings, 1989, Joint Project Management Institute/IPMA International Symposium, Atlanta, Sep. 1989, pp.16-25.*

[15] Futura International: "An Introduction to the Lichtenberg method", published on www.futuraone.com, 1998

[16] Archibald, Russell D. and Lichtenberg, Steen: "Experiences using Next Generation Management Practices - the Future has Already Begun", keynote paper, *Proceedings IPMA* 1992 World Congress*, Florens, Italy, June 1992.

[17] Lichtenberg, Steen: "Proactive management of uncertainty – using the Successive Principle", a handbook, *publ. PF Forlag, Lyngby,* Denmark, 2000, 334 p. (delivered by the author via www.lichtenberg.org).

[18] Lange, Nils: "Subjective Evaluation", Masters Thesis, *Technical University of Denmark, Department of Plan ning,* 1985, not published (in Danish).

[19] Thyregod, Poul: "Analysis of the Statistical Conditions in the Successive Principle*",* research note, *Technical University of Denmark, Institute for Mathematical Model ling,* 1982, not published (in Danish).

[20] Glahn, Jesper E.: "Untraditional Scheduling eliminates Surprises – Survey of Practical Experiences*", Proceedings, 7.th IPMA* World Congress*, Copenhagen, 1982,12 p.

[21] Berg, Peter (ed): "Controlling Governmental Investments", working group report, *Norwegian Ministry of Finance*, 1999 (in Norwegian).

LIST OF TERMS

Bayesian statistical theory is a widely accepted theory, which includes subjective probability in opposition to the classic or frequentistic statistical theory, which only accepts sets of documented data. Both theories use the same set of formulas [7].

Conditional uncertainty is the uncertainty of a local uncertain variable on condition that all other uncertain parameters are in their mean value.

Correlation or dependence coefficient. A statistical concept denoting the degree to which two separate uncertain figures follow each other. One limit is full statistical independence.

Cost Engineering is a part of Engineering Economy or Systems Economy, which deals with the cost of major assignments and related issues, see Systems Economy.

Engineering Economy, see Systems Economy.

Group Triple Estimate procedure is a procedure aimed at obtaining a neutral result by avoiding a set of pitfalls linked to subjective evaluations [17, section 5.2],. See also Triple Estimate.

IPMA, the International Project Management Association, a European and Asian-based professional organisation. See also PMI.

Mean value (also called expected value or expectation value) is a central value of an uncertain figure.

Net Present Value, NPV, is a widely used profitability criterion. It summarises all in and outgoing payments in discounted form (discounted back to the present time) for a specific system in contrast to other alternative ventures.

PMI, Project Management Institute, an American based professional organisation. See also IPMA.

Project Economy, see Systems Economy.

Standard deviation is a statistical measure of the dispersion or variation of numerical data from the mean value (see this term).

Statistical or stochastic independence, see Correlation coefficient

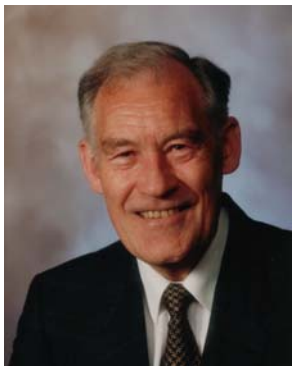
Successive Principle (also called the Lichtenberg method) is a multi-purpose management and Cost Engineering tool used to identify a realistic future result of a venture (cost, duration, profitability etc.) and the related primary uncertain issues.

Systems Economy (also called Engineering or Project Economy) is a part of economic science which deals with the profitability and other economic effects of specific ventures. See also Net Present Value and Cost Engineering.

Triple Estimate. The mean value* and standard deviation* of an uncertain figure is calculated as a weighted sum of the extreme minimum, the extreme maximum (theoretically the 1%, resp. 99% quantiles), and the most likely values. See also the Group Triple Estimate procedure.

Work Breakdown Structure, WBS, is a Cost Engineering concept. It denotes the way in which various sections or subsystems of an assignment or project are sub-divided into main items, items, and sub-items, in a hierarchical structure for planning or estimating purposes.

ON THE AUTHOR



As a researcher and consultant, Emeritus Professor Steen Lichtenberg, D.Litt. has worked successfully for many years within the area of project management and cost engineering, and specifically on safeguarding major projects against budget overruns and delays. Former president and honorary member of the International Project Management Association, IPMA*.

E-Mail: steen@lichtenberg.org.
Homepage: www.lichtenberg.org

Implementing Numerical Reasoning in ILP

Demšar, Damjan[1] and Gams, Matjaž[1]

Abstract - We introduce the 4S system capable of handling numerical problems within the general Inductive Logic Programming (ILP) framework. Previous systems had severe problems with handling numerical domains, as well as with using background knowledge and multi-relational data at the same time. The idea introduced in 4S is that regression results are integrated into an ILP program as numerical formulas. A formula is created dynamically according to the available examples and hypotheses. In this way, 4S can deal with numerical problems presented in multi-relational data by using background knowledge. The logic-based and numeric-based techniques in 4S collide in the presence of complex recursion, but we believe that the functional power of our system is less limited than related ones. We analyze the performance of 4S.

Inductive logic programming, Linear regression, Machine learning

1. INTRODUCTION

Various attempts have been made to implement Inductive Logic Programming (ILP) systems that are capable of handling numerical data [4], [22], [29], [30]. The general idea is to combine ILP with numerical manipulation in order to enlarge the space of problem domains that can be successfully tackled by ILP systems. Basic ILP has several advantages [21], such as the ability to deal with data stored in multiple relational tables and to take into account background (domain) knowledge expressed as a logic program. ILP also uses powerful expression language to describe logical relations.

Pure ILP, however, is not capable of handling numerical data [2], especially when the output "class variable" is continuous. ILP systems usually do not differentiate input and output variables, while numerical problems often involve differentiation. By itself, ILP cannot even perform such numerical functions as (in)equalities, arithmetic, trigonometry, geometry and linear regression. Of these, regression is the most relevant one in machine-learning problems since numerous real-life models can be approximated by formulas. Regression enables construction of a formula from a given set of examples, but it lacks logical mechanisms for constructing more advanced logical models on top of formulas. ILP and regression are therefore ideal for combining different advantages into an advanced integrated system, which we refer to as first order regression [14].

FORS [14] (First Order Regression System) was one of the first implemented ILP systems capable of handling numerical data. First Order Regression is an integration of (first order logic) Inductive Logic Programming and numerical regression. After first experiments with the implemented FORS system, we started with an advanced algorithm and system 4S.

The paper is organized as follows: In Section 2 we describe definitions of ILP, regression and integration. In Sections 3 and 4 we describe the 4S [6] system as an algorithm and as a fully operational program. In Section 6 we present tests on artificial and real-world problems.

2. ILP, REGRESSION AND FIRST ORDER REGRESSION

For our purposes, ILP can be defined in the following way:

- **B** is background knowledge consisting of a set of definite clauses
- E is a set of examples $E = E^+ \cup E^-$ where:
- $E^+ = \{e_1, e_2, \ldots\}$ are positive examples that are definite clauses, often ground unit clauses.
- $E^- = \{\overline{f_1}, \overline{f_2}, \ldots\}$ are negative examples that are Horn clauses, often ground unit clauses.
- $B \not\models E^+$, where $X \models Y$ denotes semantic entailment of Y by X
- Given B and E, the output is a set of definite clauses $H = \{D_1, D_2, \ldots\}$, that is a complete and consistent explanation of the examples, from a predefined language *L*. H usually satisfies at least the following conditions:
  - Each D_i explains at least one positive example. That is $B \wedge D_i \models e_j \vee e_k \vee \ldots$, where $e_m \in E^+$
  - $B \wedge H \models E^+$
  - $B \wedge H \not\models f$, $f \in E^-$

Linear regression is defined as follows:

- E is a set of examples in the form $y_n = f(x_{1,n}, x_{2,n}, \ldots, x_{m,n})$, where $x_{i,j}$ are real-valued

[1] Jožef Stefan Institute, Ljubljana, Slovenia
{damjan.demsar, matjaz.gams}@ijs.si

parameters.

- Given E, the output is a set of parameters A_0, A_1, …, A_m, that minimizes the sum of squares $\Sigma_n(y_n-(A_0+A_1x_{1,n}+A_2x_{2,n}+\dots+A_mx_{m,n}))^2$

We propose the following integration of ILP and regression:

- **B** is background knowledge consisting of a set of definite clauses
- **E** is a set of examples e_n in the form of ground unit clauses $f(y_n, x_{1,n}, x_{2,n},\dots, x_{m,n})$ where $x_{i,j}$ are either discrete, real-valued or arbitrary terms while y_j are real-valued
- Given **B** and **E**, the output is a set of definite clauses **H** = {$\mathbf{D_1}$, $\mathbf{D_2}$,… } from a predefined language ***L***, that is a good and consistent explanation of the examples. **H** satisfies at least the following conditions:
- $\mathbf{B} \wedge \mathbf{H} \models f(y'_n, x_{1,n}, x_{2,n},\dots, x_{m,n})$
- $\Sigma_n (y_n - y'_n)^2$ is "loosely" minimal
- Each $\mathbf{D_i}$ covers at least a preset number of examples.

3. 4S SYSTEM

4S is a descendant and upgrade of FORS. Some aspects of FORS and 4S are common: both systems have the same input, learning and testing examples and definitions of background knowledge predicates. They both produce a working Prolog program that describes the relation in the input data.

4S and FORS are based on a modified FOIL [27] algorithm (which is a basic covering algorithm), which was developed for searching relations between discrete classes. In order to handle continuous classes efficiently, specific modifications are needed. The main difference is lack of negative examples. We don't have any negative examples at all, so we need to redefine suitability of a hypothesis.

In ILP, a hypothesis is usually suitable when it is complete (covers all positive examples) and consistent (does not cover any negative examples). In our case we still use completeness, i.e. we usually generate hypotheses that cover the whole range of input variables values. On the other hand, since we do not have negative examples, we cannot use consistency to eliminate unsuitable hypothesis (or clauses). To eliminate unsuitable clauses, 4S uses mean squared error combined with a user defined suitability limit. The other major modification in 4S regarding classical inductive logic programming systems is based on a need to handle continuous classes. In 4S, the class variable is predicted with a linear combination of continuous variables.

The 4S algorithm is implemented as a 4s_work procedure (Figure 1), which constructs Hypothesis, i.e. H. It uses beam-search with a user-defined beam, set by default to 10.

```
H = {};
while (|E|>0) {
WorkClauses= {DefaultClause};
NewClauses = {};
repeat
OldWorkClauses=WorkClauses;
NewClauses = RefineClauses(WorkClauses);
WorkClauses=BestOf(WorkClauses∪NewClauses);
until (WorkClauses == OldWorkClauses);
D = BestClause(WorkClauses);
H = H ∪ {D};
E = E - CoveredExamples(D)
}
if (needed) H = H ∪ {FailSafeClause}
```

Figure 1. 4S_work procedure

In each step all current clauses are refined with an addition of a literal (see subsection 3.2).

After a literal is added, we calculate the coverage of the new, refined clause. From the covered examples all numerical parameters are taken (all valid combinations are generated, each representing one example for linear regression), linear regression is performed and the error is calculated. Then the output of the linear regression is transformed into a Prolog equation and added to the clause.

Each generated clause must satisfy preset conditions (e.g. the number of examples covered, variable depth…). Since these conditions are a part of algorithm parameters their explanation can be found in Section 3.4. If the error of the clause is found to be sufficiently small, refinement of the current clauses is stopped and the current clause is added to the hypothesis.

3.1. Differences to FORS

Since we had no working implementation of FORS available, the implementation of 4S, does not base on the code of FORS. but on its performance. In this subsection we list some new mechanisms that are not available in FORS.

The condition that usually stops the specialization of clauses and prevents the overfitting to the learning examples is the demand that every clause covers a minimal number of examples. However, with the addition of new literals the examples expand and split into multiple sub-examples, generated from original ones. For example when we have an example tuple (a, 2), and apply a background knowledge call of a predicate that can for that example return two possible values 10 and 15, the example tuple (a, 2) splits into 2 new example tuples (a, 2, 10) and (a, 2, 15). In this way,

examples can expand into enough sub-examples to satisfy the minimal-examples condition in each clause. In that case overfitting can occur since from only one original example several seemingly different examples can be produced. To fight this problem, we used a mechanism that treats all sub-examples that were created from the same original learning example as one example. In that case every clause must cover a certain number of original examples that were provided in the input data. 4S has therefore two ways of counting examples – one of only original examples and one of virtually created examples.

When 4S is trying to set a suitable constant for the comparison literal (Var < const) there are two possible ways of finding the value the constant should have. The first way is precise but slow. It accepts each possible value that comes into consideration. Then, specialization is performed. A new clause is generated from the old one with the addition of a comparison literal with the selected value. Linear regression is performed and average error is calculated. Only the clause with the lowest average error is accepted.

The second way is much faster but less precise. It is based on the fact that the clause under specialization already contains a linear equation E. Again each possible value for the constant is taken under consideration, and all examples covered by the original clause that satisfy the new condition. The average error is predicted using the equation E. The value with the lowest predicted average error is chosen. The specialized clause is generated, linear regression is performed, and actual error is calculated. However, it is possible that the value producing the lowest average error is not the best choice.

Also parameters that limit the depth of recursion calls and add a possibility of predicting constants and not linear equations were added.

Some of the mechanisms available in FORS were left out – for example the MDL evaluation and the usage of information whether a background predicate is determinate or not (to decide whether to add a fail safe clause or not).

3.2. Language

The hypothesis language of 4S is Prolog. A hypothesis together with the supplied background knowledge predicates is a working program in Prolog. Hypotheses generated by 4S have the following form:

```
func(Y,X1,X2,... ,Xn):-
literal1,
literal2,
...
literalm,
Y is linear_equation,
!.
```

Here Y is a class variable, X1... Xn are input data variables or attributes, and linear_equation is a linear combination of input variables and variables that we get as an output with background knowledge predicates calls.

A literal can be one of:

1. Background knowledge predicate call
2. Comparison of a continuous variable with a constant (Xi >= constant or Xi =< constant)
3. Unification of a discrete variable with a constant (Xi = constant)
4. Recursive call func(Y', X'1, X'2,... ,X'n)

3.3. Input data

The input data consists of:

1. Type declarations that can be one of three kinds: discrete, continuous (real-valued) or term (any Prolog term), e.g. type(real, continuous), where real is the name of the type.
2. Declaration of the target function/predicate, e.g. target(linear('Y','X')).
3. Variable declarations. Variables used in target predicate declaration are bound to specific types, e.g. variable('X', real).
4. Background knowledge predicates declaration includes data about types and input/output modes of the arguments, e.g. bkl(cos(-real, +real), total), where +real means that the first argument is input argument of type real, -real means that the second argument is output argument of type real.
5. Background knowledge predicates definition is actually the Prolog procedure that is used whenever background knowledge predicate is called, e.g. sin(R,A):- R is sin(A).
6. Learning data are the facts about the target predicate. Unlike classical ILP systems, 4S does not use any negative examples. An example of a learning example is lrn(linear(4.30,1)).
7. Testing data, e.g. tst(linear(4.00, 0)).
8. Parameters, e.g set(max_allowed_error, 0.001), See Table 1.

3.4. Parameters

The 4S system uses the parameters in Table 1 to guide search through a hypothesis space. By setting these parameters we adjust 4S to a particular domain. To achieve best results, we must understand the problem domain and performance of 4S. We cannot expect to do better, since it is impossible to design a uniform ILP system that can handle numerical problem

from arbitrary domains without any adjustments. Among possible solutions, setting parameters is the simplest solution. 4S can be applied with parameters set by humans or other programs.

Table 1 Parameters used in 4S

Parameter	Default	Explanation
`min_examples`	10	Minimal number of examples each clause must cover
`absulute_examples`	True	Only original examples and not subexamples count for min_examples
`max_call_depth`	10	Maximal depth of recursion
`max_literals`	10	Maximal length of a clause
`max_clauses`	10	Maximal length of a hypothesis
`max_variable_depth`	10	Maximal depth of variables
`max_allowed_error`	$1*10^{-18}$	Clause with an average error of less than that is always accepted (deemed as perfect)
`min_improvement`	0	How much specialized clause must be better from its predecessor
`allow_recursive`	False	Whether recursion is permitted
`extensive_comparing`	True	Is slower but more precise selection of comparison constant allowed
`sv_decomposition`	True	Is robust but slow SV matrix decomposition rather than fast but sensitive LU decomposition used for linear regression
`allow_regression`	True	Is regression permitted
`max_regression_vars`	-1	How many variables can be used in a linear equation (-1 means all)

4. PROBLEMS WITH INTEGRATION

The combination of recursion and linear regression causes a feedback deadlock: the linear regression process needs an instantiated result of a recursive call, whereas the recursive calls require the result of linear regression. This occurs in all functions of the form *f(n) = g(f(n-1))* where *g(m)* is a linear function.

There are a couple of ways to resolve this mutual dependency, but each introduces additional difficulties. The first way is to expand recursive calls until we reach a bound, and perform an evaluation with already instantiated returned recursive parameter. This approach is sometimes referred to as lazy or late bounding or instantiation [29]. The problem with this solution is that it demands capabilities of symbolic computing. Furthermore, there is no guarantee that this process would produce enough examples for linear regression, and no guarantee that the solution would be reasonable.

The second way to deal with the deadlock is to use the learning examples in a non-circular way. In this way, only learning examples that have known values for recursive calls (known from learning examples, or previously accepted clauses) are taken for input into the linear regression process. However, this demand is difficult to meet in real-world problems – it demands that the available learning examples cover precisely those cases that the recursive calls require. Clearly, only synthetic problem domains will have such nice properties. Since we wanted 4S to be applicable to real-world problems, we did not find this solution to be acceptable.

The first simple solution implemented in 4S is to limit the recursion and/or complexity depth. In case of too deep a recursion, the program discards the current clause and with another one.

To cope with recursion, we introduced a limited set of parameters that enable solving a specific set of predefined recursive forms. For example, solving n! can be found by limiting the linear equation to the simple form of $y = x_i$. This approach certainly cannot solve more complex cases of recursion, but can handle simple ones. In a similar way, it is possible to tune 4S so that it finds various predefined special forms. The problem with this approach is that with more complex expressions, the tuning and related hand coding gets more and more time consuming. On the other hand, we expect that a couple of simple cases suffice for expected real-life problems. In face of the complexity of the task, this seems a sensible approach.

Putting aside potential mutual deadlock between recursions and regression, other ILP mechanisms are intact in 4S. If we discard regression, 4S constructs recursive ILP programs in the same manner as other ILP systems.

5. RELATED WORK

Systems like LAGRANGE [9] and GOLDHORN [19] discover equations from numerical data, but since they are not ILP systems they lack its expressive power. There are several approaches adding numerical capabilities to ILP systems: restricting the hypothesis language to logical atoms [25], using built-in definitions for inequalities [4], [26], [27], using transformations of data to propositional level [21], using background knowledge for regression predicates [23], [29], [31] and adding regression capabilities to the ILP engine [14]. Several systems [3], [17], [18] also produce working prolog programs, that are capable of predicting numerical values, but do this by constructing a regression tree first and then translating it into a Prolog program. By taking this approach some of the power of

hypothesis language is lost.

FORS [14] is the system that first used first order regression. It combined the expressive power of ILP with linear regression. It showed that first order regression is able to produce results. While 4S is based on principles laid by FORS, it is a new system. Several new techniques were developed and some old ones from FORS, that did not perform as expected, were eliminated. In 4S we also included extensive comparing of continuous variables and the option to force the system to count only basically different examples to satisfy the minimal number of examples condition (as described in Section 3.1).

The system described in [29], [30] as well as a system described in [31] are also capable of handling numerical domains using background knowledge for comparisons, and other numerical operations including regression. Since the main goal of the designers was to keep all capabilities of the original ILP engine, several compromises had to be done. In case of numerical domains, those systems cannot navigate efficiently through the search space before predicates predicting numerical values are added to the clause. To ease the navigation through the search space and to produce meaningful clauses, the system requires the user to provide a refinement operator, cost calculation and pruning. This certainly enhances the performance of the system compared to 4S, but it also significantly decreases the ease of use and practical applicability to real-live problems. On the theoretical-ability scale, those systems outrank 4S, but judging from results presented in [29] and [30] one gets the impression that 4S is at least competitive in terms of practical use, as it does not require constant involvement by the user.

6. TEST DOMAINS

6.1. Steel grinding domain

The steel grinding domain data describes the relationship between the roughness of a steel workpiece and the sound produced by grinding that workpiece [13]. The idea is to enable automatic control of the grinding process. An example is termination of the process when performance becomes unsatisfactory [12]. Since control decisions can be quickly deduced from the roughness of the workpiece, our task was to predict roughness from the sound of grinding.

The data was collected during an experiment where the workpiece was fitted with an acceleration sensor. That data was then spectrally analyzed, and the following data was available to 4S:

- size of the area of whole spectrum
- frequency of the maximum peak
- frequency of the middle of the spectrum

During the experiment the process of grinding was stopped and the surface roughness measured 123 times. Background knowledge predicates enabling 4S to compare variables were given (<= and >=).

Table 2 Relative error of 4S on Steel grinding.

min no. of examples	all var. test	two var. test	one var. test	"no" var. test
1	0.633	0.584	0.529	0.840
2	0.656	0.594	0.518	0.629
4	5.403	1.877	1.257	0.595
6	0.550	0.470	0.536	**0.624**
8	0.509	0.461	0.558	0.634
10	0.508	0.452	0.545	0.657
12	0.509	0.476	0.544	0.643
14	0.508	0.473	0.547	0.656
16	0.500	0.465	0.544	0.660
18	0.500	0.453	0.558	0.662
20	**0.441**	**0.451**	0.565	0.662
30	0.443	0.455	0.559	0.655
50	0.501	0.516	**0.499**	0.825
80	0.457	0.464	0.674	0.965
100	1	1	1	1

Table 3 Relative error of FORS on Steel grinding.

min no. of examples	lin reg test	no reg test	lin reg, MDL test	no reg, MDL, test
1	0.74	0.79	0.73	0.62
2	9.15	0.69	0.76	0.60
4	1.01	**0.55**	0.71	0.58
6	0.81	0.60	0.66	**0.54**
8	0.78	0.58	0.63	0.59
10	0.72	0.56	**0.62**	0.56
12	0.69	0.63	0.63	0.64
14	**0.64**	0.60	0.62	0.62
16	0.70	0.57	0.70	0.57
18	0.71	0.62	0.71	0.62
20	0.74	0.64	0.74	0.64

Experiments were conducted in the form of 10 different sets consisting of 70% learning and 30% testing examples (same sets were used by FORS) and all combinations of minimal number of (original) examples in a clause (possible values 1, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 30, 50, 80, 100) and the type of linear equation (allowed are all variables, 2, 1 and no variables - a prediction of a constant). From the results on the learning and test data relative (compared to the prediction of an average learning value) average error was calculated. The results are presented in Table 2.

If we compare the results of experiments made with 4S in Table 2 and FORS [15] in Table 3 we can see that when regression is not permitted FORS (with or without MDL) performs better than 4S. However with linear regression even when using MDL FORS overfits the data. The clause with the best results by 4S (made using linear

regression with at least 20 examples covered by each clause) is small and simple, just what we would hope for.

6.2. Finite element mesh domain

Finite element method is used in engineering where stress in structures needs to be calculated. The structure is divided into many simpler elements. A set of linear equations must be solved to calculate the deformations of the elements. Since the number of the elements affects the computational time and the accuracy of the solution (both increase with the number of the elements) a suitable compromise in the partition into elements must be found. In recent years a number of very successful mesh generation algorithms was reported [20]. We should not be surprised to see a significant gap between 4S and specialized algorithms. However, our aim is to compare 4S' ability to generate meshes to that of other ILP systems.

Partitioning usually cuts the edges of the structure into multiple parts. Then the mesh dividing the elements is automatically generated, from the start points on the edges. So the task is to find the number of partitions for each edge of the structure. Since the edges can have different attributes and their partitioning is also depended on other edges and their attributes, inductive logic programming is the natural choice for solving this problem. Others have tried it before and it turned out to work well [5], [7], [8], [14], [16].

In our experiments the data from [8] was used. It describes five structures labeled from A to E. Each structure has from 28 to 96 edges. For each edge a number of partitions was determined and computationally verified by an expert [8]. Our target was a relation that describes the number of elements for each edge.

Experiments consisted of all combinations of parameter values for `min_examples` (2, 5, 10) - only original examples were counted, and `max_literals` (2, 4, 6).

At all times the maximal variable depth was set to 2, which forced 4S to use only the edges at most two steps away from the target edge (neighbors, opposites, neighbor's neighbors…). For each possible setting five experiments were made. Every time one of the structures was set aside for testing and the other four were used for learning. Results of the experiments can be found in Table 4, where a prediction is deemed correct if 4S prediction was correct after rounding to the nearest integer. If the prediction was exactly between two integers (e.g. 1.5) false prediction was assumed.

Figure 2 shows the hypothesis constructed with at least 10 examples and no more than 4 literals in a clause from structures A, B, C, D and tested on structure E. The induction took less than 15 seconds on 700 MHz Pentium III computer running Windows 2000.

We compare results of 4S with results of FOIL [27], mFOIL [10], GOLEM [24], MILP [16], FFOIL [28], FORS [14] and CLAUDIEN [5] in Table 5. Results for FOIL and FFOIL were taken from [28], for GOLEM and MILP from [16], for mFOIL and FORS from [14] and for CLAUDIEN from [5]. We can see that 4S performs better than FOIL and mFOIL, slightly better than GOLEM and CLAUDIEN, similar to MILP, and FORS and worse than FFOIL.

Table 4 Results of FEM domain experiments

Minimal number of examples	2			5			10		
Maximal number of literals	2	4	6	2	4	6	2	4	6
Structure used for testing									
A (55 edges)	22	22	22	22	20	19	21	21	22
B (42 edges)	10	12	12	11	10	7	14	10	13
C (28 edges)	5	10	10	6	5	7	8	8	8
D (57 edges)	16	21	22	14	20	12	14	21	18
E (96 edges)	20	22	22	11	6	3	9	28	4
Total correct (from 278)	73	87	88	64	61	48	66	88	65

Table 5 Comparison to other systems on FEM domain

Structure	FOIL	mFOIL Lapl	mFOIL m=0	GOLEM	MILP	FFOIL	FORS	CLAUDIEN	4S
A	16	23	22	21	21	21	22	31	22
B	9	12	12	12	12	15	12	9	12
C	8	9	9	10	11	11	8	5	10
D	12	6	6	16	16	22	16	19	22
E	16	12	12	21	30	54	29	15	22
Sum	61	62	61	80	90	123	87	79	88
%	22	22	22	29	32	44	31	28	32

7. DISCUSSION

Experiments have shown that the 4S system is capable of combining ILP and numerical regression. There are two major problems: recursion and regression together cause instability in the regression process (when recursion is used, some input arguments in the regression are dependent on output arguments). The other problem is that time complexity, which is already high in ILP, has additionally increased due to additional numeric capabilities. Further research is needed to overcome these problems and move towards a fully applicable system.

The severity of both problems can be reduced. The time complexity calls for better heuristics and optimization of the code, for example in the ILP part where construction of equivalent clauses can be avoided. The problem with recursion and

regression demands more effort. Usually ILP systems require a complete set of examples to successfully induce recursive definitions. However, a complete set of examples can be expected only in artificial domains and not in real world domains. That is why we need to create a process that combines available examples and simulates the missing examples.

In summary, 4S performs as well as other academic systems dealing with reasonable problem domains. If we compare accuracy with FORS, 4S performs similarly or even marginally better on some domains, while being more flexible and tunable to a particular domain. In comparison with more complex systems that demand inclusion of additional program code, setting parameters in 4S is far less demanding even for non-specialized users.

REFERENCES

[1] A. Appice, M. Ceci, and D. Malerba. Mining model trees: A multi-relational approach. In T. Horvath and A. Yamamoto, editors, Inductive Logic Programming, 13th International Conference, ILP 2003, volume 2835 of LNAI, pages 4_21. Springer-Verlag, 2003

[2] Bratko, I. and Džeroski, S., "Engineering applications of ILP", *New Generation Computing 13* (pp. 313-333), 1995.

[3] Blockeel, H., *Top-down induction of first order logical decision trees*. PhD Thesis, Katholike Universiteit, Leuven, Belgium, 1998.

[4] Camacho, R., "Learning stage transition rules with Indlog", *Proceedings of The Fifth International Workshop on Inductive Logic Programming (ILP-94)* (pp,273-290), Bad Honnef/Bonn, Germany: Gesellschaft für Mathematik und Datenverarbeitung Sankt Augustin, 1995.

[5] Dehaspe, L., van Laer, W., and De Raedt, L., "Applications of a logical discovery engine", in *Proceedings of the Fourth International Workshop on Inductive Logic Programming (ILP-94)*, GMD-Studien Nr. 237, 1994.

[6] Demšar, D., *Dealing witn Numerical Problems Using Inductive Logic Programming*, Master's Thesis, University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia, in Slovene, 1999.

[7] Dolšak, B., and Muggleton, S., "The application of inductive logic programming to finite-element mesh design", Stephen Muggleton. (ed.), *Inductive Logic Programming*: Academic Press, 1992.

[8] Dolsak, B., Bratko, I. and Jezemik, "A, Finite element mesh design: An engineering domain for ILP application", *Proceedings of the Fourth International Workshop on Inductive Logic Programming (ILP-94)* (pp. 305-320), Bad Honnef/Bonn, Germany. Gesellschaft für Mathematik und Datenverarbeitung Sankt Augustin, 1994.

[9] Džeroski, S., and Todorovski, L., "Discovering dynamics: from inductive logic programming to machine discovery", *Journal of Intelligent Information Systems*, 4.89-108, 1995.

[10] Džeroski, S., *Handling noise in inductive logic programming*, Master's thesis, University of Ljubljana, Faculty for Electrical Engineering and Computer Science, Ljubljana, Slovenia, 1991.

[11] Džeroski, S., *Numerical Constraints and Learnability in Inductive Logic Programming*, PhD thesis, University of Ljubljana. Faculty for Electrical Engineering and Computer Science, Ljubljana. Slovenia, 1995.

[12] Filipič, B., Junkar, M., Bratko, I. and Karaliè, A, "An application of machine learning to a metal-working process", in *Proceedings of IT-91*, Cavtat, Croatia, 1991.

[13] Junkar, M., Filipič, B. and Bratko, I., "Identifying the Grinding Process by Means of Inductive Machine Learning", *Computers in Industry*, 17(2-3), 147-153, 1991.

[14] Karalič, A. and Bratko, I., "First Order regression". In *Machine Learning*, Volume 26, Numbers 2/3. Kluwer Academic Publishers, 1997.

[15] Karalič, A., *First Order Regression*. PhD thesis, University of Ljubljana, Faculty for Electrical Engineering and Computer Science, Ljubljana, Slovenia, 1995.

[16] Kovačič, M., *Stochastic Inductive Logic Programming*, PhD thesis, Universty of Ljubljana,, Faculty of Electrical Engineering and Computer Science, Ljubljana, Slovenia, 1995

[17] S. Kramer. Structural regression trees. In Proceedings of the National Conference on Artificial Intelligence, 1996

[18] S. Kramer. Relational Learning vs. Propositionalization: Investigations in Inductive Logic Programming and Propositional Machine Learning. PhD thesis Vienna University of Technology, Vienna, Austria, 1999

[19] Križman, V., *Avtomatic discovery of structure in dynamic systems models*, PhD Thesis, University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia, in Slovene, 1998.

[20] Lavrač, N. and Džeroski, S., *Inductive Logic Programming - Techniques and Applications*, 1994.

[21] Miller, G. L.,Teng, S. H., Thurston, W. A., Vavasis, S. A. "Graph Theory and Sparse Matrix Computation", *The IMA Volumes in Mathematics and its Applications 56*, (pp 57-84), Springer-Verlag, 1993.

[22] Mizoguchi, F. and Ohwada, H., "An inductive logic programming approach to constraint acquisition for constraint-based problem solving", *Proceedings of the 5th International Workshop on Inductive Logic Programming (ILP-95)*, (pp. 297-322), Katholike Universiteit Leuven, Heverlee, Belgium, 1995.

[23] Muggleton S., Page D., "Beyond first-order learning: inductive learning with higher order logic", *PRGTR 13-94*, Oxford University Computing Laboratory, Oxford, 1994.

[24] Muggleton., S. and Feng, C. "Efficient induction or logic programs", *Proceedings of the First Conference on Algorithmic Learning Theory*, Tokyo, Japan, 1990.

[25] Page, C. D., Frisch, A. M., "Generalization and learnability: A study of constrained atoms", in: S. Muggleton (Ed.), *Inductive Logic Programming* (pp 29-56), Academic Press, London, 1992.

[26] Quinlan, R. and Cameron-Jones, R. M., "Efficient top-down induction of logic programs", *SIGART Bulletin 5* (1) 33-42, 1994.

[27] Quinlan, R., "Learning logical definitions from relations", in *Machine Learning* 3(5), 1990.

[28] Quinlan, R., "Learning First-Order Definitions of Functions", *Journal of Artificial Intelligence Research*, volume 5, (pp 139-161), 1996.

[29] Srinivasan, A. and Camacho, R., "Numerical reasoning with ILP system capable of lazy evaluation and customized search", *The Journal of Logic Programming*, Volume 40, Numbers 2/3, 1999.

[30] Srinivasan, A., "Experiments in numerical reasoning with ILP", Michie, D., Muggleton, S., and Furukawa, K., (ed.), *Machine Intelligence* 15, Oxford University Press, 1996.

[31] Srinivasan, A. The Aleph Manual. Technical Report, Computing Laboratory, Oxford University, 2000,,

Transition of E-Business to M-Business Driven by Wireless Communication

Barjis, Joseph; Barjis, Isaac

Abstract –due to wide implementation and integration, e-business has established a new paradigm for companies to capitalize on the potential of information & communication technologies mainly operating on the Internet and Web technologies. Current trend and future challenge of e-business is to conduct mobile business and bring the trade and service to where potential customers are located, and goods are produced, consumed, delivered and provided. This article provides an overall review of Mobile Businesses, as a future trend of e-businesses, with emphasis on its supporting mobile technologies and wireless networking. The discussion starts with the concept of mobile business that projects the opportunities, motivations and needs for e-business. Following this discussion, the article studies the current status of mobiles business, key hardware solutions and software business applications available on the market. Furthermore, the article discusses different mobile devices, communication infrastructure, supporting networks and other crucial components that make the businesses mobile and provides the ability to the business to be conducted at anytime and anywhere. Finally, an extended discussion is focused on issues and future developments of m-businesses along with some recommendations, and suggestions regarding mobile business (mBusiness).

Keywords: e-business, mobile business, m-business, mBusiness, wireless business, mobile devices, 3G, wireless networks

1. OVERVIEW

Increasing implementation and integration of e-business by enterprises has established a new paradigm for businesses to capitalize on the potential of information & communication technologies mainly operating on the Internet and Web technologies (Pankaj, *et al.* 2004). Current trend and future challenge of e-business is to conduct business that is mobile and bring the trade and services to where potential customers are located or goods are produced, consumed, delivered and provided. This paper is mainly focused on wireless communication and the web that facilitate e-business and conventional business to become flexible and mobile.

Mobile Business is outcome of the Advanced Communication Age and driving force of the new economy. Therefore the discussion of mBusiness is greatly aligned with the study of underlying information and communication technologies and its impact on the new economy. The increasing pace of innovations in information and communication technologies in recent years has opened a wide spectrum of new opportunities and challenges in the business industry. These opportunities demand dramatic shift towards mobility in almost every aspect of life such as retail, education, entertainment, health care, etc. Rapid developments in wireless communication technologies, mobile devices, high speed transmission mediums and broad bandwidth technology paved the road for transforming human activities towards mobility. The most noticeable impact of these evolving technologies can be seen in businesses that are preparing for another revolutionary change. First businesses have gone through transformation from traditional business to an electronic business (e-business) and now they have to addapt towards mBusiness.

Undoubtedly, for the coming years, mBusiness will remain practice of interest for industry, researchers, enterprise managers and society as a whole. This will be the business style of the 'Age of Communication' inspiring managers and enterprises for serious shift. May be there are enterprises that will experience another era of Business Process Reengineering, or mBusiness Process Engineering (mBPR)!

In order to understand how different sectors could benefit from mBusiness, it is worthwhile to mention some examples and instances where mobility has been successfully making its breakthrough. These different examples are aimed to provide an idea about the breath, depth and diversity of mobility and mobile business:

Sale and Marketing: Retail, wholesale, and mass distribution centres use mobile business environment to communicate simultaneously with branches in different locations for all business services (goods delivery, shelves refill, inventory control, warehouse management, transport and logistics).. Mobile devices help to track goods

Manuscript received July 1, 2005, contact author Joseph Barjis

Joseph Barjis is a faculty in the Department of IT, Georgia Southern University, P.O. Box 8150, Statesboro, GA 30460, U.S.A. Email: JBarjis@georgiasouthern.edu

Isaac Barjis is a professor in New York City College of Technology, City University of New York, Room P313, 300 Jay Street, Brooklyn, NY 11201, U.S.A. Email: IBarjis@CityTech.Cuny.Edu

delivery and movement from anywhere at any point of time. Large businesses such as Wal-Mart are moving towards RFID (radio frequency identification) to pursuit a better inventory control. Wal-Mart's initiative to adopt RFID technology requires top 100 suppliers to comply with RFID requirements that will create complete automatic mobile environment in dealing with millions of boxes to be daily tracked, recorded and entered into the system.

Health care: Hospitals in general and modern medical practices in particular are adapting towards mobile health care delivery. Computer-based Patient Record, also referred to as Electronic Medical Record, is a system that provides mobile working environment for physicians, staff and managers of medical practices. Each physician carries a handheld computer to access patients records, digital diagnostic images (X-rays, magnetic resonance imaging scans, ultrasound images, computerized tomography scans, digital subtraction angiography images, positron emission tomography scans), and digital surgery videos. Furthermore physicians can share and discuss digital diagnostic images and digital video clips with specialists from other hospitals across the continent, coordinate remote operations, etc. Physicians could use their handheld computer remotely to obtain information on a particular drug (side effects, drug interactions, drug information, etc.) prior to issuing a prescription. Mobile facilities can dramatically increase productivity and quality of health care services.

Education: Educational institutions implemented wireless networking environment to provide students with the flexibility to access campus resources and download academic applications at their convenient time and desired location (lab, classroom, library, cafeteria, campus garden or while watching campus games). Not being tight to the lab hours and classrooms, students are given more flexibility and opportunity to pursue their education, which in turn increases quality of education. So, campuses are going mobile within the campus area.

These three small examples illustrate different ranges of mobility within a building, within an enterprise, within a town used for a wide range of activities. These examples help analysts in extracting some important characteristics of mBusiness including range of functionality and types of mobile devices (wireless laptops, tablet PCs, smart phones, etc.). According to some authors, application of mBusiness can be distinguished as 'macro' applications in outdoor settings or 'micro' applications in indoor environments such as hospitals, libraries, hypermarkets, etc. Likewise the underlying wireless networks that support the application of mBusiness can be distinguished by its span as a local, regional or global mBusiness.

With this brief introduction, this section is concluded and the rest of the article will discuss different aspects of mBusiness, its technological components and elaborate on various facets of mBusiness.

2. OBJECTIVE OF M-BUSINESS

In general, the main concept of mBusiness is about moving enterprise's critical business to the point of sale and service or even closer, to the point of consumers.

Like its predecessor (electronic business), the concept of mobile business has been used to a wide range of application areas comprising from communication to consumer transactions and corporate services (Vos & de Klein, 2002). However, the real potential of mBusiness is much broader than merely providing service, for sale and delivery of products. A well engineered, designed and integrated mBusiness can support not only conducting business but also adds collaboration, coordination, instant communication and management features to the business. Being based on most advanced information and communication technologies, mBusiness aims to be more productive than a traditional business or a business that is supported by network of computers

Evolving from mainframe and wired network eras, mBusiness is the leading edge of the new generation (3G or third generation) of wireless networking that aims to adapt the best business and management practices, standards and styles.

2.1. M-Business Drives

The m-commerce revolution has already begun (Dholakia and Rask, 2004). The driving source behind the mBusiness can be sought from two different perspectives: from the business perspective the main motivations that pushes e-business towards mBusiness are competition for flexibility in conducting business, extending functionality and service to the business point, convenience of employees and comfort of consumers, better satisfaction, quality improvements, personalization and localization of business. In addition to aforementioned benefits in the core of the driving forces that pushes businesses toward mBusines is revenue increase and market gain.

From the technological perspective the revolutionary progress in wireless networking, information technologies and mobile devices with high-speed wireless communication has provided new infrastructure for business that resulted in more mobility.

2.2. Definition of M-Business

Mobile business is the business of future which is based on wireless infrastructure, using mobile devices that could bring critical business to the point of service and sale, with the aim of higher

productivity in wide economic sense. Depending on a focal point and perspective, mBusiness can be defined in quit different ways.

Mobile Business study and application attracted attention of many outstanding authors, scholars and researchers (e.g., Deitel *et al.*, 2001, 2003; Vos and Klein, 2002; Paavilainen, 2002; and others). In the abundance of available definitions used by different authors, it is quite challenging to find a unique definition, however the definition given by (Kalakota & Robinson, 2001) could be cited here as an example: mBusiness is "the application infrastructure required to maintain business relationships and sell information, services, and commodities by means of the mobile devices". This is one of definitions that characterizes mBusiness. Different definitions are given from different perspectives. However, regardless of definition perspective, what is generic about mBusiness is that mBusiness encompasses three essential components: Wireless networking technology (3G Networks, WLAN, WWAN), mobile devices and improved business practices (procedures). The last component is a key component of mBusiness. If the first two are facilitators, the third one is the main objective of mBusiness.

3. CURRENT STATE OF M-BUSINESS

Although application of mBusiness started not too long ago, this opportunity very quickly attracted enterprise managers, industry leaders, researchers and authors. Today a number of periodicals are adapted or founded on mBsuiness, numerous monographs are published (e.g., Paavilainen, 2002; Vos and Klein, 2002; Kalakota and Kurchina, 2004; Sadeh, 2002), annual conferences such as ICMB (International Conference Mobile Business), and a wealth of internet based resources reporting studies, results, examples and models of mBusiness. In addition, tens of IT and business consulting companies refocused their activity from electronic commerce and electronic business towards mBusiness.

Studying the overwhelming opportunities and increasing demands in mBusiness, Kalakota (2005) in his work "Mobile Business: Vision to Value" emphasizes how rapidly the emerging technologies change the way enterprises conduct their business and how dramatically mBusiness is replacing traditional business and e-business. As a result of these changes, the author states, that the question of "Should I do mobile business?" shifted to "How can mobile create business value?"

Enterprises and business worldwide are implementing mobile business solutions to accelerate their business cycles, increase their productivity, reduce the operating costs and extend their enterprise infrastructure.

The need for going mobile turned into serious competition between leading companies that provide wireless infrastructure, application solutions and mobile devices for mBusiness. Today, the following market leaders are among the top providers of applications and wireless networking infrastructure for mobile businesses.

On the software market, SAP as the world's largest inter-enterprise software company took a pioneering initiative in providing software packages for different types of mBusiness that is listed below. In getting leading position in mobile business solutions, Microsoft is adding mBusiness features to Windows.

On the wireless market, Cingular and Verizon are pioneering in providing modern wireless networking services by introducing and expanding 3G networks in major cities and metropolitans (Segan, 2005).

On the mobile devices market, Siemens (www.siemens.com), Nokia and other leading providers are introducing mobile technologies that significantly boost competitiveness of businesses.

SAP mobile business solution set includes ready-made applications that provide access to the corporate information and processes anytime, anywhere allowing use of a variety of mobile devices. Among various software packages, SAP provides mobile business applications (SAP, 2005):

SAP Mobile Time and Travel – this package gives mobile workers access to time sheet and travel management functionality.

SAP Mobile Sales – this package provides a solution for salespeople who need to perform their tasks quickly and productively.

SAP Mobile Service – this package enables field service engineers to react quickly to customer needs.

SAP Mobile Asset Management – this package allows in-house service engineers to access relevant business processes anywhere, anytime.

SAP Mobile Procurement – this package enables mobile workers to manage the entire procurement process, from price comparison to ordering.

4. M-BUSINESS FRAMEWORK

As Pelkonen and Dholakia (2004) states, mBusiness is a complex network of business relationships comprising telecommunications service providers, manufacturers of mobile devices and various third-party value-adding companies. For more accurate understanding of the mBusiness framework, one may think of mBusiness as a two-level framework (as shown in figure 1), where the upper level is the business level (processes, procedures, standards) and the bottom level is the IT infrastructure (hardware and software) that supports business to carry out its mission and tasks and sets business in motion.

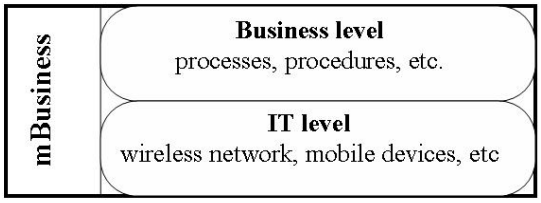


Figure 1. mBusiness IT infrastructure

From a high level perspective, one can consider that the business level is a variable and the IT level is relatively constant. Business level is variable, because it represents different types of businesses, whereas the IT infrastructure, in its general functionality, will remain the same for most types of mBusiness but differ in constituent details or nuts and bolts (range, size, structure, configuration, potential, complexity etc).
IT infrastructure of mBusiness includes hardware (wireless networks and mobile devices) and software (mobile business, office and enterprise applications). The hardware or mobile devices and wireless networks will be discussed further in subsequent sections.

4.1. Mobile devices

One of the main characteristics of mobile devices used in eBusiness is the communication or networking facility and capacity to receive, transmit and process different types of data (text, audio, video) at high rate. In addition to high speed data exchange, ease of handling, portability and size are important in these devices. Furthermore, these devices must be able to download essential business, office and enterprise applications and have sufficient memory to run them.
Currently devices used in mBusiness are wireless handheld computers, laptops, PDAs, tablet PCs, smart phones, Blackberry, etc. These devices should be capable to run high-performance business, office and enterprise applications such as multi-media, full-motion video, wireless teleconferencing and use connection over wireless networks using Wi-Fi, GPRS, Bluetooth or other advanced connections. For a better idea and distinction between ordinary mobile devices and business quality devices and better idea about business quality mobile devices, one could look at the features provided with Nokia 9500 that have a full set of critical business tools, full keyboard, with a large memory capacity and versatile network connections (Nokia, 2005):
Browse the Internet in rich full color, on a wide, easy-to-read screen. Work with office documents - not just email and memos, but presentations and databases too. Get them from your corporate network via Wireless LAN or EDGE for fast mobile access. Keep your Personal Information Management data in sync and up-to-date calendar and contacts - with PC Suite and SyncML, you can easily exchange data between your Nokia 9500 Communicator and a compatible PC.

4.2. Wireless Networks

Although not in mass application yet, but in the future mBusiness will be operating on third generation wireless networks, 3G, that provide high speed download and upload rate. The speed of transmission in these networks using mobile devices is at the level of DSL connections. Because data costs on 3G networks are lower than on traditional networks (Solheim, 2005) more and more enterprises will shift into using 3G networks as main infrastructure.
Currently, mBusinesses are based on different types of networking technology:
Personal area network using Bluetooth technology that enables short-range device-to-device wireless connections within a small space (office, a desktop, a personal space).
Wireless local area networks using Wi-Fi technology, based on IEEE 802.20, 802.11 standards, which supports a wireless connection to a network from inside a home or from a hotspot in a building, campus, or airport.
Wireless metropolitan area networks using WiMax, based on IEEE 802.16 standard, will enable any remote worker to make a wireless connection anywhere in a range up to 50 kilometers.
Wireless wide area networks or 3G networks provide the highest available bandwidth for mobile devices. Although, theoretical rate of transmission is 2Mb, but for practical purposes the transmission speed is like DSL which enables users to download text, audio, video, web contents, and send or receive e-mail while in motion.
4G technologies promise to integrate different modes of wireless communications – from indoor networks such as wireless LANs and Bluetooth, to cellular signals, to radio and TV broadcasting, to satellite communications.

5. ISSUES OF M-BUSINESS

In the previous sections we discussed advantages and benefits of mBusiness, however, enterprise managers and business owners ought to be aware of issues and challenges before undertaking the transformation initiative.
It would not be less than an illusion, if consider transition from traditional or eBusiness to mBusiness as just shifting from wired environment to wireless communication and networking. The transformation into mobile business is not just about moving from wired environment to wireless, from desktops to handheld computers or from office to field. Also the challenge of adapting mBusiness is not only about acquisition and implementation of best wireless technology or awareness of emerging mobile technology, but it is rather a multi-

dimensional issue where technology is only one facet of it. Challenges that may require more profound study are transformation of the business and enterprise, shift in the mind-set (e.g using mBusiness as a different way of doing business rather than considering mBusiness as an implementation of best business practices), improved procedures, higher quality, etc. Feeny (Feeny, 2001) suggests rethinking business processes, models, customer relationship and the whole supply chain in pursuit of highest productivity and customer satisfaction, should be part of taking advantage of e-opportunities. Thus, the challenge of going mobile is much complicated than application and implementation of mobile technology.

Different sectors of business may have their own features and attributes, however some of the common issues for enterprises to carefully study while embarking on competing for faster implementation of mBusiness are:

From business's perspective:

How well mBusiness opportunities are studied?

How much Business Process Reengineering, ERP system changes, customer relationship management is required to go mobile?

What they expect from mBusiness, direct profit or quality of service? How qualitative values can be turned into quantitative values?

How much patience they have before harvesting first fruits of benefit?

Mobile business should not be considered as an immediate way for profits.

Are the employee ready and trained to go mobile, or they will resist against?

From technology perspective:

What will happen with existing IT infrastructure?

How carefully the problems of interfacing, integration and legacy systems are studied?

While going mobile, an important challenge is security of mBusiness, is this issue studied?

How well the connectivity and management of mobile devices, security and updates issues are studied?

From consumer's perspective:

What is the impact of mBusiness on consumers?

Is the transition for consumers straightforward or painful?

Transition cost: Wither to mobilize few employees or the whole enterprise. How much will it cost for an enterprise with tens of thousand employees?

What is the price and benefit of a mobile employee (mEmployee)?

For each of the mentioned categories, list of such questions can be much longer than shown here. These are just some of the issues not including public, political and legislatives related issues.

6. CONCLUSION

This article provided a brief overview of mBusiness, opportunities that mBusiness opens and challenges accompanying these opportunities.

The article also discussed mBusiness in connection to its underlying technology. In this part different technological components of mBusiness were introduced and discussed. Along with technology, the article provided brief information about some mBusiness software solutions.

6.1. Future of M-Business

With the arrival of 3G wireless networks on market and development of powerful mobile devices, shift in the business and the needs for more ubiquity, in few years you will be: participating in and watching TV quality business meeting of your corporation on your laptop while in the air (flying back home); coordinating and managing your business while enjoying on the beach; evaluating and managing a project and assigning new tasks while interacting with nature; able to process loan applications while on the road; conducting other serious business activities on the streets and fields. What is most amazing, neither your manager, nor employees or colleagues will realize that all these times you were miles away from your office, because mBusiness will provide you with facilities like being all the time at your office.

In short, you will be carrying your office or enterprise in your briefcase because you will be doing mBusiness. Your office and business will be residing at your fingertips, without being at your office. If all the business services could be managed from the comfort of home or anywhere else then the need to maintain big office buildings will also become questionable. All these changes will be available in the near future because mBusiness is the business of future, but at the present time, there are more than enough challenges and issues in order to adapt and transform traditional way of doing business to mBusiness. We will see how mBusines can increase the productivity and comfort to the employees, employers and consumers in coming years.

REFERENCES

[1] Deitel, H., Deitel, P., & Steinbuhler, K. (2001). e-Business and e-Commerce for Managers. Prentice Hall, Upper Saddle River, New Jersey

[2] Deitel, H.; Deitel, P.; Nieto, T.; Steinbuhler, K. (2003). The Complete Wireless Internet and Mobile Business Programming Training Course. Prentice Hall PTR.

[3] Dholakia, N.; Rask, M. (2004). Configuring M-Commerce Portals for Business Success. Nan Si Shi, ed., Mobile Commerce Applications, Hershey PA: IRM Press, 2004, pp. 76-94

[4] Feeny, D. (2001). Making Business Sense of the E-Opportunity. Sloan Management Review, 42(2), 41-51.

[5] ICMB: The Fourth International Conference on Mobile Business (ICMB), July 11th -13th, 2005,Sydney, Australia: http://www.mbusiness2005.org/

[6] Kalakota, R. (2005). Mobile Business: Vision to Value. On http://www.kalakota.com/speakingtopics/mbusiness.htm

[7] Kalakota, R. and Robinson, M. (2001). mBusiness: The Race to Mobility. McGraw-Hill Publishing Company, New York.
[8] Kalakota, R.; Kurchina, P. (2004). Mobilizing SAP: Business Processes, ROI and Best Practices. Mivar Press, Inc. B·
[9] Nokia, accessed in March 2005 on: http://www.nokia.com/mobilebusiness/americas/
[10] Paavilainen, J. (2002). Mobile Business Strategies. Addison-Wesley.
[11] Pankaj B., L.S. Murty, T.R. Madanmohan, R. K Bandi (2004). Project Mi-Net - An Inter-Organizational E-Business Adoption Study. Annals of Cases on Information Technology. Volume 6. Idea Group Inc.
[12] Pelkonen, P.; Dholakia, N. (2004). Understanding Emergent M-Commerce Services By Using Business Network Analysis: The Case of Finland. Nan Si Shi, ed., Wireless Communications and Mobile Commerce, Hershey PA: IRM Press, 2004, pp. 105-131
[13] Sadeh, N. (2002). M Commerce: Technologies, Services, and Business Models. Wiley; 1st edition.
[14] SAP, accessed in March 2005 on: http://www.sap.com/solutions/mobilebusiness/index.epx
[15] Segan, S. (2005). Wireless Without Borders: Networks for Those on the Go. PC Magazine. 24(5), p90
[16] Siemens, accessed in March 2005 on: www.siemens.com
[17] Solheim, S. (2005). 3G Gets Real. eWeek. 22(3), p28.
[18] Vos, I. and de Klein, P. (2002) The Essential Guide to Mobile Business. Prentice-Hall PTR, Upper Saddle River, NJ.

ON the Functional Characterization of "Intelligent" Materials

Yehia Haddad

Abstract— ***The paper reviews possible forms of intelligence that may be incorporated in various classes of engineering materials and structures. Basic mechanisms of intelligence are described, and implementation of these, as well as pertinent algorithms and techniques are illustrated. Possible analytical approaches to the functional characterization of intelligent material systems are discussed. For this type of material, it is logical to conclude that the micromechanical approach would generally overweigh the more conventional models of macromechanics. A knowledge-based experimental system, based on a pattern-recognition methodology, is ultimately introduced as a feasible technique to determine quantitatively the mechanical response states of intelligent material systems using non-destructive stress-wave propagation testing.***

Index Terms — Characterization, intelligence material, structure, mechanical response, micromechanics, microstructure, pattern-recognition. quasi-static, dynamic.

1. INTRODUCTION

Engineering materials are used either for their inherent structural strength or for their functional properties. Often a feed back control loop is designed so that the mechanical response of the material is monitored and the environment that is causing such a response can be controlled. The evolution of a new kind of material termed "*Intelligent*", "*Smart*", or "*Adaptive*", e.g. [1-6], witnesses a significant development in materials science. This concept aims at creating an artificially designed material, a so-called '*tailored*' material, having several functions in itself as a sensor, a processor, an actuator and feedback functions in combination with the inherent response characteristics of the parent material substance under consideration. The typical characteristics of an intelligent material may be stated to reflect that *'its properties, structural composition, function (and/or systematized functions) can adapt to the changes required in the environment and/or operating conditions'.* Such properties can be realized if the material has some built-in intelligence such as self-diagnosis, prediction / notification, self-repair and self-learning (i.e., with an ability to recognize, discriminate and backup). The concept of an '*intelligent/smart/adaptive*' material is related to a wide area of research, i.e., interdisciplinary fields such as medicine, biomaterials, polymers, metals, semiconductors, ceramics, mechanical and electronic engineering.

An "*Intelligent*" or "*Smart*" material may be defined as "*that material which senses any environmental change and responds to it in an optimal manner"*, e.g., Rogers *et al.* [6]. From this definition and the analogy of the *bionic* system of humans and animals, it can be seen that the following mechanisms may be essential for any material to be made intelligent (see Figure 1):

- A sensing device to perceive the external stimuli (e.g., skin which senses thermal gradients, an eye that senses optical signals, etc.), termed "*sensor*" function.
- A communication network by which the sensed signal would be transmitted to a decision-making mechanism (e.g., the nervous system in humans and animals), referred to as "*memory*" function.
- A decision-making device which has the capability of reasoning (e.g., the brain), designated as "*processor*" function.
- An actuating device, which could be inherent in the parent material or externally coupled with it (e.g., stiffening of muscles in humans and animals to resist any strain due to external loading), referred to as "*actuator*" function.

All of the above mechanisms need to be active in real time applications, so that the material could respond intelligently. Another important factor in the overall process is the time of response. This is the interval between the instant when the sensor senses the stimulus and that of the actuator response. An optimum time interval is crucial in the design of an intelligent material and obviously would depend on the particular application being considered.

Manuscript received October 25, 2005. This work was supported by NSERC – Natural Sciences and Engineering Research Council of Canada.

Yehia Haddad is Professor of Mechanical Engineering at the University of Ottawa, Ottawa, Canada.
(Email: yhaddad@uottawa.ca).

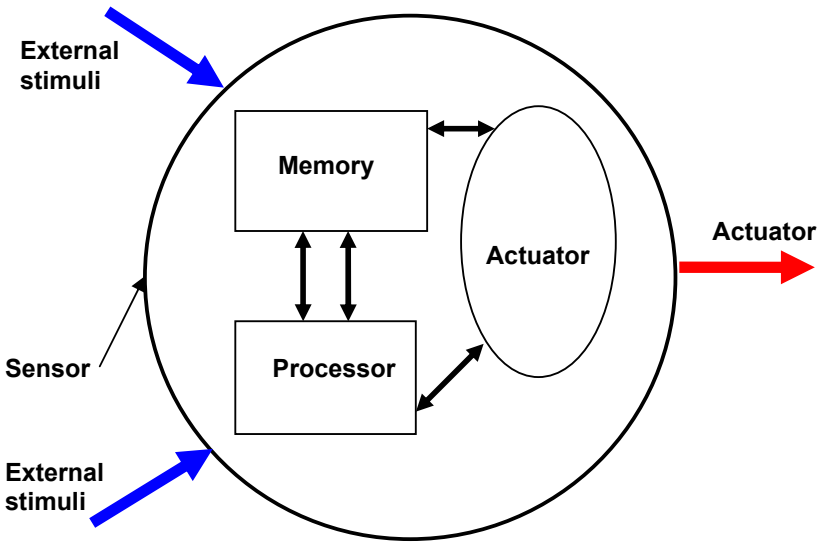


Figure 1. Concept of an "intelligent" material.

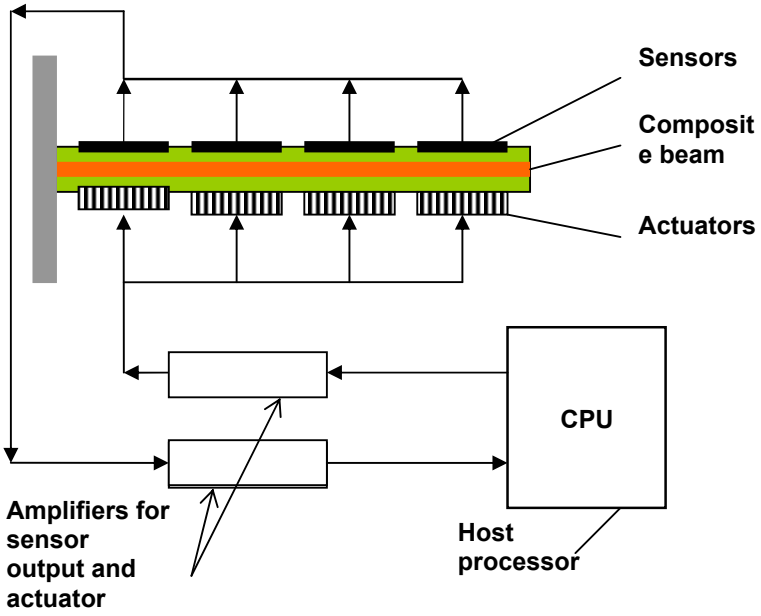


Figure 2. Incorporation of sensor, processor and actuator functions in an intelligent composite beam.

2. *INTERNAL FUNCTIONS OF AN INTELLIGENT MATERIAL*

2.1 Sensor Function

The concept of a sensor function in a smart material is defined as the ability of the material to sense the response, of self, to the externally imposed environmental factors such as mechanical loading, temperature, humidity and electrical inputs. An example of this function is that of an optical fibre sensor [7] embedded in a composite material. Such sensor diagnoses the mechanical disturbance imposed on the material by generating an output which can be further measured and analysed.

2.2 Memory and Processor Function

This mechanism stores the signals which are sensed and transmitted earlier by the sensor function. The characteristics of these signals are then compared with pre-stored acceptable values acquired during the '*training*' process of the processor. The training process may be carried out using an artificial intelligence technique, e.g., a pattern recognition method [8, 9]. Typically, this function is in the form of executable artificial intelligence software that could produce a logical output in the form of an electrical voltage that

could be amplified and used to activate an actuator mechanism; see, e.g. Takagi [10].

2.3 Actuator Function

This mechanism is coupled with the material or the structure member. It produces an output corresponding to the signal received from the processor function. This output is usually in the form of a restoring stress, strain or change in temperature, or stiffness (response) of the actuator mechanism. Such change in response would be designed to neutralize the effect of the change in the environment on the material, thus enabling the material to adapt continuously to its environment. A typical intelligent composite cantilever beam which comprises of sensor, processor and actuator functions is illustrated, as a possible example, in Figure 2 above.

3. SOME CLASSES OF INTELLIGENT DEVICES

Different forms of substances that could be incorporated into the material as sensors and actuators could be designed, for instance, as piezoelectric and piezoceramic devices. Optical fibres also are often used as sensors. Shape memory alloys, shape memory polymers and electrorheological fluids, among others, can be employed as actuators

TABLE 1. Piezoelectric sensors and actuators

Type	Piezo-effect	Input	Output	Applications
Piezo ceramic (PZT)†	Direct	Stress	Voltage	Sensors for mechanical loading
	Converse	Voltage	Strain	Actuators for deformation control
Piezo electric polymer (PVDF)††	Direct	Mechanical loading (static and dynamic)	Voltage	Sensors for static and dynamic loadings. Also, as passive vibration absorbers
	Converse	Voltage	Strain	Strain rate control

† Lead zirconate titanate piezoelectric ceramics
†† Polyvinylidene fluoride

.

3.1 Piezoelectric and Piezoceramic Devices

Piezoelectric and piezoceramic materials could be used as sensors and actuators in intelligent materials and structures. These devices can convert a mechanical signal to an electrical voltage, and vice versa.

A piezoelectric material is a crystal in which electricity or electric polarity is produced by pressure. Conversely, a piezoelectric material deforms when it is subjected to an electric field. The first characteristic expresses the so-called "*direct*" effect, while the second expresses the "*converse*" effect. Following the above characteristics of a piezoelectric crystal, if the pressure on the crystal is replaced by a stretch, the sign of the electric polarity would be reversed accordingly. This is determined by the crystal structural "*bias*" which establishes whether a given region on the surface is subjected to a positive or a negative mechanical effect. In the converse effect, the same unidirectional aspect determines the sign of deformation when the direction of an electric field input is reversed in the crystal. It is this reversal of sign of mechanical strain with that of the electric field that distinguishes piezoelectricity from electrostriction. Table 1 describes the utilization of direct and converse piezoelectric effects as applied to the sensor and actuator functions of such classes of intelligent materials [11-13].

3.1.1 Piezoelectric sensors and actuators

As mentioned in the foregoing, mechanical displacement and electrical voltage are the varying parameters of the intelligent material when using piezoelectric sensors and actuators. Mechanical disturbance is converted into electrical voltage by a piezoelectric sensor [14, 15]. On the other hand, a piezoelectric actuator is activated by an electrical input to produce specific mechanical effect (e.g., vibrations or strains) through properly designed control algorithms. Such mechanical effect would then be used to compensate or control undesired effects such as deflections, excessive vibrations caused by the external stimuli acting on the engineering material or structural member with which the intelligent material is incorporated.

3.1.2 Piezoelectric polymers as intelligent sensors and actuators

Polyvinylidene fluoride (PVDF), for instance, is a piezoelectric polymer that can be used for sensor/actuator functions. The piezoelectric polymer may be embedded inside a structural

member to actively control the vibrations and/or strains by dissipating the elastic energy imposed on the member. Such attenuation is achieved by converting a large fraction of the elastic energy into electric energy using the piezo-electric coupling effect and then dissipating the electrical energy using a simple resistive element, e.g. [15].

3.2 Optical Fibers as Sensors

Optical fibres can be used effectively as sensors in intelligent materials. Optical fibres may be classified, in general, into the following two types:

i) An extrinsic optical fibre sensor that operates only as a transmitting medium for light but it does not perform any of the sensing functions.

TABLE 2. Applications of optical fibres

Variable	Methodology	Examples of applications
Stress	Photoelastic effect	Fibre composites embedded with optical fibres can detect mechanical loading & vibrations
Strain	Change in optical power due to deformation	Strain could be sensed in structures embedded with optical fibres
Temperature	Thermal change in refractive index	Thermal state of fibre composites could be monitored during manufacturing by embedded optical fibres

ii) An intrinsic optical fibre sensor that utilizes some intrinsic property of the fibre to detect a phenomenon or to quantify a measurement.

Glass and silica fibres form a basis for a broad range of sensors. These fibres utilize the pertinent material properties to provide signals, indicative of external stimuli such as force, temperature and strain that are to be measured. The intrinsic properties of glass and silica qualify fibre optics as smart materials. Optical fibres are capable of performing as a sensor as well as a transmitter of the sensor's signal. In this, optical wave guides may be embedded in the material composite and be used to determine the levels of dynamic strain to which the structural member is subjected to. This is carried out by using the change in the optical power transmitted in the fibre due to the induced strain in the structural member and processing the resultant signal; see e.g. [16, 17]. A list of intrinsically measurable variables associated with the use of optical fibres is given in Table 2 above.

3.3 Shape Memory Alloys (SMA's)

Shape memory alloys have emerged as an appropriate choice for situations involving dynamic control of large structures, which would often require vibration suppression and strain control. Such vibrations and strains are induced by an adverse environment.

Shape Memory Alloys possess response behaviour in dependence of the state of loading/strain they are subjected to, or the thermal environment in which they are loaded. If such alloys are deformed at one temperature, they will completely recover their original shape when their thermal state is raised to a higher temperature. On the other hand, if these alloys are constrained during recovering, they can produce a mechanical effect (a recovery response) that is in relation with their temperature of transformation. Several alloy systems exhibit the phenomenon of shape memory. A number of such alloy systems and their characteristics are given in Table 3 below.

In the case of "*one way shape memory effect*", an SMA wire, for instance, deformed below the temperature of the Martensitic end of transformation temperature can regain its original shape when heated to a temperature above that of the Austenitic transformation temperature. But when cooled again to the temperature of its Martensite start of transformation, the wire recovers its original configuration and the material does not assume any longer the earlier deformed shape. In the case of '*two way shape memory effect*', however, a deformed SMA material below its Martensitic transformation temperature would regain its undeformed configuration when heated to a temperature above the temperature of Austenite end of transformation. However, the undeformed configuration spontaneously attains its deformed configuration when cooled below its Martensitic end of transformation temperature. The specimen can, however, recover its undeformed configuration when heated to temperatures

above its Austenitic end of transformation temperature. Thus, it is possible to produce two geometric configurations of the material,

TABLE 3. Alloy systems exhibiting shape memory effect

SME-alloy systems	Transportation temperature	Recovery force for 2% strain in Kg/mm^2
Nitinol[1]	373 K	17
Cu-Zn-Al[2]	350 K	9
CANTIM 75[3]	480 K	14

[1] 49.93% nickel and 50.03% titanium.
[2] 25.9% zinc, 4.04% aluminium and rest is copper
[3] 11.68% aluminium, 5.03% nickel, 2.00% Manganese, 0.96% titanium and rest is copper.

by subjecting it to thermal cycling. The latter is termed as the "*trainability of two way shape memory effect*".

Thermomechanical environment may subject materials to cyclic thermal loadings, leading to fatigue and other undesirable mechanical effects. If the shape memory material is made to alter its mechanical properties with respect to a mechanical loading, many of the induced strains, due to such loading, could be controlled. In this case, the thermal environment is sensed by an incorporated sensor and the SMA-material acts as an actuator by changing its mechanical response properties when heated (e.g., by passing an electric current through the SMA material); see, for instance [18-20].

In a multi-layered composite laminate with embedded SMA-fibres, excellent vibration suppression could be achieved when the laminate is subjected to dynamic loading. Varying the mode shapes of induced vibration could be also achieved by varying the stiffness of SMA-fibres. This is accomplished by utilizing the large force created on constraining the micromechanical phase transformation from a deformed state to an undeformed state. It is also possible to use SMA-fibres as simple thermomechanical actuators rather than integrating them into a fibre-matrix system. This is achieved by coupling the thermomechanical actuator to the structural member externally. By ensuring proper coupling between the actuator and the structural member, the effects of the SMA actuator could be transferred to the parent material. Thus, shape memory alloys can be used effectively as actuators in intelligent material systems, or structures, when coupled with proper sensor and control algorithms; e.g. [21].

3.4 Material Intelligence Using a Viscoelastic Microstructure

Polymeric materials are generally viscoelastic in response behaviour and have the capability of changing their dynamic properties (storage modulus E', loss modulus E" and loss tangent *tan* ϕ) with variations in environmental factors such as temperature, frequency and time; e.g. Haddad [22]. Thus, polymeric materials, in general, would have a smart/intelligent function capability. This is accomplished by a sensor/actuator mechanism

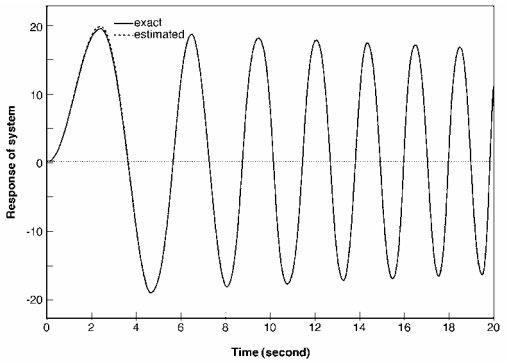


Figure 3. The exact and estimated responses from the first-order discrete time system for a particular first-order dynamic system with a first-order dynamic sinusoidal input.

that could be incorporated in a structural member so that external stimulus such as mechanical vibrations could be sensed. By incorporating a control mechanism, the dynamic moduli of the polymeric material could be made to change so that the material could adapt itself to the new environment. This could be achieved also by shifting the loss factor ($tan\ \phi$) towards the frequency spectrum that matches the imposed vibrational frequency, so that the absorption of the imposed vibrations would be maximized. This shifting could be carried out by varying the loss modulus (E") or the loss factor ($tan\ \phi$) of the polymeric material, which will be acting in this case as a damper, with respect to temperature or frequency.

For a linear dissipative system, it is well recognized that the behavioural functions characterizing both quasi-static and dynamic responses are interrelated, e.g. [22-24]. Quasi-static experiments, to determine such response functions, require, however, considerably long periods of time to be performed. To overcome such inconvenience, dynamic methods are recently attracting the attention of researchers. Gibson, *et al.* [25], for instance, presented a method by which experimental dynamic data are used to determine both quasi-static and dynamic response behaviour of the material. In their method, the complex moduli were obtained first from vibration measurements by employing Fast Fourier Transform technique. Then, the quasi-static time-dependent properties were calculated from the experimentally determined dynamic properties by employing a numerical integration algorithm. In this context, Haddad and Yu [24] considered that the viscoelastic material as a dynamic system, whereby a relation was established between the quasi-static response functions and corresponding frequency functions of a specifically proposed dynamic system. In the frequency domain, an analytical model was assumed for the frequency response function of the system, then, a discrete-time system analysis was developed to estimate the order and parameters of the proposed model. The proposed model was shown to be efficient and powerful; see, e.g., Figure 3 above.

3.5 Shape Memory Polymers

Shape memory polymers are unique polymeric materials which can recover their original shape before deformation at lower temperature (below the material's glass-transition temperature T_g), upon heating them to a temperature above their glass transition temperature T_g ; see e.g., Yoshiki and Shun-Ichi [27]. This is an apparent advantage over ordinary polymers. An ordinary polymer when stressed, may not recover completely to its original undeformed configuration if the stress is released (or its temperature is raised), thus, resulting in permanent deformation. In a shape memory polymer, however, the recovery loop is completed upon heating, thus, a shape memory polymer is able to revert back to its original shape without undergoing any permanent deformation.

3.6 Electro-rheological Fluids

The viscosity of certain fluids is influenced by the applied electric field. This phenomenon, termed "*Electroviscous Effect*", was first reported around the turn of the century; e.g. Duff [26]. Researchers have found, for instance, an increase in the viscosity of conducting polar liquids of up to 100%, upon application of an electric field of the order of 1-10 kV/cm. For the electroviscous effect to occur, both polar molecules and conducting impurity ions would be needed to be present. Large increases in viscosity, due to an applied electric field, for suspensions of finely divided solids in low viscosity oils was found as early as 1949. This effect, termed "*Winslow Effect*", is attributed to field induced fibre formation of the particles between the electrodes, thereby requiring additional shear stress for flow; see e.g. Conrad and Sprecher [27]. This phenomenon has recently been termed as "*Electrorheology*" and has been applied in the development of actuator mechanisms in intelligent materials. When used with suitable sensors and control algorithms, electrorheological fluids can be made to change their properties by subjecting them to an electric field.

With reference to Fig. 4, below, an engineering structural member which contains electro-rheological fluid, when not activated, has a very low composite stiffness. This state represents the undisturbed configuration. When an environmental input (e.g. mechanical loading or a difference in thermal gradient) causes, for instance, deflection in the structural member, it would be desirable to increase the stiffness to control or limit the occurring deflection. This is achieved by sensing the external mechanical loading through incorporated sensors. The sensed signal is then processed to a microprocessor, which activates an auxiliary electric input to produce a desirable voltage. This voltage, when applied to the electrorheological fluid contained in the mechanical structural member, increases the viscosity of the fluid, thus, practically converting it into a solid. As a result, the overall stiffness of the structural member is increased. The above said process could be made to take place in a matter of $1/1000^{th}$ of a second.

4. FUNCTIONAL CHARACTERIZATION OF "INTELLIGENT" MATERIALS

4.1 Macromechanics vs. Micromechanics

In assessing the mechanical response behaviour of an intelligent material system, macromechanics and micromechanics models are used.

In the *continuum mechanics* approach, the actual microstructure of the intelligent material system is disregarded and the medium is

pictured as a "*continuum*" without gaps or empty spaces. Hence, the configuration of the assumed continuous medium is described by a continuous mathematical model whose geometrical points are identified with material particles of the actual physical medium. Further, when such a "continuum" changes its configuration under some boundary conditions, such change is assumed to be continuous, i.e., neighbourhoods evolve into neighbourhoods. Thus, the

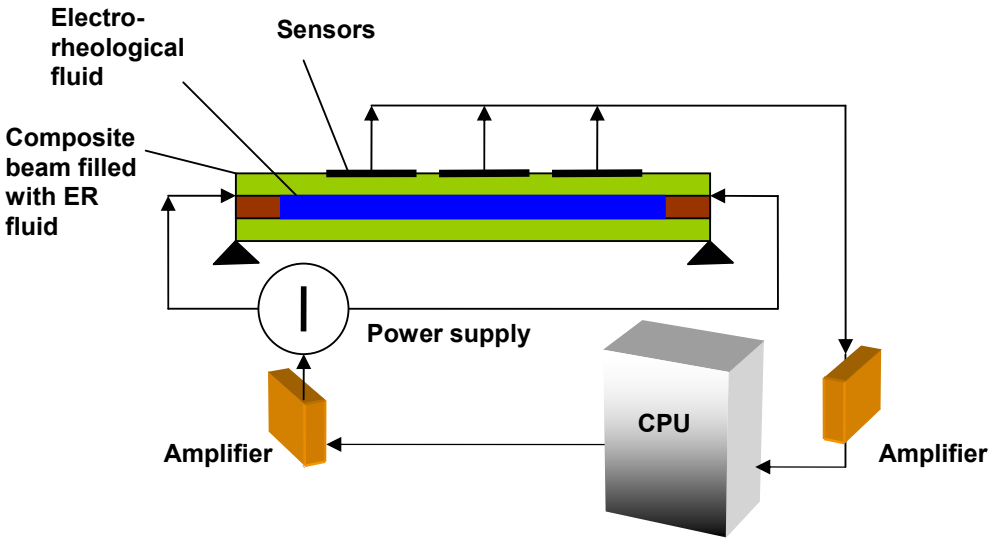


Figure 4. Electrorheological fluid as an actuator in a smart beam.

mathematical functions entering the analysis of the pertinent deformation process, are assumed to be continuous functions with continuous derivatives. Any creation of new boundary surfaces, such as those developed by interfaces between the sensor, actuator and the parent material may, then, be seen as extraordinary events that might require alternative formulations outside the realm of continuum mechanics. An example of intelligent material systems that often treated with continuum macromechanics is a cantilever beam whose response behaviour is controlled by two layers of piezoelectric material layers, e.g. [17].

In the *micromechanical* approach, however, the macroscopic medium is considered to consist of constituent structural elements. Such elements are often seen to be interacting with each other, and, hence their individual responses are considered to be mutually inter-dependent. The behaviour of an ensemble of such elements may be studied using, for instance, deterministic or statistical (stochastic) micromechanics; see, Haddad [8].

Laminated fibre composite materials, for instance, with their high specific modulus, high specific strength, and the ability to tailor them for a specific application, offer definite advantages as potential "*adaptive*", i.e. "*intelligent*" material systems. In this context, modelling of such response, from a micromechanical point of view, could provide in-depth understanding of the microstructural mechanisms that might be available for possibly controlling the mechanical performance of engineering structures made of such materials. Thus, in this context, the effects of selected microstructural parameters on the damping and stiffness of a class of "*adaptive*" viscoelastic fibre-composite systems may be examined. Subsequently, simultaneous optimisation of the adaptive material system characteristics, e.g., damping and stiffness, could be carried out; Haddad and Feng [28].

4.2 Pattern-Recognition and Classification Methodology

A knowledge-based experimental system; e.g. Haddad [8, 29], may be used to determine quantitatively the mechanical response state of an intelligent engineering material system, as based on quantitative non-destructive testing combined with "*pattern recognition" and "classification*" methodologies. In the pertaining experimental procedure, stress waves are "*simulated*" in the microstructure of the material system to resemble acoustic emission waves After propagating through the microstructure, the waveforms are captured, identified, and then classified as belonging to various classes, where each class represents one of different response states of the tested material-property. This experimental approach has been proven to be powerful in determining quantitatively numerous material response states in both homogeneous and heterogeneous classes of engineering materials that were subjected *priori* to unknown static, quasi-static and dynamic types of loading.

5. SUMMARY

Intelligent materials have the ability to control and hence improve the performance of engineering structures. Although the concepts of the techniques described in this article were discovered decades ago, only recently that such techniques have emerged as potential applications in intelligent materials methodology. The formulations for piezoelectric effect indicate the nature of direct and converse effects and their possible use in sensor and actuator technologies. Discussions relating to shape memory alloys, viscoelastic materials, shape memory polymers and electrorheological fluids, illustrate the possible usage of these materials as actuators in smart material systems. The increase in stiffness of shape memory alloys and the change in the dynamic moduli of shape memory polymers, and viscoelastic materials, in general, with temperature and/or frequency offer distinct advantages in controlling the static and dynamic states of the engineering systems and structures. In addition, the development of different feed back mechanisms, as based on control algorithms, and the increase in sophistication of microprocessor technologies and pattern recognition methodology will definitely play an important role in the advancement of processor function in this new field of science. It is emphasized that mechanical behaviour of intelligent material systems should be analysed using micromechanics. In this, an experimental quantitative approach based on pattern recognition and classification methodologies is proven to be useful.

ACKNOWLEDGMENT

The author wishes to sincerely acknowledge the research grants provided to him by Natural Sciences and Engineering Research Council of Canada (NSERC). The scientific and technical assistance of AES-Advanced Engineering Solutions (Ottawa, Canada) is also gratefully appreciated.

REFERENCES

[1] Rogers, C.A., “Workshop summary”, Virginia Polytech. Inst. & State Univ., ARO Smart Materials, Structures & Mathematical Issues, 1988, pp.1-9.

[2] Ahmed, I., “Smart structures and materials”, Virginia Polytech. Inst. & State Univ., ARO Smart Materials, Structures & Mathematical Issues, 1988, pp. 13-16.

[3] Takagi, T., “Recent research on intelligent materials”, J. Intelligent Material Systems and Structures 7, 1996, pp. 346-52.

[4] Muto, N., Yanagida, H., Nakatsuji, T. *et al.* “ Design of intelligent materials with self-diagnosing function for preventing fatal fracture”, Smart Mater. Struct. 1, 1992, 324-29.

[5] Holnicki-Szule, J. and Rodellar, J., editors, “Smart Structures: Requirements and Potential Applications in Mechanical and Civil Engineering”, *Kluwer* (Springer), NATO Advanced Research Workshop, Warsaw, Poland, 1998.

[6] Rogers, C.A., Barker, D.K. and Jaeger, C.A., “Introduction to smart materials and structures”, Virginia Polytech. Inst. & State Univ., ARO Smart Materials, Structures & Mathematical Issues, 1988, pp.17-28.

[7] Main, R.P., “Fibre optic sensors – Future light”, Sensor Review 5(3), 1985, pp.133-9.

[8] Haddad, Y.M., *Mechanical Behaviour of Engineering Materials,* Volumes I & II, *Kluwer* (Springer), 2000.

[9] Haddad, Y.M., “An intelligent non-destructive inspection system for the quantitative determination of material response states: A multidisciplinary approach”, IPSI’2005-France, 2005.

[10] Takagi, T., “A concept of intelligent material”, U.S.-Japan Workshop on Smart/Intelligent Materials and Systems, Honolulu, Hawaii, 1990, pp. 3-10.

[11] Sung-Kyu Ha, Charles K. and Chang, F-K, “Analysis of laminated composites containing distributed piezoelectric ceramics”, J. Intell. Mater. Syst. & Struct. 2, 1991, pp.59-71.

[12] Crawley, E.F. and Luis, J.D., “Use of piezoceramics as distributed actuators in large space structures”, AIAA, 1985, pp. 126-32.

[13] Crawley, E.F. and Luis, J.D., “Use of piezoelectric actuators as elements of intelligent structures”, AIAA 25(10), 1987, pp. 1373-85.

[14] Bailey, T. and Hubbard, J.E., “Distributed piezoelectric polymer active vibration control of a cantilever beam”, J. Guidence 8(5), 1985, pp. 605-11.

[15] Ikeda, T., “Fundamentals of piezoelectricity” Oxford University Press, 1990.

[16] Claus, R.O., McKeeman, J.C., Mary, R.G. and Bennet, K.D., “Optical fibre sensors and signal processing for smart materials and structures”, Virginia Polytech. Inst. & State Univ., ARO Smart Materials, Structures & Mathematical Issues, 1988, pp.29-38.

[17] Haddad, Y.M. and Frédéric, T., “Modélisation et matériaux intelligents”, Internal report in cooperation with Université Polytech’ Grenoble, Grenoble, France, *Ubiversity of Ottawa*, 2004.

[18] Delaey, L., Krishnan, R.V., Tas, H. and Warlimont, H., “Review – Thermoelasticity, pseudoelasticity and the memory effects associated with martensitic transformation”, J. Mats. Sci. 9, 1521-35.

[19] Cross, W.B., Kariotis, A.H. and Stimler, F. J., “Nitinol characterization study”, NASA Contractor Report, NASA CR-1433, 1969.

[20] Jackson, C.M., Wagner, H.J. and Wasilewski, R.J., “55-Nitinol – The alloy with a memory: Its physical metallurgy, properties and applications”, NASA report, NASA-SP5110, NASA, Washington, D. C., 1972.

[21] Yoshiki, S. and Shun-Ichi, H., "Development of Polymeric Shape Memory Material", Mitsubishi Tech. Bulletin, No. 184, *Mitsubishi Heavy Ind. Ltd*., New York, 1988.
[22] Haddad, Y. M., *Viscoelsticity of Engineering Materials*, *Kluwer* (Springer), 1995.
[23] Haddad, Y. M. and Feng, J., "On the optimization of a class of adaptive composite systems: A multidisciplinary approach", IPSI' 2005 – Spain, 2005.
[24] Haddad, Y.M. and Yu, P., "A dynamic system approach to the characterization of a class of rheological materials at base-temperatures, Proceedings, 32nd Annual Conference on Thermal Analysis and Applications, NATAS, Williamsburg, Virginia, 2004.
[25] Gibson, R.F., Hwang, S.J. and Sheppard, C.H., "Characterization of creep in polymer composites by the use of frequency-time transformations", J. Composite Materials 24, 1990, pp.441-53.
[26] Duff, A.W., "The viscoelasticity of polarized dielectrics", Phy. Rev. 4, 1896, pp. 23-38.
[27] Conrad, H. and Spreher, A.F. "Characteristics of ER fluids", Advanced Mats. Conf., Colorado, Feb. 25-27, 1987, pp. 63-76.
[28] Haddad, Y.M. and Feng, J., "On the optimization of the mechanical behaviour of a class of composite systems under both quasi-static and dynamic loading", AMPT' 1999 and IMC' 16, Dublin, August 3-6, 1999, pp. 1351-61.
[29] Haddad, Y.M., "Towards determining non-destructively the ultimate strength of the material in a quantitative manner", AES Technical Reviews (in Press).

Dr. Yehia Haddad is a senior professor of mechanical engineering at the University of Ottawa. He was granted his Ph.D. in Mechanical Engineering from McGill University in 1975. Since then, he has been actively doing scientific research in Canada, Japan, France and Finland. He published extensively in well-recognized international scientific journals, and presented a large number of scientific papers in national/ international conferences and symposia. Dr. Haddad authored, in addition, six engineering monographs on applied mechanics, mechanics of solids and advanced engineering materials. He also contributed to a number of engineering handbooks and encyclopedia. Further relevant information is available on the website: www.genie.uottawa.ca/~haddad

Comparative Studies of Shortest Path Algorithms and Computation of Optimum Diameter in Multi Connected Distributed Loop Networks

Vasappanavara, Ramesh; Prasad, EV; Seetharamanath, MN

Abstract. Multi connected Distributed Loop (MCDL) networks with multiple hops offer smaller diameters, path lengths and better fault–tolerance. These networks have extensive uses in LAN, parallel processing, and multi processing environments. Results exist for finding shortest path between a single pair in O(δ) time where δ is diameter of distributed and in O(h/g + log h), and g is GCD(N, h), where N is number of nodes h is hop size, without prior knowledge of diameter. This paper critically analyses both these approaches and benchmarks these algorithms for suitability for adopting to MCDL Networks with large number of nodes. Results also exist for computing diameter in time O(log r) , where 0≤ r < h and h is hop length , for certain cases with restrictions imposed . But, in general, there is no closed form for diameter. In this paper we present a novel method to find out optimum diameter given N by employing curve fit using power series equations. We provide a closed form of expression for Opt Diameter = √N/2 for 1≤ N ≤112 ,rounded to nearest integer. The error is observed to be 0.035%.

Index Terms— *Algorithm , Diameter , Distributed Systems, Optimum Diameter , Shortest path , Un directed double loop networks.*

1. INTRODUCTION

A distributed loop network G (n; h_1, h_2) [8] is an undirected graph with vertex set Zn = {0, 1 …. n-1} and edge at E = E_1 U E_2 where n, h_1, h_2 are positive integers.

$$E_1 = \{(u, u \pm h_1) \mid u \in Z_n\}$$
$$E_2 = \{(u, u \pm h_2) \mid u \in Z_n\}$$

[2] provides survey of Distributed Loop networks. A distributed loop network has two types of links h_1 and h_2 where every link h_1 connects u^{th} node to node (u ± h_1) mod n and link h_2 connects node u to node (u ± h_2) mod n. If either h_1 or h_2 is equal to 1, then we have a ring network with some additional links added homogenously to it. These networks are called Multi Connected Distributed Loop Networks (MCDL) and are denoted as G (n; 1, h). For example figure 1 is G (20; 1, 5). While designing the network, it is important to consider the performance requirements of the network in terms of network parameters such as; communication delays, shortest path lengths and average path lengths These parameters depend on various network characteristics such as; hop size, diameter. Hence, finding shortest path is of paramount importance in MCDL networks. A number of algorithms have been proposed in order to find the shortest paths for routing the packets in MCDL networks. In [7,12,13] authors have presented details regarding comparative studies of shortest path algorithms , fault-tolerance and delay performance studies for multi connected distributed loop networks. Several researchers also considered directed double loop networks [3, 8, 9], which have directions assigned to the four links of each node. In a directed double loop network, for each pair of links of same hop (i.e., a pair of links consisting of ±1-links only or consisting of ±h-links only), one of the two links is an inward arc and the other link is

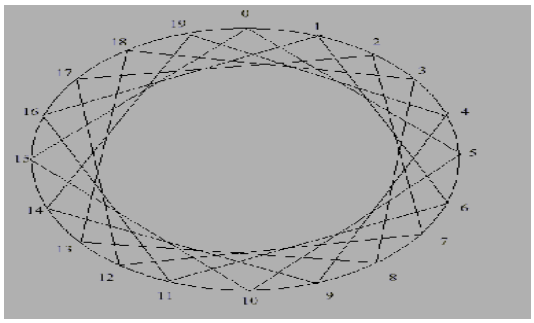


Fig 1. Distributed loop with n = 20 and h = 5

an outward arc, uniformly throughout the network. Raghavendra et al [8, 9 ,10] proposed a forward loop and backward hop (FLBH) network

Manuscript received May 10 , 2005.

Ramesh Vasappanavara* is with the Computer Science & Engineering Department, Gayatri Vidya Parishad Collge of Engineering,Visakhapatnam,India.(email:ramesh_vasappanavara@yahoo.com).

EV Prasad is with the Computer Science Department , JNTU College of Engineering Kakinada. Email: deevprasad@yahoo.co.in

MN Seetharamanath is with the Computer Science & Engineering Department,AndhraUniversityCollegeofengineering,Visakhapatnam ,India.(email:mns_nath@yahoo.com).

using +1- and −h- links and further discussed on reliable loop topologies and their performance studies. In [1], Arden and Lee proposed a chordal ring in which hop connections are formed by connecting node i to (i+h) mod N or to (i-h) mod N, depending on whether i is even or odd. In [5] Chen Bao-Xing et.al presented a method to compute the diameter of undirected double loop networks with certain restrictions, i.e when $q \le r$ and $a > (b+1)q+1$, where q, r, a and b satisfy $N = qh+r$, $0 \le r < h$, and $h-r = br + a$, $0 \le a < r$, and the algorithm runs in time O(log r), where $0 \le r < h$ and h is hop length. In our study we concentrate on algorithms , that do not need prior knowledge of diameter , i.e algorithm at [4] which has O (h/g + log n) time where g = GCD (n, h) and algorithm at[11] which has a time complexity of 'd' time steps where 'd' is the diameter of the network.

The MCDL network has a total of four link combinations denoted as (+1, +h), (-1, +h), (+1, -h), (-1, -h). The sub graphs formed by each of these link combinations can be represented as G_1 (n; +1, +h), G_2 (n; -1, +h), G_3 (n; +1, -h), G_4 (n; -1, -h). The shortest path from node l to m in G (n; +1, +h) can be represented as sp_{++} (l, m). Similarly shortest paths in the remaining link combinations can be represented. The sp (l, m) in MCDL is equal to sp (l, m) = Min[sp_{++}(l, m) + sp_{+-}(l, m) + sp_{-+}(l, m) + sp_{--}(l, m)]. Due to property of isomorphism among G_1, G_2, G_3 and G_4 [4], G_1 and G_4 are isomorphic and similarly G_2, G_3. Hence one does not need all four combinations but only two combinations are sufficient due to isomorphism. So only G_1, G_2 are sufficient for shortest paths.. The diameter of a network is defined to be the length of the maximum shortest path in the network. So we can represent the diameters of the MCDL is dia (n, h) = Max[sp(l, m)] for all l, m pairs in Z_n. The diameter of the network is dependent on the hop size and as the hop size varies the diameter value changes. The minimum of all diameters obtained with different hop sizes for a given 'n' is called optimal diameter, od(n) = $Min_{\forall h}$ [dia(n,h)]. It may be noted that the 'od(n)' may be obtained at one or several values of h. There is no closed form equation for the shortest path, diameter and optimal diameter. Hence algorithms [4,6,11] exist. [6] requires knowledge of diameter to compute shortest path in O(δ) time , where δ is diameter. Algorithms at {4,11] do not require knowledge of diameter .They are compared for checking the suitability for application to large undirected double loop networks in the next section.

2. EVALUATION OF ALGORITHMS

In any configuration, the paths through which the packets are routed from the source to the destination are of utmost importance [4,5] and if the packets were not routed through the shortest paths, the performance of the system would be less than expected [14]. Therefore, in most networks, the shortest path information is calculated, stored in some form or the other and used later for routing. We consider algorithms [4] and [11] in terms of their time complexities and suitability for un directed double loop networks.

2.1 GCD Based Algorithm

The time complexity of this algorithm, as mentioned in [4], is O(h/g + log (h)) where g = GCD (n, h) and h is hop size. This algorithm's time complexity is independent of the diameter. The algorithm computes shortest path only for (+h, +1) and (+h, -1) combination of links. Isomorphic property of network can be employed for finding the paths along the other two link combinations. Finally the minimum of these four is the shortest path. A detailed algorithm is found in [4]. Hence using isomorphic property, we need to calculate a total of 2*(n-1) paths as stated above. Therefore, total time to fill the array, is of the order O (2*(n-1)*(h/g + log (h)). To calculate the shortest path between one source destination pair, we need to access only one single array element i.e. c. If 'p' paths are calculated it is p*c. therefore, total time is O (2*(n-1)*(h/g + log (h) + p*c).

2.2 COMPUTATION OF OPTIMAL DIAMETER (OD) AND HOP-SIZES(H) CORRESPONDING TO 'OD' USING ALGORITHM [4] GIVEN NUMBER OF PROCESSORS (N) AS input**.**

Algorithm:

1. Read 'n' ; // n is number of nodes in MCDL.
2. h ← 2 // h is hopsize.
3. s ← 0 // s is source node.
4. for d = 1 to n-1 // d is destination node.
   do
   a) sp_{++} (0,d)
   b) sp_{-+}(0,d)
   c) sp_{--}(0,d) ← sp_{++} (0, n-d) // G_1, G_4 are isomorphic
   d) sp_{+-}(0, d) ← sp_{-+}(0, n-d) // G_2, G_3 are isomorphic
   e) sp(0,d) = Min {a, b, c, d}
   end for
5. dia_h = max (sp(0,d)) ; // After all the destinations have been considered, compare all the shortest paths obtained and assign the maximum to dia_h
6. h ← h + 1; If (h < n-1) go to step 3.
7. od = Min (dia_h) ; // Compare all values of dia_h and assign the minimum value of diameter to 'od'. This is called 'h_{od}'.

Table : 1 for nodes 5 to 50 values of od & h_{od}

Nodes	od	h_{od}
5	1	2,3
6	2	2,3,4
7	2	2,3,4,5
8	2	2,3,4,5,6
9	2	2,3,4,5,6,7
10	2	4,6
11	2	3,4,7,8
12	3	2,3,4,5,6,7,8,9,10
13	2	5,8
14	3	3,4,5,6,8,9,10,11
15	3	3,4,5,6,9,10,11,12
16	3	4,6,10,12
17	3	3,4,5,6,7,10,11,12,13,14
18	3	4,5,7,11,13,14
19	3	4,5,7,8,11,12,14,15
20	3	8,12
21	3	6,8,13,15
22	3	6,16
23	3	5,9,14,18
24	4	4,5,6,7,9,10,14,15,17,18,19,20
25	3	7,18
26	4	4,6,7,10,11,15,16,19,20,22
27	4	4,5,7,8,10,11,16,17,19,20,21,22,23
28	4	5,6,8,11,12,16,17,20,22,23
29	4	5,6,8,11,12,17,18,21,23,24
30	4	8,12,18,22
31	4	7,9,12,13,18,19,22,24
32	4	7,9,12,20,23,25
33	4	5,6,7,9,13,14,19,20,24,26,27,28
34	4	6,28
35	4	10,25
36	4	8,10,26,28
37	4	8,14,23,29
38	4	16,22
39	4	7,11,28,32
40	5	6,7,9,11,12,15,16,17,23,24,25,2829,32,33,34
41	4	9,32
42	5	9,12,16,18,24,26,30,33
43	5	5,8,9,12,16,17,18,19,24,25,26,27,31,34,35,38
44	5	8,10,12,32,34,36
45	5	6,7,8,10,13,33,36,38,39,40
46	5	6,7,8,10,13,33,36,38,39,40
47	5	7,10,13,14,18,20,27,29,33,34,37,40
48	5	18,20,28,30
	5	9,11,14,35,38,40
50	5	9,11,14,19,21,22,28,29,31,36,39,41

2.3 Computation of Optimal Diameter (od) and hop -sizes corresponding to'od' using Algorithm [11] given number of processors (n) as input.

Algorithm :

1)i ← 0

2)s_i = { 0 }

3)I ← i + 1

4)s_i = set formed by adding 1, h, -h and -1 mod n to each element x in s_{i-1} such that if y € s_{i-1}, y ∉ s_i

5)Mark distances to all elements of s_i by adding 1.

6)If s_i does not contain n elements, which is number of nodes then Goto 3.

7)End

The algorithm was used for 'n' between 5 and 1000. to find the 'od' and the results are tabulated in Table 1. From the table it can be seen that as the size of the network increases (i.e. 'n' increases), the value of 'od' increases as expected. This is because the processors are farther apart and the paths required become longer. It can also be seen that the values of 'hod' the hop sizes where OD occurs also increase with 'n'. This indicates that for larger networks, larger hop sizes should be preferred and these larger hops have greater chance of providing shorter paths. We have fitted a linear graph into these values of 'od' versus n as given in Figure 2. The equation od = 0.017924*n + 6.1869 fits it and is of linear relation line. The values of 'od' for n = 5 to 1000 have been calculated using the above equation. Similarly, we have also fitted a non-linear power series graph, Fig 3, into these values of 'od' as od = 0.6943 * n0.5029 . The values of 'od' for n = 5 to 42 have been calculated using the above equation, rounding the value of 'od' to the nearest integer. We have drawn errors in linear and non linear fits in Fig:4 and observed that non-linear approximation gives a better results. Hence this non-linear approximation fit has been extended to the range of 1 to 1000.The respective equations for the ranges and the corresponding errors have been tabulated as below in Table 2. We have also observed that row1 of table2 can be approximated to governing equation Opt dia = Round(√ N)/2 for 1≤ N ≤112 ,rounded to nearest integer. The error is observed to be 0.035%

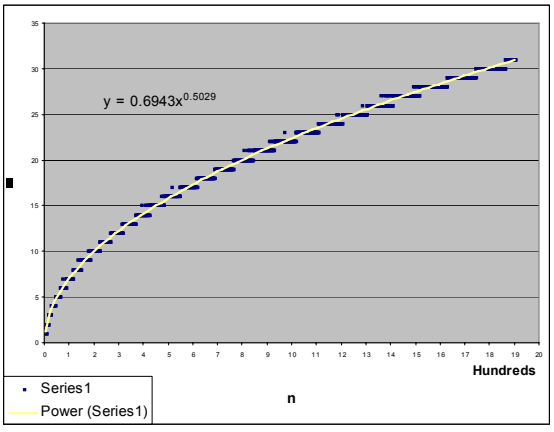


Fig 2. Linear approximation

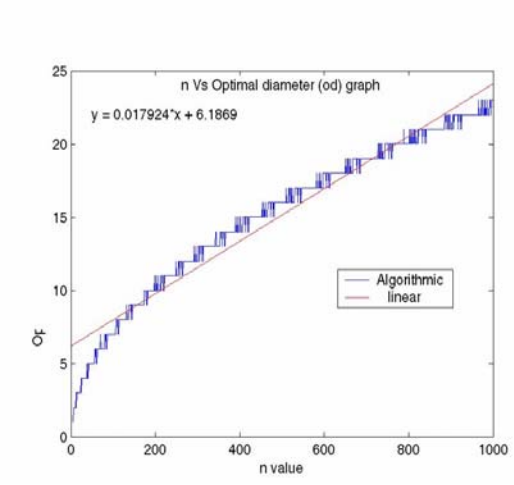


Fig:3 Non-Linear approximation

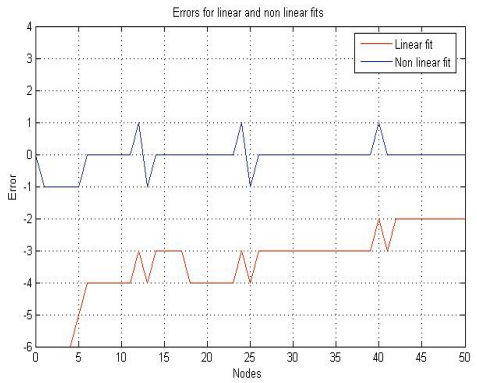


Fig 4 : comparison of linear and non linear fits

Slno	Nodes Range	Equation (power series) Opt.dia	avg% error
1	1-112	0.672*X^0.5109	0.035
2	113-172	0.5928*X^0.5122	0.03
3	173-262	0.8179*X^0.4561	0.05
4	263-300	1.9566*X^0.3148	0.05
5	301-418	1.9393*X^0.3148	0.008
6	419-454	0.067*X^0.8788	0
7	455-480	14*X^7E-11	0.08
8	481-516	0.06815*X^0.8788	0
9	517-610	16.203*X^-0.0019	0.01
10	611-682	0.523*X^0.5381	0.03
11	683-758	23.185*X^-0.0384	0
12	759-838	24.266*X^-0.0365	0.01
13	839-972	18.271*X^0.0134	0.01
14	973-1000	0.2359*X^0.6557	0

Table 2: Average % error for computing Optimum diameter given number of node (n) using algorithm and governing equation given by power series.

2.4 Algorithm [11]

This algorithm [11] does not give the number of 'h' and '1' links used; rather it provides us shortest path length. This information about the number of h and 1 links is necessary for the implementation of various fault-tolerant algorithms [4,5]. This algorithm fills an array of size 'n' with an integer which is the shortest distance from node 0.The time complexity of this algorithm is O (d) where d is the diameter of the network, however for the computation of the shortest path the knowledge of the diameter is not required. Once the array is filled, the k^{th} element gives the shortest path from node 0 to node the k. This can be extended to find the shortest path between nodes u and v basing on the observation that shortest path from u to v is equal to shortest path from 0 to (v-u) mod n. . Once the array is filled, the number of array elements that need to be accessed is variable and can be a maximum of d elements, where d is the diameter. Therefore, the total time required to calculate shortest path is d*c, where c is the time taken to access one single array element. The total time to fill the array and calculate one shortest path is O (d + d*c). If 'p' paths are to be calculated, time is d + d*p*c. As diameter is dependent on number of processors and hop size, the time complexity increases with increase in diameter and hence the number of processors.

3.0 Time Complexities

For [11], time required to fill up the array is O (d). For [4], the time to fill the array is the time required to calculate the shortest paths O (h/g + log h). These algorithms were implemented on IBM iv Pentium Machines and the time actually consumed by these algorithms computed by varying number nodes up to 6000. . The results are at Fig 5. It is observed that algorithm [4] performs better and displayed a flat response , indicating the independence of diameter for computing the shortest path for large n. The algorithm at [11] time consumed raises exponentially confirming O(δ) nature of the algorithm. Its performance is better only for networks with less number of nodes.

Charts 1 and 2 show the times we calculated for paths p = 50 and p = 200 respectively where p is number of shortest paths computed.. The network considered is 200 nodes and hop sizes were h= 4, 5, 6, 8, 10. The diameters obtained were, 16 and 14 respectively. 'D' indicates algorithm at [11] and 'h/g+log h' implies algorithm at [4]. Charts 3 and 4 show the time performance of the two algorithms for a network having 200 processors, with hop sizes 4 and 10 respectively. Here we have considered number of path lengths as parameter .We can see that time taken in [11] increases as the number of nodes increases, where as [4] with isomorphic property, gives a flat response, independent of the diameter and number of nodes.

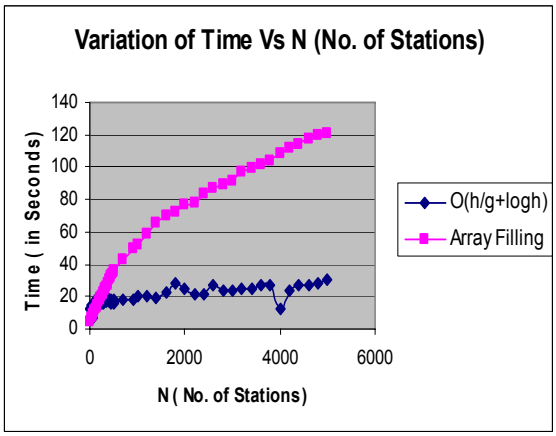


Fig 5: Time units consumed by algorithms [11 , 4]

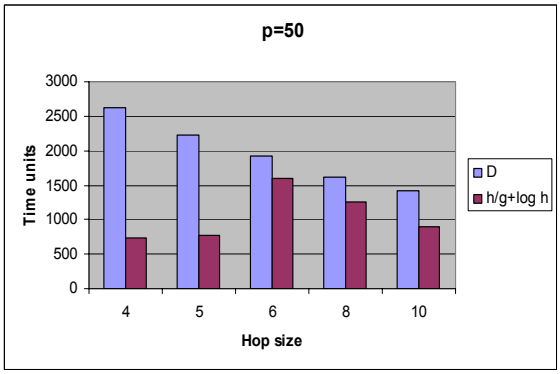


Chart 1: Time units (delay) vs. hop size, n=50

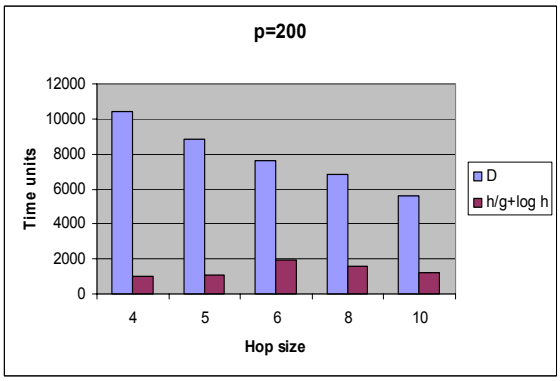


Chart 2: Time units (delay) vs. hop size, n=20

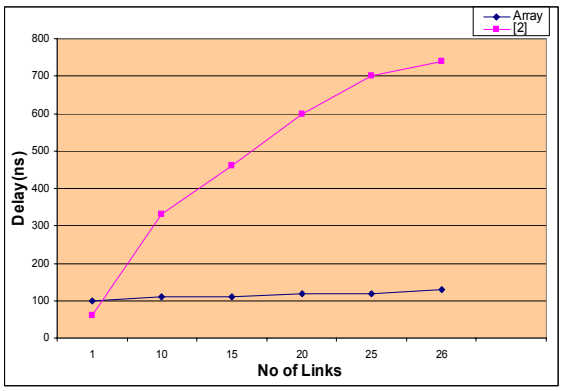


Chart 3:Delay vs. no of links traversed n=200, h=4

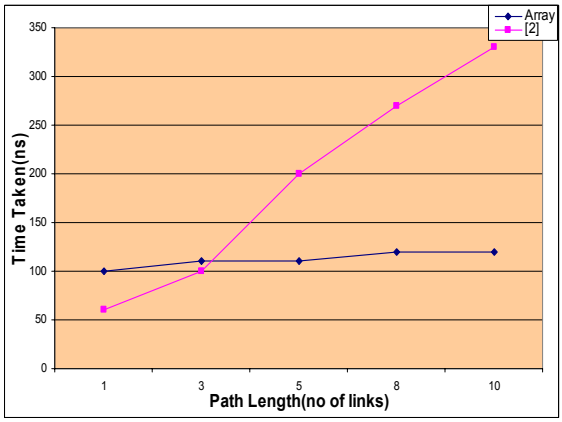


Chart 4: Delay vs. no of links traversed (path lengths) n=200, h=10

4.0 Conclusions

This paper considered MCDL networks , of the type G(n,+h,+1)G(n,+h,-1)G(n,-h,+1) G(n,-h,-1). We have critically analyzed the shortest path algorithms at [4 ,11]], that do not demand prior knowledge of diameter, for their suitability for adoption practical implementations on MCDL networks. Algorithm at [4] takes O[h/g+log(h)] where g = GCD(n,h) which is independent of diameter and computes shortest path in terms of type and number of links. Algorithm at [11] takes 'd' (diameter) time steps. Since diameter is dependent on the number of processors and hop size, the time complexity increases with increase in number of processors. Algorithm at [4] owing to diameter independence and flat and constant response times with increase in no of processors and path lengths is apt for computing shortest paths in MCDL networks than [11]. We have presented a novel method to find out optimum diameter, given n , the number of nodes , by employing curve fitting using power series. It is observed that the predicted values from the curve fitting approach and from the algorithm are almost same with an average error of 0.05 %

ACKNOWLEDGMENT

Authors would like to thank Gayatri Vidya parishad college of Engineering for providing all the facilities and encouragement and Prof(Dr.) NB Venkateswarlu for his valuable suggestions.

5.0 References

[1] B.W. Arden , H. Lee, Analysis of chordal ring networks , IEEE Trans. Compt. C-30(4)(1981)291-295.

[2] J.C. Bermond , F.Comellas , D.F.Hsu, Distributed loop computer networks : a survey . Journal of Parallel and Distributed Computing 24(1995) 2-10.

[3] F.T. Boesch, J.F Wang. Reliable circulant networks minimum transmission delays . IEEE Trans. Circuits Systems CAS-32(1985)11286-1291.

[4] N. Chalamaiah and B. Ramamurthy, "Finding shortest paths in distributed loop networks", Information rocessing letters, 67, pp. 157-161, 1998.

[5] CHEN Bao-Xing, Wenjun Xiao , Behrooz Parhami Diameter formulas for class of undirected double loop networks, Journal of Interconnection Networks, Vol. 6, No. 1 (2005) 1-15

[6] Krishnendu Mukhopadhyaya and Bhabani P. Sinha, "Fault-Tolerant Routing in Distributed Loop Networks", IEEE Trans. on Computers, Vol. 44, No. 12, 1995.

[7] G Praveen Babu, Ramesh Vasappanavara and Nimmagadda Chalamaiah "Comparative Study of shortest path routing algorithms in MCDL Networks" Proceedings of Asia Pacific International Conference ObComAPC-2004: Parallel and Distributed Computing Technologies, (December, 13-15, 2004:Vellore Institute of Technology, Vellore, India.

[8] Raghavendra and M Gerla,AAvizienis, "Reliable loop topologies for large local loop networks", IEEE Trans on Computers, 34(1) pp 46-55,1985.

[9] C.S Raghavendra , J.A. Silvester, double Loop Network architecture – a performance study , IEEE Trans. Comm COM-3392) (1985) 185-187

[10] S. Raghavendra , J.A. Silvester, A survey of multi-connected loop topologies for local computer network. Comput. Networks ISDN Systems 11(1986) 29-42

[11 Raj Katti and V.V. Bapeswara Rao, "An array based technique for routing messages in distributed loop networks", presented at : ISCAS 2002 IEEE International Symposium on Circuits and Systems, 26 -29 May, 2002, pages V 9-12, Raj Katti and V V Bapeswara Rao.

[12] Ramesh Vasappanavara, Sasikiran Kandula and Nimmagadda Chalamaiah "Fault Tolerance Studies" for wormhole routing in Multi connected Double Loop Networks", Asian Applied Computing Conference AACC-2004, Khatmandu, Nepal. (October, 28-31, 2004, Springer Veriag LNCS 3435, pp198-204)

[13] Ramesh Vasappanavara, Sasikiran Kandula and Nimmagadda Chalamaiah "Delay Performance Simulation Studies in MCDL Networks" International Conference on Systemics, Cybernetics, and Informatics ICSCI-2005, Hyderabad, India. (January, 06-09, 2005 Proceedings pp948-953).

[14] Suresh Chalasani and Rajendra V Boppana , " A Frame work for Designing a Deadlock-free wormhole routing Algorithms", IEEE Tr on Parallel and Distributed systems Vol 7 No 2 Feb 1996.

Authors

Ramesh Vasappanavara received his B.Tech degree in Electronics and Communication Engineering from Andhra University (AU) in 1977. He received M.Tech degree in Computer Science from Indian Institute of Technology, Kharagpur in 1984. He has worked as Commander, Naval Systems for and Defense R&D for twenty-five years prior to joining as Professor and Head of the Department, Department of Computer Science, GVP college of Engineering, Visakhapatnam, India in 1999. His research interests include parallel and distributed systems, multimedia and network security

EV Prasad received his PhD from Indian Institute of Technology Chennai India in 1990. He has served as Professor and head of Department of Computer Science at College of Engineering Anantapur , Andhra Pradesh India. Since June 2005 , he is working as Vice Principal JNTU College of Engineering Kakinada . He has several research publications to his credit. He has served as member review committee for IEEE Transaction on Reliability . His areas of research interests include fault tolerant studies , Software engineering , and soft computing.

M N Seetaramnath received his PhD from Andhra University Visakhapatnam in the year 1972. He has worked as Systems Programmer at Electronics Corporation of India Limited. Subsequently in the year 1977, he joined as Asst Professor in the Department of Computer Science at Indian Institute of Technology Kharagpur . Since 1985, he is Professor Department Computer Science and Engineering Andhra University College of Engineering Viakhapatnam. His areas of interest include Operating Systems, Systems Programming, Parallel and Distributed Computing, and Data Warehousing and Data Mining.

Automatic Injection Profile Prediction by Soft Computing Methods

Wei, Mingzhen; and Sung, Andrew H.

Abstract—*Soft computing methods are well known for their applications in the petroleum industry. Injection profile is one of the most important factors that need be taken into consideration in major oilfield development or related decision making. Injection profile is affected by multiple geological and developmental factors, which form complicated nonlinear relationships that are difficult to model by using conventional approaches. In this paper, an intelligent system is presented to construct fuzzy inference systems (FIS) automatically by integrating the fuzzy subtractive clustering and the Adaptive-Network-based Fuzzy Inference System (ANFIS). The method was tested by injection profile prediction in the Daqing Oilfield in China. A data cleaning strategy, the TANE algorithm, was applied to identify and to remove inconsistency in the raw data set collected from the oilfields. Results demonstrated that cleaned data produced more robust FIS and achieved higher prediction accuracies. The same approach can be applied in intelligent systems in resolving complicated Internet problems.*

Index Terms—*soft computing, ANFIS, fuzzy subtractive clustering, approximate dependency mining, conflicting data patterns*

1. INTRODUCTION

Soft computing techniques have been applied in many areas of the petroleum industry, such as reservoir characterization [1,2,3], well log interpretation [4,5,6], production prediction [7] and treatment optimization [8].

Injection profile of injection wells in water flooding oilfields is one of the most important factors in oilfield development. In water-flooding oilfields, injected water drives petroleum fluid (oil, gas and water) to move towards the wellbore by pushing the oil/gas/water in the porous media underground. Understanding injection profiles significantly aids in analyzing key petroleum production problems, such as residual oil distribution, residual reserve estimation, water flooding efficiency, water injection balancing and so on.

Many methods can be applied to obtain injection profiles in oilfields [9, 10], such as sealed coring, sidewall coring, logging data interpretation, C/O spectral logging, numerical simulation and comprehensive analysis of static and dynamic data from the oilfield development. Most of those methods, except for numeric simulation, are for obtaining injection profiles by in place measurement and interpretation. They are expensive and time-consuming. In addition, it is impossible to obtain injection profiles whenever and wherever they are needed for improving oil recovery (IOR) purposes. Reservoir numeric simulation models the oil/gas production by combining petroleum fluid flow and other models. By properly modeling reservoir and matching the history production data, reservoir simulation produces injection profiles in the production history and predicts injection profiles in the given future. However, reservoir simulation has its own inherent problems, including that: 1) modeling multiple parameters and the integration of sub-models are difficult; 2) history matching is actually a trial-and-error and time-consuming process which depends on reservoir simulation expertise intensively; and 3) reservoir simulation sometimes encounters difficulties in modeling actual reservoir features due to built-in limitations in models. In addition, time-consuming post-processing is required to obtain injection profile data from reservoir simulation results. Considering that injection profiles are required in many different IOR projects, it would be nice to have handy data available when it is required.

Injection profile prediction using soft computing methods was reported in [11]. The paper integrated the subtractive clustering and the ANFIS methods to construct optimized FIS automatically using available data from the Daqing Oilfield. In this paper, we briefly introduce the profile injection problem in water flooding oilfields and describe the problem modeling considerations in more detail. In order to enhance the performance of soft computing methods, the TANE algorithm was applied to identify implicitly conflict data patterns in the raw data. The similar problem modeling procedure and same data were used. Compared with results presented in [11], improved injection profile prediction accuracy was achieved. It shows the importance of data quality in problem modeling using soft

Manuscript received September 17, 2005. This work was supported in part by the New Mexico Petroleum Recovery Research Center in New Mexico, USA.

M. Wei, the response author, is with the Computer Science Department, New Mexico Institute of Mining and Technology, Socorro, New Mexico (email: mingzhen@nmt.edu).

Prof. A. Sung is with the Computer Science Department, New Mexico Institute of Mining and Technology, Socorro, New Mexico (email: sung@cs.nmt.edu).

computing methods, especially when the available training data set is small. The same method can be applied in other complicated problems in the petroleum industry, such as production predict and reserve evaluation.

The rest of this paper is organized as follows. Section 2 presents considerations in soft computing method selection and the problem of injection profile prediction; Section 3 presents the methodology used in injection profile modeling and prediction; Section 4 summarizes the experimental results and Section 5 concludes the paper.

2. PROBLEM STATEMENT

In water flooding oilfields, the distribution of injected water along the producing strata, the injection profile, is tightly related to oil/gas production from producing wells. Water injectivity in active producing strata is affected by many parameters, such as permeability of formations, the communication of injection and producing wells, injection pressure differences, well patterns, and so on. Because the fluid flow in the porous media in reservoir follows the non-linear Darcy's Law and these factors interact complicatedly, it is difficult to model their relationships with the injection profile data by conventional approaches. To improve the injection profile prediction efficiency and accuracy, soft computing methods are applied.

2.1 Parameter Selection

In problem modeling using soft computing methods, problem formulation and decomposition are equally important. Parameter selection determines how the problem is modeled and resolved. In the influential parameter selection, following points should be considered:

- Selected parameters affect the target problem and the target parameter. Strong relationships, linear or non-linear, must exist among selected parameters and target variables;
- Selected parameters must be well-populated and corresponding data must be as clean as possible. Since the soft computing methods model problems based on available data, the data availability and quality are essential for successful modeling.

In order to model and predict injection profiles, above mentioned factors, formation permeability, communication of injection and producing wells, well patterns and production setups, should be considered. In order to filter proper influential parameters, injection profile data from 25 wells, totally 218 active strata, was analyzed. Following parameters were selected:

- Gross sand thickness near the wellbore of injection wells, denoted as h_{gross1};
- Net sand thickness near the wellbore of injection wells, denoted as h_{net1};
- Gross sand thickness near the wellbore of nearby producing wells, denoted as h_{gross2};
- Net sand thickness near the wellbore of nearby producing wells, denoted as h_{net2};
- Spacing distance between injection wells and surrounding producing wells, denoted as d.

The first four parameters reflect the communication between injection and producing wells. Well spacing distance reflects the effect of well pattern and production criteria, the larger the well spacing, the smaller the injection capability.

Formation permeability of active strata is another key factor that affects the water injectivity and the injection profiles. Injected water moves faster in the direction of higher permeability, and breaks through in producing wells in high-permeability zones. Studies on available data show that absolute permeability of active strata is positively related to the sand type in the Daqing Oilfield. Sand types are embodied by the thickness of sand and communication of injection and producing wells, as shown in Table 1. In addition, permeability is not widely available in our tested area. Therefore, permeability is not considered in problem modeling.

Thickness of sand-body (m)	<0.5 gross	≥0.5 gross	0.2-0.5 net	0.5-1.0 net	1.0-1.5 net	1.0-1.5 net
Average permeability (um^2)	0.037	0.123	0.264	0.802	1.064	2.181

Table 1: Relationship of permeability and sand thickness in the active producing strata

2.2 Problem Formulation

The relationship of injection profile and selected parameters is not obvious. Injection profile is calculated by summing up relative water intakes of producing wells perforated in each active stratum, formulated as follows:

$$ri_i = RI \times ratio_i / \sum_i ratio_i \quad (1)$$

$$ratio_i = h_{net2_i} + (h_{gross2_i} - h_{net2_i}) / 3.3 \quad (2)$$

where i=1,2,…, refers to one of surrounding producing wells of injection wells. Hence in the problem modeling, the input is h_{gross1}, h_{net1}, h_{gross2}, h_{net2}, d, and the output is ri_i. The resulting FIS models the relationships of input and output data.

The injection of an active stratum is calculated as follows:

$$RI_k = \sum_i ri_{ki} \quad (3)$$

where k is the index of a producing stratum of an injection well, and i is the index of surrounding producing wells of the injection well in the producing stratum.

Predicted injection rates are compared with corresponding measured rates. If two rates have small difference, as in

$| RI_k^{predicted} - RI_k^{measured} | \le threshold$, where *threshold* is defined based on requirement in the petroleum industry, it is correctly predicted. Prediction accuracy is evaluated by percentage of correctly predicted injection rate using following formula:

$$accuracy = \frac{\#\ of\ \ | RI^{predicted} - RI^{measured} | \le threshold}{\left\| predicted\ set \right\|} \times 100 \quad (4)$$

In injection profile prediction, a bias of 2% to injection rate meets the precision requirements. Therefore, 2% is taken as a threshold to calculate the prediction accuracy.

3. THE METHODOLOGY

3.1 Soft Computing Method Selection

Fuzzy logic (FL) was introduced by Zadeh in 1965 [12], which processes data using partial set membership rather than crisp set membership. Similar to neural networks, FL is able to generate definite conclusions based on vague, ambiguous, imprecise and missing input information. One essential task in constructing proper FIS is to provide correct fuzzy membership functions and fuzzy rules. It is a time-consuming task, and requires profound expertise for a given problem. In addition, it is oftentimes difficult to convert domain knowledge into if-then fuzzy rules, even for domain experts.

Some methods [13,14] have been proposed to learn fuzzy membership functions and fuzzy rules and further FIS automatically by analyzing available input-output data. ANFIS [15] is a sophisticated neuro-fuzzy system and is able to model complicated fuzzy relationships. It learns the Sugeno-Takagi (or ST) FIS [16] using forward and backward passes, as shown in Fig. 1. The ST FIS is formatted as equations (5) and (6). ANFIS is an efficient approach to construct FIS for a given problem using available data for training.

$$R^1 : if\ \ x_1\ \ is\ \ A_1^1\ \ and\ \ x_2\ \ is\ \ A_2^1, then\ \ f_1 = p_1 x_1 + q_1 x_2 + r_1 \quad (5)$$

$$R^2 : if\ \ x_1\ \ is\ \ A_1^2\ \ and\ \ x_2\ \ is\ \ A_2^2, then\ \ f_2 = p_2 x_1 + q_2 x_2 + r_2 \quad (6)$$

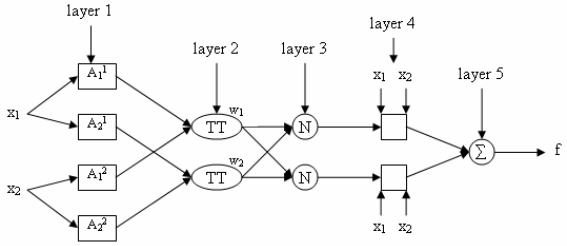


Fig. 1. The equivalent ANFIS architecture of the first order ST fuzzy inference system in equations (5) and (6).

To aid the construction of ANFIS, a clustering approach, the fuzzy subtractive clustering [17], is employed to cluster available data into fuzzy clusters. After fuzzy clusters are generated, an FIS can be generated based on the clustering results. Fuzzy subtractive clustering works better than other clustering methods for FIS. The resulting FIS can be refined further by the ANFIS.

In using the fuzzy subtractive clustering method to a given problem, main parameters that are required to adjust are the influential radii of a cluster center in the multiple dimensional data space. The method assumes that all data falls within a unit hyper-box. Smaller influential radii generate more clusters for further process, and larger radii generate fewer clusters. Different radius combinations generate different FIS. Hence it is important to select proper influential radii for a given problem.

In our method, the fuzzy subtractive clustering and ANFIS are integrated to model complicated real-world problems. This combination has been applied in automation control [18]. In order to save the trial-and-error radius adjustment, an automatic optimization process was employed to obtain the best combination of influential radius in our approach. Given injection profile prediction problem, a range of influential radius of 0.3~0.7 was used based on the precision requirement. Optimized radius combination was selected according to the root mean square errors (RMSE) from the training and validating data sets.

3.2 Conflict Data Pattern Identification

To identify conflict data patterns in the raw data, the TANE algorithm [19] was applied. The TANE algorithm analyzes the functional or approximate dependencies of different attributes in given data source. The results tell whether the key (e.g. primary key or compound key) is properly selected or data patterns are properly associated. Interesting results were observed, as reported in Section 4.

The TANE algorithm requires partition of continuous values for interesting domains. Data should be pre-processed before being put into the TANE algorithm. Based on the precision requirement for the injection profile prediction, the gross thickness and net thickness of strata near the wellbore of injection and producing wells are kept in its original representation. Well spacing is processed into discrete numbers using following rules:

$$d' = \begin{cases} 150 & 125 \le d < 175 \\ 200 & 175 \le d < 225 \\ 250 & 225 \le d < 275 \\ 300 & 275 \le d < 325 \end{cases} \quad (7)$$

The relative injection was kept in precision of 1%, which meets the precision requirement in the petroleum industry.

After all these processing, the TANE algorithm was run in APPROXIMATE mode. By running the TANE algorithm, four approximate dependencies are discovered using given error threshold. To

identify exceptional tuples, it is required to investigate the equivalence partitions of both left-hand and right-hand sides of approximate dependency. Results are shown in Section 4.

3.3 Injection Profile Modeling & Prediction

In modeling the injection profile problem, the fuzzy subtractive clustering and the ANFIS were integrated, and a parameter optimization strategy was applied to achieve best performance FIS for training and validating data. Problem modeling follows steps in Fig. 2.

Main data set used in this paper is from the South V District and the South II District in the Daqing Oilfield. It covers 10 injection wells, 53 active producing strata, and 45 producing wells. In problem modeling, totally 363 data points were sampled from the raw data. After data cleaning, 356 data points were left for analysis. In problem modeling, sampled data points were separated based on injection wells. Data from nine injection wells were used to construct the FIS; the left data were used to validate the resulting FIS. Average validation accuracy is calculated based on different training and validating data sets. In injection profile prediction, un-involved profile data from the same development district was used to calculate the prediction accuracy.

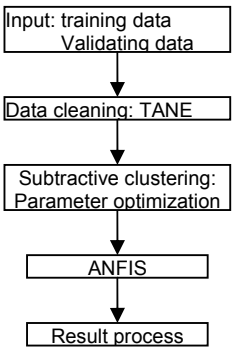


Fig.2. Flowchart of injection profile modeling using the fuzzy subtractive clustering and ANFIS methods.

4. EXPERIMENTAL RESULTS

4.1 Sample Data Set

Fig. 3 shows the complicated relationships of five selected parameters with relative injection.

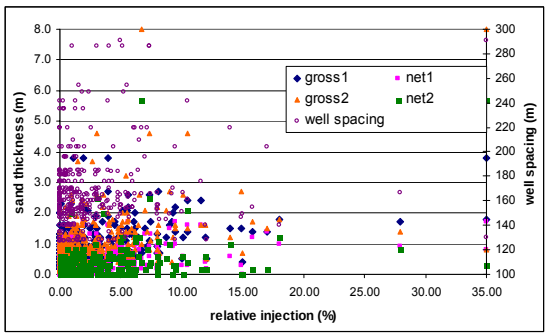


Fig.3. Relationships of selected parameters to relative injection of producing oil strata, where gross1 stands for h_{gross1}, net1 for h_{net1}, gross2 for h_{gross2}, net2 for h_{net2}, and well spacing for d.

4.2 Conflict Data Patterns in the Raw Data

Using the TANE algorithm in the APPROXIMATE mood and error threshold of 0.07, following approximate dependencies are identified, as shown in Table 2. Some other approximate dependencies can be discovered by using other error thresholds. The first dependency is the most important one because it shows that selected five parameters have consistent association relationship with the water intake per active layer except a few data tuples. To simplify and ensure the correct approximate dependency analysis, only the first dependency was analyzed for identifying conflict data patterns.

By analyzing the first dependency, following conflict tuples were identified out, as listed in Table 3. Highlighted pairs have large pattern conflicts, and bolded highlighted patterns were removed from the data set to study the effect of conflict patterns on the prediction models.

Index	Approximate Dependencies	# of Rows to Delete
1	$h_{gross1}\ h_{net1}\ h_{gross2}\ h_{net2}\ d' \rightarrow RI$	25
2	$H_{gross1}\ h_{net1}\ h_{gross2}\ h_{net2}\ RI \rightarrow d'$	20
3	$H_{gross1}\ h_{net1}\ h_{gross2}\ d'\ RI \rightarrow h_{net2}$	24
4	$H_{gross1}\ h_{gross2}\ h_{net1}\ d'\ RI \rightarrow h_{net2}$	23

Table 2: Approximate Dependencies Detected Using the TANE algorithm

Index	h_{gross1}	H_{net1}	h_{gross2}	h_{net2}	d'	RI
1	0.2	0	0.4	0	150	0
2	0.2	0	0.4	0	150	2
3	***0.2***	***0***	***0.6***	***0.5***	***150***	***0***
4	***0.2***	***0***	***0.6***	***0.5***	***150***	***3***
5	*0.2*	*0*	*0.4*	*0.2*	*150*	*0*
6	***0.2***	***0***	***0.4***	***0.2***	***150***	***5***
7	0.4	0.2	0.4	0	150	0
8	0.4	0.2	0.4	0	150	1
9	0.5	0	0.4	0	150	0
10	0.5	0	0.4	0	150	2
11	0.5	0.5	0.5	0.5	150	0
12	0.5	0.5	0.5	0.5	150	1
13	0.5	0	0.8	0	150	0
14	0.5	0	0.8	0	150	2
15	*0.5*	*0.4*	*1.0*	*0.4*	*150*	*1*
16	***0.5***	***0.4***	***1.0***	***0.4***	***150***	***6***
17	*0.6*	*0.2*	*0.5*	*0.2*	*200*	*1*
18	***0.6***	***0.2***	***0.5***	***0.2***	***200***	***6***
19	***1.3***	***1.1***	***1.2***	***0.4***	***150***	***0***
20	***1.3***	***1.1***	***1.2***	***0.4***	***150***	***3***

Table 3: Conflicting tuples identified by analyzing the first approximate dependency in Table 2.

From the results, it is obvious that the data set contained conflict relationships and associations among parameters. Some of them contained serious problems. For example, for specific pattern of h_{gross1}, h_{net1}, h_{gross2}, h_{net2} *and* d', the relative injection per active layer bear large difference, as <0.2, 0, 0.4, 0.2, 150, 0> in fifth tuple and <0.2, 0, 0.4, 0.2, 150, 5> in the sixth tuple, and <0.5, 0.4, 1.0, 0.4, 150, 1> in the

fifteenth and <0.5, 0.4, 1.0, 0.4, 150, 6> in the sixteenth. With same parameters, differences of relative injection were up to 5%. With these conflict data in the training data, it is difficult to model correct pattern association among these parameters.

4.3 Experimental Results

This method was applied to predict the injection profiles in the South II and V Districts in the Daqing Oilfield. Averagely, the prediction accuracy was up to 82%. Generally, prediction accuracy is high for thick and high- injectivity strata, and thin and low-injectivity strata; prediction accuracy is low for thick but low-injectivity and thin but high-injectivity strata. This is because that the assumption on the relationship of thickness of strata and their permeability is invalid in those cases.

The results were applied in analyzing the residual oil distribution and designing the secondary re-perforation in the South II District in the Daqing Oilfield. Promising results were observed, with increased recoverable reserve of 5,370,000 tons and increased recovery of 4.5%.

In this section, detailed results are reported to demonstrate the effectiveness of the new approach, compared with the FFBP Neural Networks, and the benefit of data scrubbing in problem modeling using soft computing methods.

Results using the Raw Data: The injection profile prediction was mainly implemented by the new approach and FFBP neural networks. FFBP neural networks are well known for their capability in modeling non-linear and complicated problems. The problem with neural network is that it is difficult to set up a proper neural network, including architecture and parameters, for a given problem. Fig. 4 shows results from different neural networks, either with different architecture or with different parameters. It is obvious that, with same training and testing data sets, different FFBP neural networks generate different validating results. The optimization is difficult. Table 4 shows the validating results of these neural networks.

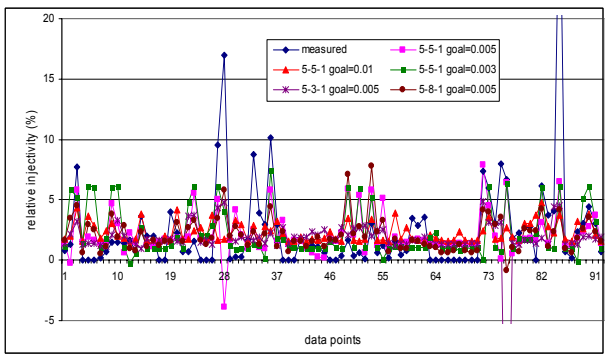


Fig. 4. Validating results from different FFBP neural networks for sampled testing data.

Modeling Methods	Threshold	
	1(%)	2(%)
5-5-1 goal=0.005	34.78	72.82
5-5-1 goal=0.01	35.87	72.82
5-5-1 goal=0.003	43.48	72.82
5-3-1 goal=0.005	42.39	72.82
5-8-1 goal=0.005	38.04	79.34
New approach	43.48	76.34

Table 4: Validating accuracies obtained from the new approach and different FFBP neural networks using the raw data

Results in Table 4 reflect the complexity in optimizing neural network setups. However, it is fairly easier to obtain an optimized model using the new approach. For our method, although it may not obtain highest accuracy in all thresholds, it achieved higher accuracy in most cases.

Table 5 shows the prediction accuracy using the FFBP and the new approach for the N2-D2-B447 well. The N2-D2-B447 is an injection well in the South II District. From its injection profile data, it has thin strata having large injectivity, such as S_{24a} and S_{25}, and thick strata having small injectivity, such as S_{216} and P_{21a}. These strata have poor prediction accuracy. Results show that, on same level of accuracy in validating process, new approach achieved higher accuracies under different precision requirements. Fig. 5 shows the predicted injection profiles of the well N2-D2-B447 using new approach and FFBP under configuration in the Table 4.

θ	5-3-1 goal=0.005	New Approach
1	27.8	41.7
2	41.7	75.8

Table 5. Injection profile prediction accuracy for the well N2-D2-B447 using the new approach and the FFBP neural network using the raw data

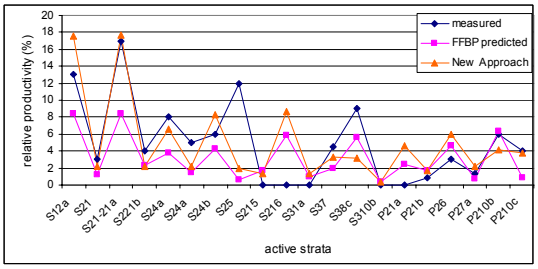


Fig.5. Predicted injection profiles for the N2-D2-B447 using the raw data.

Results with Cleaned Data: Results with cleaned data were obtained by re-running above experiments after removing bolded highlighted conflicting data patterns in Table 3. Table 6 shows the validating accuracy from the new approach and different FFBP neural networks using cleaned data. Table 7 shows the prediction accuracies from the new approach and the FFBP

neural network configured as "5-3-1 goal=0.005." These results show great improvements on accuracies under thresholds θ=1 and 2 for both the new approach and FFBP neural networks, and demonstrate the importance of consistent data used for problem modeling.

Modeling Methods	Threshold	
	1(%)	2(%)
5-5-1 goal=0.01	54.12	74.12
5-5-1 goal=0.005	54.12	75.29
5-5-1 goal=0.003	54.12	78.82
5-3-1 goal=0.005	57.64	77.65
5-8-1 goal=0.005	60	74.12
New approach	50.24	83.33

Table 6: Validating accuracies obtained from the new approach and different FFBP neural networks using the cleaned data

θ	5-3-1 goal=0.005	New Approach
1	30.56	54.44
2	72.22	83.33

Table 7: Prediction accuracy of the well N2-D2-B447 using the new approach and the FFBP neural network using the cleaned data

5. CONCLUSIONS AND DISCUSSIONS

Modeling complicated real-world problems is always full of challenge. In this paper, a new problem modeling approach is presented. The approach integrates two soft computing methods, the fuzzy subtractive clustering and the ANFIS, which enables it to construct a FIS automatically from training data set. A parameter optimization step is adapted to aid in generating optimized clustered data and hence the optimized FIS.

In problem modeling using soft computing methods, quality of training data is the key for stable results. Raw data obtained from real practice or process is always messed with different problems. Conflicting data is dangerous, especially when sampled data sets are small in size. It is shown by the prediction accuracy improvement in the practical problem.

The problem modeling strategy is justified by being applied in a critical problem in oilfield development – the injection profile prediction. Based on targeting problem, the detailed modeling procedure and methods could be different.

Compared with neural networks, this approach provides more promising results for intractable and complicated real-world problems. The new approach can also be applied to resolve internet related problem modeling, which are the hot topic of this decade.

ACKNOWLEDGMENT

The authors greatly appreciate the Bureau of the Daqing Oilfield, China for providing their data for research purpose. Financial support from the New Mexico Petroleum Recovery Research Center (PRRC) and the Computer Science Department of New Mexico Tech is also acknowledged.

REFERENCES

[1] Ouahed, K. El., Tiab, D., Mazouzi, A., Jokhio, S. A., "Application of artificial intelligence to characterize naturally fractured reservoirs," the SPE International Improved Oil Recovery Conference in Asia Pacific, 2003.

[2] Li, H-J, Li, F-S, Sung, A.H. "A fuzzy inference algorithm for lithology analysis in formation evaluation," Lecture Notes in Artificial Intelligence, Vol. 1821 (Intelligent Problem Solving: Methodologies and Approaches, R. Loganantharaj, et al Ed.), pp.290-295.

[3] Tamhane, D., Wong, P.M., Aminzadeh, F., Nikravesh, M., "Soft computing for intelligent reservoir characterization," the SPE Asia Pacific Conference on Integrated Modelling for Asset Management, 2000.

[4] Janoski, G, Sung, A. H., Grigg, R, "Petroleum Reservoir Simulation with Fuzzy Control", in Intelligent Engineering Systems through Artificial Neural Networks, Vol. 9, C.H. Dagli, et al Editors, ASME Press, New York, (Nov. 1999), pp.611-616.

[5] Lbatullin, R.R., Lbragimov, N.G., Khisamov, R.S., Podymov, E.D., Shutov, A.A., "Application and method based on artificial intelligence for selection of structures and screening of technologies for enhanced oil recovery," the SPE/DOE Improved Oil Recovery Symposium, 2002.

[6] Soto, R., Garcia, E., Torres, F., Perez, G.S., "Permeability prediction using hydraulic flow units and hybrid soft computing systems," the SPE Annual Technical Conference and Exhibition, 2001.

[7] Weiss, W. W., Balch, R. S., "How artificial intelligence methods can forecast oil production," the SPE/DOE Improved Oil recovery Symposium, 2002.

[8] Liu, Y., Bai, B., Li, Y., "Optimization design for conformance control based on profile modification treatments of multiple injectors in a reservoir," the SPE Symposium, 2000.

[9] Towler, B. F., "Fundamental principles of reservoir engineering," *Richardson, Tex.: Henry L. Doherty Memorial Fund of AIME, Society of Petroleum Engineers*, 2002.

[10] Lake, L. W., Carroll, H. B., "Reservoir characterization," *Orlando: Academic Press*, 1986.

[11] Wei, M., Sung, A. H., Wang, J., Liu, L., Zhang, S., Li, X., "Predict injection profile of an injector using a soft computing method," the 11th Abu Dhabi International Petroleum Exhibition and Conference, 2004.

[12] Zadeh, L. A., "Fuzzy logic, neural network, and soft computing," Communications of ACM, Vol. 37, pp.77-84, 1994.

[13] Keller, J., M., Tahani, H., "Implementation of conjunctive and disjunctive fuzzy logic rules with neural networks," International Journal of Approximate Reasoning, Vol. 6, pp.221-240, 1992.

[14] Takagi, H., Hayashi, I., "Artificial neural network driven fuzzy reasoning," International Journal of Approximate Reasoning, Vol. 5, pp.191-212, 1991.

[15] Jang, J., R., "ANFIS: Adaptive-network-based fuzzy inference system," IEEE Transactions on System, Man and Cybernetics, Vol. 23(3), pp.665-685, 1993.

[16] Takagi, T., Sugeno, M., "Fuzzy identification of systems and its application to modeling and control," IEEE Transactions on System, Man and Cybernetics, Vol. SMC-15, pp.116-132, 1985.

[17] Chiu, S., "Fuzzy Model Identification Based on Cluster Estimation," Journal of Intelligent & Fuzzy Systems, Vol. 2(3), Spet. 1994.

[18] Farag, W., A., Quintana, V., H., Lambert-Torres, G., "A genetic-based neuro–fuzzy approach for modeling and control of dynamical systems," IEEE Transactions on Neural Networks, Vol. 9, pp.756-767, 1998.

[19] Huhtala, Y., Kärkkäinen, J., Porkka, P., Toivonen, H., "TANE: an efficient algorithm for discovering functional and approximate dependencies," the Computer Journal, Vol. 42(2): pp.100-111, 1999.

Approximation Method for Extended Routing with Clue (ERC) Protocol over Meshed Networks

Viriyapho, Piboonlit; and Jittawiriyanukoon, Chanintorn

Abstract—***This paper compares the performance of the distributed IP lookup with conventional IP lookup process by applying to expandable meshed network. We simulated up to three meshed nodes network. Then, the performance matrices such as throughput, mean queue length (MQL) and utilization factor, are collected in order to explain the effect of expanding network (scalability) versus the performance of both networks. Results indicate that RC network outperforms vis-à-vis the conventional IP network in terms of expanding the network size. Moreover, the performance of the RC network is not dropping much compared to those of the IP network. This can be proven by extending the "O" notation as of a packet holding time, applying for both RC network and IP network. From mathematical point of view, the "O" notation as of time in packet holding for both networks will be derived. The calculations obtained from the notation are relevant to the simulation results. Furthermore, an approximation method is introduced. Results from the approximation method are close to results from simulation.***

Index Terms—***IP Routing, Routing with a Clue, Distributed IP Lookup, Mesh Topology and Extended RC.***

1. INTRODUCTION

ROUTING with a Clue[1] is the newly evolving routing technology introduced in the new millennium. Applying the distributed IP lookup process, the technique can reduce the excessive time used in redundant IP lookup for each node. Routing with a Clue utilizes a small overhead unit called "a clue" to transmit the IP lookup parameters from the current node to the next-hop node. The next-hop node learned from the clue shows where to start its lookup processes. Distributed IP lookup was already compared with the label switching process in Multiprotocol Label Switching (MPLS) [2], [15]. The two techniques share similar concept in adding a small unit, a label in MPLS and a clue in distributed IP lookup, onto the packet header to help routing the packet to the destination.

In this paper, an overview will be firstly introduced in section 2. The simulation will be briefly cited in section 3. Results from simulation are shown in section 4. An approximation method and the calculation are discussed in section 5. Section 6 will focus on the comparison between results from approximation and results from simulation. Conclusion is written in section 7.

Manuscript received September 27, 2005 and revised January 17, 2006.

Mr.Viriyaphol is with the Department of Telecommunications Science, Faculty of Science and Technology, Assumption University, Bangkok, Thailand (e-mail: p4519497@au.edu).

Dr. Jittawiriyanukoon is a leader of data communication research group in Faculty of Science and Technology, Assumption University, Bangkok, Thailand (e-mail: pct2526@gmail.com).

2. BACKGROUND

There are several directions to improve the throughput of routing packets from source to destination. For example, new hardware may be developed to increase forwarding interval [11] and IP lookup process [14], [16], parallel processing may be applied to process IP forwarding [10], new data structures may be used to store and manipulate the IP prefixes [6], [7], [9], [12], and new searching methodology [13] and algorithms [17] may be invented in order to process the packet faster. However, IP network is widely used and accepted worldwide. The new technique invented must be compatible with the conventional IP network.

Distributed IP lookup [1] is the extension of IP routing by adding 5 bits overhead into IPv4 packet [3]. The 5-bit clue is the encoded prefix of the packet destination address. The clue is used to acknowledge the next hop router where the router should begin searching. All prefixes are stored in a Trie [1] or Patricia data structure. The root of the tree represents the empty string. Each edge going to the left from a vertex represents 0, and an edge going to the right represents 1. Not all the vertices in the tree represent prefixes. The ones that represent prefix are specially marked. All the leaves of a trie are marked because those leaves that are not marked to be the prefix must be removed from the system.

Traditional IP lookup is performed by scanning the destination address bit by bit and matching it along the path of a trie. The computational cost for IP lookup is O(log L) where L is an IP address length [4],[5].

In this paper, distributed and conventional IP lookup processes are firstly introduced then compared in the three different mesh networks. We applied a mesh network of 4 nodes, then, we cascaded another mesh network with an additional mesh consisting of 2 nodes each.

3. Simulation Model

In this paper, we simulate Routing with a Clue network and conventional IP network by utilizing the model as shown in figure 1. Only the first mesh consists of four nodes then another two meshes (consisting 2 nodes each) are attached to the first mesh. Each link has the maximum capacity limit to 1 Mbps. As the traditional RC network, at each node the packet holding time is next to zero and can be neglected. An extended RC (ERC) network with inclusive packet holding time will be discussed in section 4.

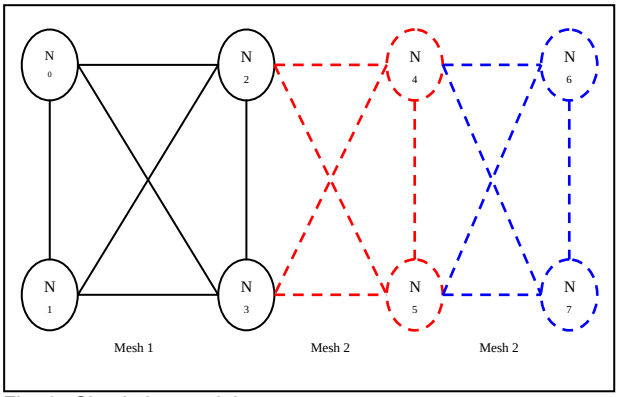


Fig. 1. Simulation model.

3.1 Input Traffic

We varied the input data rates for the simulation into several rates ranging from 500 kbps to 2 Mbps. The arrival process of the input traffic follows Poisson distribution. [2], [6].

3.2 Queue

All queues in front of each node found in this paper are assumed to be of First-In-First-Out (FIFO) discipline. Both queue capacity and the waiting time in queue are unlimited.

3.3 Packet Holding Time

The binary search is applied to the IP lookup [4] process resulting in the computational cost of O(N*log L) steps where L is the IP address length and N is number of prefixes used by hash tables. However, in distributed IP lookup, the processes are distributed to each node along the path then only O(log L) computational cost will become. Therefore, in ERC network, this packet holding time O(log L) is included in the account.

4. Simulation Results

4.1 One Mesh

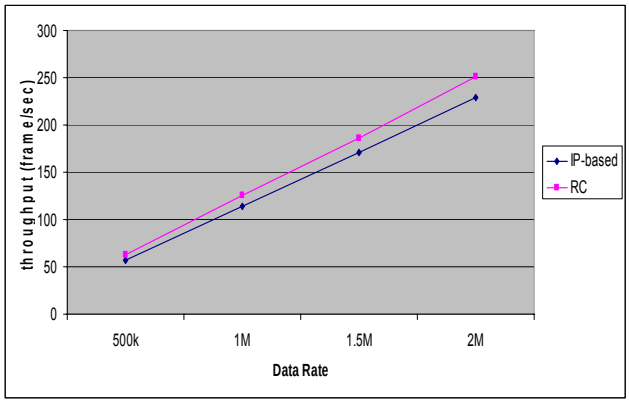


Fig. 2. Throughputs for 1 mesh.

From figure 2, the throughput of RC network is outnumbered by that of the conventional IP network in every input data rate. This is because the IP lookup process has been distributed to every node along the path and has to be repeatedly performed at every single node. Since this IP process repetition will not affect throughputs of both networks much, there will be no significant differences while the number of nodes is small as well as at low input data rates.

4.2 Two Meshes

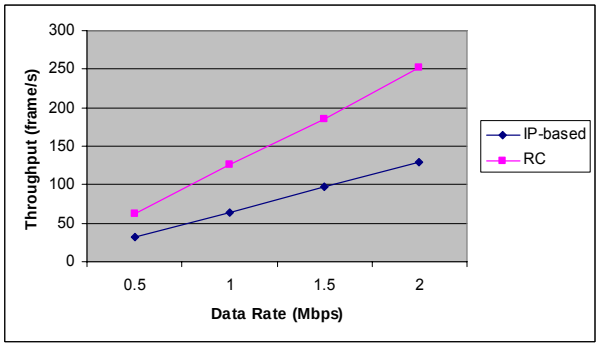


Fig. 3. Throughputs for 2 meshes.

In the two-mesh network, the difference between the throughputs of RC and IP networks can be observed more obviously than in one mesh network as shown in figure 3. As the node is increased, the throughput of IP

network dropped, but RC network still maintained the level of its throughput to remain unchanged from the one mesh network. The reason is because RC network's performance relies only on address length, but for conventional IP network's performance relies on both the address-length and numbers of prefixes. In the two-mesh network, increase in number of nodes results the increasing of prefixes in data transmission.

4.3 *Three Meshes*

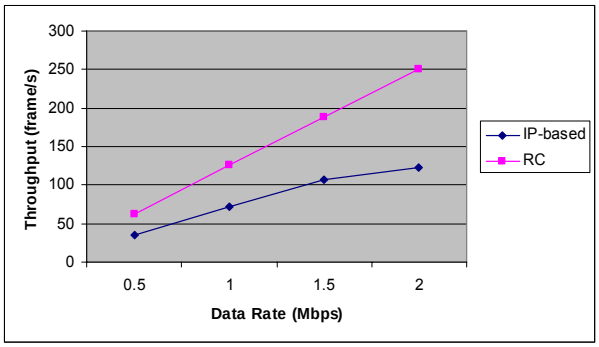


Fig. 4. Throughputs for 3 meshes.

The results from every parameter in every data rates and every node still indicate that distributed IP lookup outperforms compared to conventional IP lookup. RC throughputs are distinctly higher than IP. In the three-mesh network, the increasing number of nodes has a major impact on the performance of conventional IP network. The throughputs of IP network dropped dramatically because of the growing numbers of prefixes, while it has no effect on the RC network's performance. From figure 4, RC throughputs remain unchanged compared to those results from one-mesh and two-mesh networks.

5. *Approximation Method*

1.

In the previous section, the simulation of two address lookup techniques has been investigated. The results of simulation illustrate the difference of the two techniques. Several parameters have been carried out from the simulation program. However, to develop, to use, and to learn how to use the simulation programs are time-consuming processes. As the simulation is complicated and resource-consuming process, an approximation [18] tool can be an easy process that is less time-consuming and utilizes small numbers of calculation or computing process. The approximation method then will be introduced in this section.

Each node in the model as shown in figure 1 will be assumed to accommodate an input as Poisson arrival process.

$$f(x) = \frac{1}{x!}\lambda^{x} e^{-\lambda}\ , \text{x=1,2,3....} \quad (1)$$

where λ is a mean arrival time. In the simulation, input data rate λ will be ranging from 0.5 Mbps up to the capacity of the network 2 Mbps. Another assumption is based upon the service distribution in each node. Here the Erlang distribution is employed.

$$g(x) = \frac{\mu^{n} x^{n-1} e^{-\mu x}}{(n-1)!} \quad (2)$$

where x=1,2,3…and n is a positive integer.

In the simulation, μ will be ranging by the cost O(N log L), where N is total number of prefixes used by the forwarding table and L is the address length.

From the assumption of input (equation 1) arrival process and the service rate (equation 2), mean queue length (MQL_0) at the very first node can be estimated by

$$\mathrm{MQL}_0 = \frac{\lambda^2 E[X^2]}{2(1-\lambda E[X])}\ . \quad (3)$$

Note that an output from a particular node will proceed to next node then becomes an input arrival process to next adjacent node.

$$\mathrm{MQL}_k = \frac{\lambda_k{}^2 E[X^2]}{2(1-\lambda_k E[X])}\ . \quad (4)$$

Similarly, utilization factor (ρ) at node k can be approximated by

$$\rho_k = \frac{\lambda_k}{\mu_k}\ . \quad (5)$$

Finally, throughput (THRU) at node k can be computed by

$$\mathrm{THRU}_k = \lambda_k\,(1\text{-}E(T_k))\ , \quad (6)$$

where $E(T_k)$ is an average time T_k that IP frames reside in node k.

The iterative computation of each node's (k=0,1,2,…,k-1) performance measures (MQL_k, UTILIZATION(λ_k) and $THROUGHPUT_k$) using equation 1-6 has been developed. The node-by-node estimation is decomposed based upon the above calculation formulae. Algorithm used for this iterative computation is listed in figure 5. Results from approximation method by altering number of nodes (mesh) are then compared vis-à-vis simulation results. It can be seen clearly that approximation results are outperforming in the line of simulation as shown in figure 6.

```
/************ALGORITHM************/

BEGIN  /*** Use formulae to compute
              performance measures for
              each node in meshed network ***/
    FOR n = 0 TO k-1 DO /*** k nodes ***/
      Compute mean queue length MQL_n
      Compute utilization factor UTIL_n
      Compute throughput THRU_n

    END FOR

END
/***** END OF ALGORITHM **********/
```

Fig. 5: Algorithm of approximation method.

The advantages of employing the approximation method are not only time consuming, but also involve computational cost. The approximation technique needs least computational cost because the output parameters can be iteratively computed by mathematic functions. In some cases, the approximation method may give an enormously different result due to a bottleneck situation (ρ_k is approaching 1). Unfortunately the approximation method will not be a useful tool for the case as such. It is also clearly seen that if the percentage of error between approximation and simulation can fall in the range of 10 percent then the approximation results will be very similar.

The approximation method will be applied for the distributed IP lookup technique introduced in the previous section, and will be compared to results from the simulation program. From our simulation model, we will work out the computational cost for both techniques, and calculate the percentage of error.

6. COMPARISONS

In this section, we will investigate the approximation and simulation results applying to the extended Routing with a Clue (ERC) network. Here the packet holding time is taken into account. The experiments are based on one- mesh to three-mesh network, and varying the data rates from 500 kbps to 2 Mbps.

6.1 One Mesh

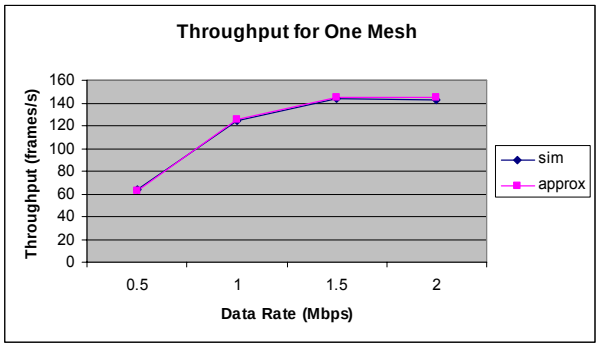


Fig. 6. Throughputs for 1 mesh.

From figure 6, we observe that approximation throughputs are very close to the simulation results. With the percentage of error shown in table 1, the approximation results error ranges from 0.22-1.7% which is very small compared to the simulation results.

Table 1. Throughputs Percentage of Error.

Data Rate (Mbps)	One Mesh Error (%)	Two Meshes Error (%)	Three Meshes Error (%)
0.5	**1.7**	**0.19**	**-0.68**
1	**-0.22**	**1.01**	**0.65**
1.5	**-0.57**	**-0.63**	**-1.9**
2	**-1.17**	**-1.45**	**-0.76**

6.2 Two Meshes

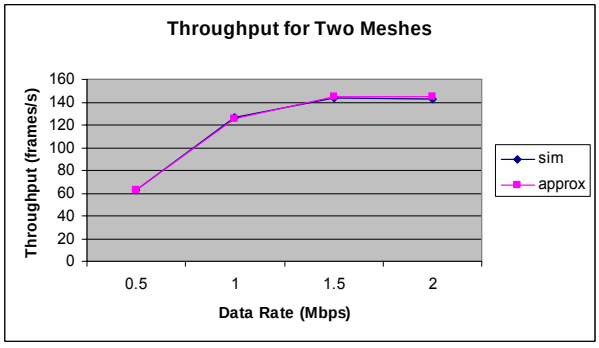


Fig. 7. Throughputs for 2 meshes.

We can obviously see from figure 7 that throughputs from both approximation and simulation techniques are still very close with the error between 0.17 to 1.4% as listed in table 1. Percentage of error for mean queue length and utilization factor is also listed in table 2 and 3.

Table 2. MQL Percentage of Error for Two Meshes.

Data Rate (Mbps)	One Mesh			Two Meshes		
	Sim	Approx	Error (%)	Sim	Approx	Error (%)
0.5	11	0	100	8	0	100
1	54	0	100	36	0	100
1.5	4125	4242	**-2.84**	150	0	100
2	4184	4200	**-0.38**	132	0	100

Table 3. Utilization Factor Percentage of Error for Two Meshes.

Data Rate (Mbps)	One Mesh			Two Meshes		
	Sim	Approx	Error (%)	Sim	Approx	Error (%)
0.5	43	43.12	**-0.28**	43	43.12	**-0.28**
1	88	86.25	**1.99**	88	86.25	**1.99**
1.5	100	100	**0**	98	100	**-2.04**

6.3 Three Meshes

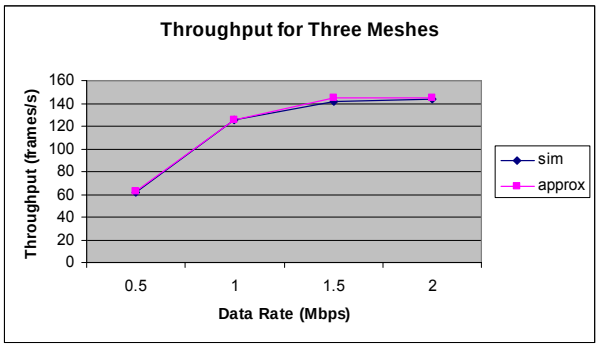


Fig. 8. Throughputs for 3 Meshes.

The throughputs of three meshes coming from both techniques still confirm the close of approximation technique with small percentage of error less than 2% as shown in figure 8.

With the data in table 5, the error percentage for approximation techniques is less than 3%.

From experiments on throughputs and utilization factor, the percentage of errors ranges from zero, exactly the same as the simulation results, to 4%. In MQL approximation, there are some cases that the result of the technique equal to zero which directly affect the calculation of percentage of error. The result from zero value in MQL approximation causes 100% of error. Except in the special case of zero MQL, the approximation technique results are less than 9% of error as listed in table 4.

Table 4. MQL Percentage of Error for Three Meshes.

Data Rate (Mbps)	One Mesh	Two Meshes	Three Meshes
	Error (%)	Error (%)	Error (%)
0.5	100	100	100
1	100	100	100
1.5	**8.74**	100	100
2	**-3.78**	100	100

Table 5. Utilization Factor Percentage of Error for Three Meshes.

Data Rate (Mbps)	One Mesh	Two Meshes	Three Meshes
	Error (%)	Error (%)	Error (%)
0.5	**-2.67**	**-0.28**	**2**
1	**0.86**	**0.86**	**-0.29**
1.5	**0**	**-2.04**	**-3.09**
2	**0**	**0**	**-2.04**

With acceptable error percentage, approximation technique needs no computational cost or simulation time according to table 6. In simulation technique, up to 100 sec of simulation time is used to conduct the simulation output, but no time is needed for approximation technique. Moreover, to develop or to learn how to use the simulation program is a complicated task, and time-consuming process. We propose the application of the easier approximation method, which requires only mathematical knowledge to solve some formulae derived from techniques needed for performance comparison. The error of 4-9% from the technique can be overcome by quick results achieved from the technique.

Table 6. Computational Cost for Simulation and Approximation Techniques.

Data Rate (Mbps)	One Mesh		Two Meshes		Three Meshes	
	Sim (s)	Approx (s)	Sim (s)	Approx (s)	Sim (s)	Approx (s)
0.5	100	0	100	0	100	0
1	100	0	100	0	100	0
1.5	50	0	100	0	80	0
2	40	0	40	0	40	0

7. CONCLUSIONS

When the data rates are increased, throughputs in both RC and IP networks also increased, but the throughputs of RC network are better than that of IP network. The network traffic fluctuated into the first mesh causing nodes in the first mesh to process more traffic than the following meshes, so performance parameters for nodes in the first mesh are not as good as those in the following meshes. We can obviously perceive that nodes in the last mesh will have very low traffic, so there are very few packets waiting in their queues. Moreover, in the meshed network, as we increase the number of nodes by cascading another two nodes to the existing mesh network, RC network can maintain its throughputs in all meshes. In ERC network, RC network with O notation for packet holding time, the performance relied only on the address length of the packets. In contrast, the performance of IP network relied on both address length and number of prefixes in the table. As number of nodes increases, the number of prefixes in the hash table also increases, so we can assume that number of nodes directly influences the performance of IP network. This results in the IP network's throughputs to be lower as the number of nodes in the backbone is increased.

With simulation technique, excessive time has been used for developing the simulation programs, and for simulating events in the model before we could get the performance measures from the simulation program. An approximation method has been proposed. Three performance measures, throughputs, MQL, and utilization factors can be collected from the approximation method. It applies mathematics to carry out all performance measures in no time. The error from the approximation compared to simulation technique ranges from 4% to 9%.

REFERENCES

[1] Afec, Y., Bremler-Barr, A., and Har-Peled, S., "Routing with a Clue," *IEEE/ACM Transactions on Networking* , 6, 2001, pp. 693 – 705.

[2] Viriyaphol, P. and Jittawiriyanukoon, C., "Evaluation of Routing Techniques over IP-based Network," *Proceedings of IASTED EuroIMSA*, 2005, pp.158-162.

[3] Chandramenon, G. and Varghese, G.," Trading Packet Headers for Packet Processing," *IEEE Trans. on Networking* , 4, 1996, pp 141-152 .

[4] Waldvogel, M., Varghese, G., Turner, J., and Plattner, B., "Scalable High Speed IP Routing Lookups," *Proceedings 27th ACM SIGCOMM*, 1997.

[5] Doeringer, W., Karjoth, G., and Nassehi, M., "Routing on Longest-Matching Prefixes," *IEEE/ACM Trans. on Networking*, 4, 1996, pp 86-97.

[6] Srinivasan V. ,and Varghese, G., "Faster IP Lookups using Controlled Prefix Expansion," *Proceedings ACM SIGCOMM*, 1998, pp. 1-10.

[7] Degermark, M., Brodnik, A., Carlsson, S., and Pink, S., "Small Forwarding Tables for Fast Routing Lookups," *Proceedings ACM SIGCOMM*, 1997, pp.1-14.

[8] Khoshnevis, B., "Discrete Systems Simulations", *New York: McGraw-Hill Publishing*, 1994.

[9] Sahni, S. and Kim, K., "An O(log n) Dynamic Router-Table Design," *IEEE Trans. Computers*, vol. 53, 2004, pp. 351-363.

[10] Akhbarizadeh, M. J. and Nourani, M., "An IP Packet Forwarding Technique Based on Partitioned Lookup Table," *Proceedings of IEEE Intl Conference on Communications*, May 2002.

[11] You, C., Chu, Y., Li, C., Su, H., "Fast IP Routing Lookups and Updates in Hardware," *Proceedings of Intl. Symposium on Communications*, November 2001.

[12] Buchsbaum, A., Fowler, G., Kirshnamurthy, B., Vo, K. and Wang J., "Fast Prefix Matching of Bounded Strings," *Proceedings ACM SIGACT ALENEX03*, January 2003.

[13] Kim, K. and Sahni, S., "IP lookup by binary search on length," *Journal of Interconnection Networks*,Vol. 3,2002, pp. 105-128.

[14] Taylor, D., Turner, J., Lockwood, J., Sproull, T., and Parlour, D., "Scalable IP Lookup for Internet Routers," *IEEE Journal on Selected Areas in Communication*, Vol 21, No. 4, May 2003.

[15] Rosen, MPLS Architecture, RFC 3031.

[16] Gopalan, K., and Chiueh, T., "Improving Route Lookup Performance Using Network Processor Cache," *Proceedings SC2002 High Performance Networking and Computing*, November 2002.

[17] Ioannidis, I., Garmar, A., and Atallah, M., "Adaptive Data Structures for IP Lookups," *Proceedings of IEEE INFOCOM*, March 2003.

[18] Yate, Roy D. and Goodman, David J. "Probability and Stochastic Processes," *John Wiley & Sons*, 1999.

Mr. Piboonlit Viriyaphol received the B. Eng. degree in Computer Engineering from Kasetsart University, Thailand, and the M.Sc. degree in Computer Information System from Assumption University, Thailand. He is currently a Ph.D. candidate in Telecommunications Science at Assumption University of Thailand. His research interests include IP network, network performance evaluation, routing algorithm and network of queue.

Mr. Chanintorn Jittawiriyanukoon received the B.Eng degree in electrical engineering from Kasetsart University, Bangkok, Thailand in 1984, the M.Eng. and D.Eng. in communications engineering from Osaka University, Osaka, Japan in 1987 and 1990 respectively. He is presently an assistant professor of faculty of science and technology, Assumption University of Thailand. He is also a leader in a data communication research group at the faculty. His research interests include satellite communications, performance evaluation of parallel processing, fast routing algorithm and high speed networks.

Reviewers by Countries

Australia
O'Dwyer, Mark F.; University of Wollongong
Brown, Simon; University of Tasmania
Currie, Carolyn V.; University of Technology Sydney
Pereloma, Elena; Monash University
Pinto, Dale; Curtin University
Soar, Jeffrey; University of Southern Queensland
Sweller, John; University of New South Wales
Unnithan, Chandana R.; Deakin University

Austria
Geser, Guntram; Salzburg Research
Mulrenin, Andrea M.; Salzburg Research
Wöß, Wolfram; Johannes Kepler University Linz

Belgium
Daele, Peter Van; IMEC-Ghent University
Tianrui, Li; The Belgium Nuclear Research Centre

Brazil
Fernandes, Marcia Aparecida; University of Uberlandia
Wazlawick, Raul Sidnei; UFSC

Canada
Bokhari, Asghar; McMaster University
Feng, Franc; University of British Columbia
Hansebout, Robert R.; McMaster University
Holroyd, Carin; Asia Pacific Foundation of Canada
Jirgens, Karl E.; University of Windsor
Jung, Benjamin; University of Victoria
L'Ecuyer, Pierre; Université de Montréal
McLeod, Bob
Peters, Dennis K.; University of Newfoundland
Vaillancourt, Remi; University of Ottawa
Wesolowsky, George O.; McMaster University
Wolfe, Dalton L.; Lawson Health Research Institute

Chile
Bedini, Alejandro; Practia

China
Cao, Bing-Yang; Tsinghua University
Chung, Walter W.C; Hong Kong Polytechnic University
Guo, Zeng-Yuan; Tsinghua University
Lau, Henry C. W.; Hong Kong Polytechnic University
Li, Lemin; University of Electronic Science
Luo, Yingwei; Peking University
Tsang, Albert H.C.; Hong Kong Polytechnic University
Yin, TANG Chak; Hong Kong Polytechnic University
Zhuoqun, Xu; Peking University

Croatia
Jovic, Franjo; University of J.J. Strossmayer

Czech Republic
Brandejsky, Tomas; Czech Technical University
Kaluza, Jindrich; Technical University of Ostrava
Merunka, Vojtech; Czech University of Agriculture

Finland
Mattila, Minna Marita; Tampere Technical University
Repo, Petteri; National Consumer Research Centre
Tico, Marius; Nokia Research Center

France
Lepage, Alain; University of Technology of Compiègne
Rossi, Fabrice; INRIA Rocquencourt

Germany
Baum, Dieter; University of Trier
Brunkhorst, Hauke; Flensburg University
Buss, Claudia; University of Trier
Hirzinger, Gerd; Institute of Robotics and Mechatronics
Köhler, Stefan; Infosim GmbH & Co. KG
Morschett, Dirk; Saarland University
Nordmann, Alfred; Darmstadt Technical University
Ragia, Lemonia; Technische Hochschule Aachen
Rennings, Klaus; ZEW Mannheim
Tschofenig, Hannes; SIEMENS Corporate Technology
Unland, Rainer; University of Duisburg-Essen
Weiß, Christian; University of Würzburg

Greece
Kosmidis, Mary H.; Aristotle University of Thessaloniki
Moulianitis, Vassilis C.; University of Patras

Hungary
Jakabfi, Anna; Eötvös Loránd University
Lorincz, Andras; Eötvös Loránd University

India
Chatterjee, Arnab; Saha Institute of Nuclear Physics
Maheshwari, Neeru; Birla Insititute of Technology
Roy, Debanik; Bhabha Atomic Research Centre
Sreeram, Pazynur Agneswaran; Bose National Centre

Israel
Levine, Adina; Bar-Ilan University
Fund, Zvia; Bar-Ilan University
Haberman, Bruria; Holon Institute of Technology
Koren, Shira; Bar Ilan University

Italy
Appice, Annalisa; University of Bari
Chella, Antonio; University of Palermo
Chella, Antonio; University of Palermo
Ferraresi, Pier Marco; CERP
Fogli, Daniela; University of Brescia
Furcolo, Pierluigi; University of Salerno
Jacobsen, Eric; Università degli Studi di Bologna
Kenett, Ron S.; University of Torino
Mello, Paola; University of Bologna
Papini, Pier Luigi; Università di Bologna
Robustelli, Anton Luca; University of Salerno
Salomoni, Paola; Universita' di Bologna
Santucci, Fortunato; Università de L'Aquila

Japan
Sasakura, Mariko; Okayama University
De Marco, Giuseppe; Fukuoka Institute of Technology
Fujimoto, Isao; Toyota Technological Institute
Okamoto,Toshio;University of Electro-Communications
Matsuo, Hiroshi; Nagoya Institute of Technology
Robb, Thomas N.; Kyoto Sangyo University
Tada, Yasunori; Osaka University

Jordan
Salhieh, Sa'Ed M.; The University of Jordan

Malaysia
Lee, Goi Chai; Curtin University of Technology
Yeap, Boon Han; Monash University

México
Domínguez, Humberto Ochoa; Universidad Autónoma
Mortera-Gutierrez, Fernando; ITESM-CCM

Netherlands
Bregt, Arnold; Wageningen University
Eldin, Amr Ali; Delft University of Technology
Leemans, Rik; Wageningen University and Research
Vermetten, Eric; University Medical Center

New Zealand
Elley, Raina; University of Otago
Huff, Sid; Victoria University of Wellington
Scornavacca, Eusebio;Victoria University of Wellington

Norway
Hoogensen, Gunhild; University of Tromsø

Pakistan
Khan Sial, Sultan Muhammad; Lahore University

Poland
Pakszys, Elzbieta; Adam Mickiewicz University
Pakienkiewicz, Julek; Warsaw University
Qatar
Kassim, Norizan Mohd; University of Qatar
South Korea
Kim, Cheol-Han; Daejon University
Singapore
Bressan, Stephane; National University of Singapore
Yau, Wei-Yun; Institute for Infocomm Research
South Africa
Blignaut, Pieter; University of the Free State
Sweden
Barse, Emilie Lundin; AerotechTelub AB
Stenfors-Hayes, Terese; Karolinska Institutet
Switzerland
Camponovo, Giovanni; University of Lausanne,
FISCHBACH, Nicolas; COLT Telecom
Goyette, Stéphane; University of Fribourg
Mutti, Carlo; Datamars SA
Papadimitratos, Panagiotis; EPFL
Spaccapietra, Stefano; EPFL
Studer, Thomas; University of Bern
Taiwan
Chang, Yang-Lang; National Taipei University
Chen, Hengsheng; Chinese Culture University
Hsieh, Ching-Tang; Tamkang University
United Kingdom
Emes, Michael; Mullard Space Science Laboratory
Rees, Philip Howell; University of Leeds
Walker, Stuart D.; University of Essex
Ali, Maaruf; Oxford Brookes University
Bailey, Stephen J.; Glasgow Caledonian University
Barnes, Stuart J.; University of East Anglia
Brewster, Stephen; University of Glasgow
Coolen, Frank; Durham University
Coolen-Schrijner, Pauline; Durham University
Jiang, Jianmin; University of Bradford
Kelbert, Mark; University of Wales-Swansea
MacDonald, Theodore; London University
Wright, Patricia; Cardiff University
USA
Assavapokee, Tiravat; University of Houston
Backer, Larry Catá; Pennsylvania State University
Beck, Kent; Three Rivers Institute
Benotsch, Eric G.; University of Colorado
Bloomfield, Peter; North Carolina State University
Champion, Jane Dimmitt; The University of Texas
Christie, William G.; Vanderbilt University
Chun, Mark W.S.; Pepperdine University
Cole, David N.; Leopold Wilderness Research Institute
Conn, Cynthia Anne; Northern Arizona University
Coppola, Jean F.; Pace University
Dai, Qizhi; Drexel University
Derzon, Jim; Econometrica Inc.
Dholakia, Nikhilesh; University of Rhode Island
Dunn, Mary; University of Texas
Dunn, Phillip C.; University of South Carolina
Embley, David W.; Brigham Young University
Erdogan, Sevki; University of Hawaii
Eskicioglu, Ahmet; CUNY Brooklyn College
Freedman, Immanuel
Gefen, David; Drexel University
Goldstone, Jack A.; George Mason University
Han, Ju; Lawrence Berkeley National Lab
Hansen, Bill;
Hasselmo, Michael Erik; Boston University
Hunt, Anthony; University of California
Kitts, Christopher; Santa Clara University
Lazarevic, Aleksandar; United Tech. Research Center
Lee, Chin C.; University of California
Lee, James; The George Washington University
Lin, Zhibin; University of Virginia
Lindner, James R.; Texas A&M University
Lorenzi, Nancy M.; Vanderbilt University
Loyall, Joseph P.; BBN Technologies
Luo, Xin; Mississippi State University
MacLachlan, Douglas L.; University of Washington
Maher, Julie; Oregon Department of Human Services
Mazumdar, Subhasish; New Mexico Tech
Melody, Alexander W.; Ball State University
Matthews, Harry R.; University of California
McKenney, Paul E.; IBM Linux Technology Center
Middleton, Blackford; Harvard Medical School
Miller, Robin Lin; Michigan State University
Moore, Michael Grahame; Pennsylvania University
Morell, Jonathan A.; The Altarum Institute
Motter, Adilson E.; Northwestern University
Neuberger, John W.; University of North Texas
Newman, Peter ; Colorado State University
Niemi, Richard; University of Rochester
Nolle, Daniel E.; Comptroller of the Currency
North, Michael J.; Argonne National Laboratory
Osterman, Karen F.; Hofstra University
Othayoth, Raghunath; Hewlett-Packard Company
Ott, Margaret G.; Tyler Junior College
Peha, Jon M.; Carnegie Mellon University
Plass, Jan L.; New York University
Podichetty, Vinod; Cleveland Clinic Florida
Poltrock, Steven; The Boeing Company
Prabhakar, Salil; Digital Persona Inc.
Prato, Tony; University of Missouri-Columbia
Raje, Rajeev; Indiana University
Reeve, Trevor; Federal Reserve System
Reynard, Ruth; Trevecca Nazarene University
Roiger, Richard J.; Minnesota State University,
Roozendaal, Benno; University of California
Rosen, Lauren; University of Wisconsin
Salovey, Peter; Yale University
Sands, Daniel Z.; Harvard Medical School
Schott, Marshall; University of Houston
Schwabacher, Mark; NASA Ames Research Center
Sears, Andrew; UMBC
Shah, Biren N.; University of Louisiana
Sittig, Dean F.; Kaiser Permanente, Northwest
Smith, Daniel T.; Purdue University
Takayama, Leila; Stanford University
Tang, Yuchun; Georgia State University
Tarjan, Robert E.; Princeton University
Thompson, Dale R.; University of Arkansas
Troutt, Marvin D.; Kent State University
Usher, John M.; Mississippi State University
Wagstaff, Samuel S.; Purdue University
Wang, Ke; Columbia University
Wang, Xinwei; The University of Nebraska-Lincoln
Weeks, Michael; Georgia State University
Xie, Qing; University of Maryland

IPSI Team

Advisor for Mathematics-Related Problems:
Jelena Miletic and Maja Etinski
Advisors for IPSI Tutorials:
Tanja Kovacevic and Tanja Petrovic
Advisors for IPSI Research:
Milos Kovacevic and Milos Milovanovic
Advisors for IPSI Developments:
Zoran Babovic and Darko Jovic

Welcome to IPSI BgD Conferences and Journals!

http://www.internetconferences.net
http://www.internetjournals.net

IPSI-2006 MONTREAL
Montreal, Quebec, Canada
June 30 to July 3, 2006
With Visits to Jazz and Laughs Festivals

IPS-USA-2006 NEW YORK
Harvard Club of New York,
27 West 44 Street,
New York, USA
July 3 to 6, 2006

IPS-USA-2006 BOSTON
Hotel@MIT, Cambridge, Boston
Massachusetts, USA
July 6 to 9, 2006

IPSI-2006 ITALY
Rome and Hotel Castello Chiola,
Loreto Aprutino in Abruzzo,
relatively near Rome, Italy
August 19 to 26, 2006

IPSI-2006 BELGRADE
Belgrade, Serbia
August 28 to 31, 2006
With Visits to Famous Serbian Monasteries

IPSI-2006 LONDON
Hotel Beaufort, Knightsbridge,
London, England
Aug 31 to Sep 3, 2006

IPSI-2006 MONTENEGRO
Hotel Sveti Stefan, Montenegro
Sep 23 to Sep 30, 2006

IPSI-2006 VENICE
Venice, Italy
October 12 to 15, 2006

IPSI-2006 SLOVENIA
Hotel Toplice,
Bled Slovenia
Nov 30 to Dec 3, 2006

IPS-USA-2007 HAWAII
Hilton Waikoloa
Big Island, Hawaii, USA
January 6 to 9, 2007

IPS-USA-2007 CALIFORNIA
Palo Alto in Silicon Valley,
near San Francisco,
Hotel Stanford Terrace,
California, USA
January 9 to 12, 2007

IPSI-2007 MARBELLA
Hotel Puente Romano
Marbella, Spain
February 22 to 25, 2007

IPSI-2007 AMALFI
Hotel Santa Caterina
Amalfi, Italy
March 22 to 25, 2007

IPSI-2007 LAPLAND
Rovaniemi,
Santa Claus World Capital,
Finland
March 29 to April 1, 2007

IPSI-2007 FRANCE
Hotel de la Cite,
Carcassonne, France,
UNESCO Heritage,
April 26 to 29, 2007

IPSI-2007 SPAIN
Hostal de la Gavina,
S'Agaro, Costa Brava
near Gaudi Barcelona, Spain
May 3 to 6, 2007

IPSI-2007 TOKYO
M.I.T., Tokyo, Japan
May 31 to June 3, 2007

IPS-USA-2007 NEW YORK
Harvard Club of New York,
27 West 44 Street,
New York, USA
July 2 to 5, 2007

IPS-USA-2007 BOSTON
Hotel@MIT, Cambridge, Boston
Massachusetts, USA
July 5 to 8, 2007

IPSI-2007 ITALY
Rome and Hotel Castello Chiola,
Loreto Aprutino in Abruzzo,
relatively near Rome, Italy
August 18 to 25, 2007

IPSI-2007 FIRENCE
Firence, Italy
Aug 25 to Aug 28, 2007

IPSI-2007 BELGRADE
Belgrade, Serbia
September 19 to 22, 2007
With Visits to Famous Serbian Monasteries

IPSI-2007 MONTENEGRO
Hotel Sveti Stefan
Sep 22 to Sep 29, 2007

IPSI-2007 VENICE
Venice, Italy
October 11 to 14, 2007

IPSI-2007 SLOVENIA
Lake Bled in Alps,
Slovenia
Nov 29 to Dec 2, 2007

The publication of this journal
is supported in part by
the Ministry of Science of Serbia

CIP – Katalogizacija u publikaciji Narodna biblioteka Srbije, Beograd

ISSN 1820 – 4503 =
The IPSI BGD Transactions on Internet Research
COBISS.SR - ID 119127052