

The IPSI BgD Transactions on Advanced Research

Issues in Computer Science and Engineering

A publication of
IPSI Bgd Internet Research Society
New York, Frankfurt, Tokyo, Belgrade
January 2005 Volume 1 Number 1 (ISSN 1820 - 4511)

Special Issue - Selected best papers from the IPSI-BgD conferences in 2004 and before

Guest Editor's Introduction:	Wisdom of Education for Globalization	
	Milutinovic, Veljko.....	3

Table of Contents:

Issues in Wireless Security Based on AES Hardware Implementation	
Bojanic, S. Carreras, C., Díez, J. M., and Nieto-Taladriz, O.	6
Programming Language Concepts for Global Computing	
Ferreira, U.	11
Information Extraction from Texts: Adapting a System for Summarization of News Reports to the Domain of Bioinformatics	
Gawronska, B.	20
easyTransport: An Interoperable and Secure e-ticketing Model based on Contactless Smart Cards	
Gómez Skarmeta, A.F., Martínez Pérez, G., and Yago Sánchez, C. M.	29
Adaptive Collaboration: The Road Map Leading Telework to a More Advanced and Professional Working Format	
Hori, M. and Ohashi, M.	36
Molecular Processes as a Basis for Autonomous Networking	
Krüger, B. and Dressler, F.	43
H1B Visa and the Policy Vignette for 21st Century Dislocated Workers in the United States	
O'Lawrence, H.	51
Factors affecting supplier adoption of governmental reverse auctions: an exploratory case study	
Radhakrishnan, A., Davis, S., and Davis, B. R.	57
Quantifying the Learning Efficiency of Programming Learning and Its Implications	
Shyu, Y.	66
Agile Systems for Clinical Research	
Wade, T. D. and Murphy, J.R.	74
Reasoning Procedure and Implementation for Logic Programs as Managing Schemes to Extract Demand	
Yamasaki, S., Iwata, K., and Sasakura, M.	83

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 €20 plus an annual subscription fee of €100 (air mail printed matters delivery).

Member copies of Transactions are for personal use only

IPSI BGD TRANSACTIONS ON ADVANCED RESEARCH

www.internetjournals.net

STAFF

Veljko Milutinovic, Editor-in-Chief	Sanida Omerovic, Journal Manager
Department of Computer Engineering University of Belgrade POB 35-54 Belgrade, Serbia (381) 64-1389281 (tel)	IPSI BgD Internet Research Society
vm@etf.bg.ac.yu	sanida@internetjournals.net

EDITORIAL BOARD

Michael Flynn	Dawn Jutla	Hojjat Adeli
Stanford University, Palo Alto, California	Sant Marry's University, Halifax,	Ohio State University, Ohio,
USA	Canada	USA
Dino Karabeg	Borko Furht	Massimo Piccardi
Oslo University, Oslo,	Florida Atlantic University, Boka Raton, Florida,	Sidney University of Technology, Sidney,
Norway	USA	Australia
Andrea Domenici	Victor Gonzalez	Jean-Luc Ganascia
University of Pisa, Pisa,	University of Oviedo, Gijon,	Paris University, Paris,
Italy	Spain	France
Tan Kok Kiong	Ryuichi Takahashi	Bob Crisp
National University of Singapore,	Hiroshima University, Hiroshima,	University of Arkansas, Fayetteville, Arkansas,
Singapore	Japan	USA
Estela Blaisten-Barojas	Erich Neuhold	Hironori Fujii
George Mason University, Fairfax, Virginia,	Fraunhofer IPSI, Darmstadt	Fujii Labs, M.I.T., Tokyo,
USA	Germany	Japan
Charles Milligan	Veljko Milutinovic	Chip Rutledge
StorageTek, Colorado,	IPSI, Belgrade	Purdue Discovery Park, Indiana,
USA	Serbia	USA

Wisdom of Education for Globalization

Veljko Milutinovic

1. Essential Issues

As a university professor, this is what I tell to my students when they ask (about their possible success strategies and tactics in the era of globalization):

Index Terms—*compatibility, globalization, intelligence dimensions.*

1.1

When you grow up your own children, set up the priorities as follows:

a. Ethical Strength: If a word is given, it should be kept at any cost, even if conditions in the environment change, and keeping the word becomes difficult, or costly, or against our own interests.

b. Multidimensional Personality Strength: First, psychological strength, by having them grow in what appears to them as a harmonic family. Second, physical strength, by running them through an exhaustive athletic program. Third, socio strength, by having them constantly interact with the surrounding in real-problem situations.

c. Multidimensional Academic Strength: First, school (formal education). Second, languages (English is the key to businesses, and other languages are keys to hearts). Third, hobby (preparation for old age); something to keep them busy after retirement – the earlier it is developed - the better it is.

Many would react fiercely to the advocating that academic strength should be the third, rather than the first priority. However, I strongly believe that, after the

critical age is over (teens), things turn upside down, and ethics form the fundamentals, while formal education becomes the mechanism to reach the skies of professional success.

1.2

After the formal education is over (for some, this is Ph.D., for others, this is B.Sc.), it is a must that the young person goes away from its native environment, because the talents can be fully developed only outside the nest of the native culture. However, after the "talent development phase" is over, the young person should return back to the native environment (if conditions permit), since that is where the life is compatible with personality needs (global success can still be achieved, via the Internet). Moving the roots (from one culture to the other) is painful, and could be tragic in the last part of the life (it is tough to be an old person anywhere, but it is the toughest among the people different from you).

1.3

The "going away" helps develop three dimensions of human intelligence:

a. Vertical Dimension: Deep professional knowledge.

b. Horizontal Dimension: Knowledge about other scientific fields, other cultures, other languages, etc. This dimension is important for creativity. Fachidiots are never creative.

c. Perpendicular Dimension: Look-ahead, or ability to anticipate (problems that may create loss or opportunities that may create benefit), get prepared for what is to come, and minimize the loss (if problems) or maximize the benefit (if opportunities). This dimension is especially important for financial success (if one anticipates now what will be a good offer/demand scenario after a certain time interval, and uses that

Manuscript received November 2, 2004.

Veljko Milutinovic is with the School of Electrical Engineering, Serbia and Montenegro (email:vm@etf.bg.ac.yu).

time interval to become an expert far what is to be in demand a lot, and doable by few, the market value of the expert knowledge may become so high to bring a financial fortune).

The third dimension is the most difficult one to build; takes talent, travel, and time, and very few achieve it before their strength is over.

1.4

Do not take standard routes! That approach brings incremental benefits. Nonstandard routes, if properly selected, and if the strength is not lacking, may bring extraordinary benefits. Do not teach your child to become a good fisherman (who brings fish for dinner each evening); teach it to become a good sailor (who brings nothing for dinner each evening, but becomes rich in some uncomprehensible way).

1.5

Pay attention to learning, but more to forgetting. Human knowledge is "decayable goods/tissue" and the garbage has to be disposed periodically; otherwise, it piles up in one's brain, and disables one from creating optimal decisions. This applies both to the life wisdom and the professional wisdom. Stay away from "success models" that your parents impose on you by copying from the times of their "full gears." Also, remember, when the technology changes, optimal solutions change as well (e.g., in a Silicon model, carry-lookahead adder is faster than the ripple-carry adder for all except the lowest word lengths; for a GaAs model, ripple-carry adder may be faster for word lengths as large as 16 or 32 bits).

1.6

When you are to form a team, pass the interested candidates first through the filter of ethics, second through the filter of personal strength (the major three dimensions elaborated earlier in this text), and apply the filter of academics criteria only at the very end, i.e., only on those who have passed successfully through the first two filters (ethics and personality). The worst damage to the business is made by intelligent but unethical and of a deviated personality (e.g., socially unadjusted).

Only stupid but well-intended and hard-working can do worse than that.

1.7

When the time comes to "sell" the results of your work, present the results (by using an appropriate output processor) in the

way that helps pass the essence as efficiently as possible (by doing it compatibly with the input processor of those to absorb). Compatibility has three major forms:

a. **Structural Compatibility:** Each presentation must have the following 10 points: (1) Introduction, to tune the audience for faster comprehension of what follows; (2) Problem statement, which also tells why is the problem important; (3) About the existing solutions, and what are their drawbacks; (4) About the proposed solution, and why is it not supposed to have the same drawbacks; (5) Conditions and assumptions of the analysis to follow; (6) Details of the solutions to be compared - proposed versus existing; (7) Analytical modeling, to show the essence; (8) Simulation analysis, to show performance or speed, (9) Implementation analysis, to show complexity or price; (10) Conclusion from the performance/complexity point of view.

b. **Semantic Compatibility:** Each presentation should use the symbols compatible with the semantics of the subject matter. One picture is worth 1000 words, a done MPEG "picture" is worth 1000 JPEG "pictures".

c. **Syntactical Compatibility:** Each presentation must have the form that matches the essence; for example, if a bullet of a PowerPoint presentation spreads over two or more lines, each line must be a different thought, with line boundaries and thought boundaries at the same places.

* * * * *

At all times, keep in mind that one possible system of human knowledge implies the following seven layers:

- a. *Philosophy*
- b. *Logic*
- c. *Mathematics*
- d. *Sciences*
- e. *Engineering*
- f. *Commonsense*

Religion is where we put the un-comprehensible.

I also tell the students that the higher is the layer they reach in their professional lives, the lower is the level they reach on their bank accounts.

2. PRACTICAL ADVICES:

2.1

Make sure that at birth you select parents who place the children education on the top of the list of their priorities. A good alternative is no parents at all (if you survive, you have good chances to be more successful than anybody from the first group).

2.2

Make sure that you understand that it is much less important what is written about you (what people read in your CV), and much more important what is being told about you (what people talk about you, and how people view you as a complete person, with many of the components not found in typical CV).

2.3

Make sure that you know how to learn from your students.

2.4

Make sure, when things go wrong, that your attitude is "who knows why is this good for me," and when things go superb, that your attitude is "I know that this will not last for ever!"

2.5

Make sure that you have a good balance of work, sports, and hedonism; otherwise, your success does not make much sense. Too much work can ruin both the physical and the mental health.

2.6

Make sure that you recognize the moment when you have to be patient!

2.7

Make sure that you have a good source of inspiration at all times!

3. CONCLUSION

I usually wait for 30 years before I ask my students about the feedback. Any comments or additions to all of the above are welcome!



Veljko Milutinovic is known as the co-architect/co-designer of the World's first 200MHz microprocessor, for DARPA, back in 80s, a decade before Intel. In 80s he taught at Purdue University, USA. Since 1.1.1990. he teaches at the University of Belgrade, Serbia. He is a Fellow of the IEEE.

Issues in Wireless Security Based on AES Hardware Implementation

Slobodan Bojanic, Carlos Carreras, Juan M. Díez, and Octavio Nieto-Taladriz,

Abstract— While short-term IEEE 802.11 wireless security solution (TKIP) accommodates existing hardware, the long-term security solution called CCMP is targeted at new hardware designs. This paper is related to FPGA/ASIC hardware architectural options in implementation of required AES crypto algorithm. We are addressing different design criteria like high-speed, low-cost, or modes of operation like CCM. These issues are also of interest for wireless sensor networks whose secure low-power operations are referred by another IEEE standard (802.15.4).

Index Terms—AES, FPGA, security, wireless

1. INTRODUCTION

IEEE standard 802.11-1999 has the ability to handle wireless traffic quite easily, but its WEP (wireless equivalent protection) security scheme has been flawed. Consequently IEEE 802.11 Task Group i (TGi) offered the security solution in two phases: short-term security solution called Temporal Key Integrity Protocol (TKIP) and long-term security solution called Counter Mode with CBC-MAC Protocol (CCMP). TKIP accommodates existing hardware and fixes all known WEP vulnerabilities, but provides only minimal security on deployed equipment and degrades performance [1].

CCM is a generic authenticate-and-encrypt block cipher mode. It provides authenticated encryption combining Counter

(CTR) mode encryption and CBC-MAC authentication using a single key. It assumes 128 bit block cipher – IEEE 802.11i uses AES. It is intended for packet environments with no attempt to accommodate streams. This method proves to be viable for several reasons; there is no known patent encumbrances, the modes have been used and studied for a long time and have well-understood cryptographic properties, and they provide good security and performance, whether implemented in hardware or software [2].

Since most 802.11i implementations of AES will be in hardware [3], we intend in this paper to consider different aspects of AES hardware design. Therefore, the rest of the paper is organized as follows: in Section 2 we introduce the CCM protocol and its implications to hardware design; in Section 3 the high performance AES implementation is presented while in Section 4 low cost and in Section 5 low power consumption aspects of the AES implementation are treated. Conclusions are drawn in Section 6.

2. AES-CCM MODE

In CCM mode, the majority of the time of the protocol is spent on computing the AES algorithm. AES is used to generate the cipher text from the header of the 802.11 package as well as the package payload. Therefore, it is incentive to have a hardware assisted ASIC or FPGA for computing the AES cipher text. This will alleviate the computing power from the main processor. One possible implementation of the CCMP protocol is to have the main processor in charge of the MAC layer and ASIC/FPGA device running simultaneously performing the AES encryption/decryption algorithm.

The Advanced Encryption Standard (AES)

Manuscript received July 14, 2004. This work was supported in part by the Spain Ministry of Science and Technology under Grant TIC2003-09061-C03-02 and the “Ramón y Cajal” program.

The authors are with the Electronic Engineering Department, Technical University of Madrid, Spain (e-mails: {slobodan, carreras, jmdiez, nieto}@die.upm.es). The contact person is S. Bojanic.

is likely to become a *de facto* worldwide encryption standard commonly used to protect all means of secret communications during the next several decades. AES (Rijndael) is a symmetric block cipher with a variable key size (128, 192 and 256 bits) and variable input/output block size where only a 128-bits block size is required by the AES Standard. Rijndael is a substitution-linear transformation cipher based on S-boxes and operations in the Galois fields [4].

The implementation of the encryption round of Rijndael requires the realization of four component operations: ByteSub, ShiftRow, MixColumn, and AddRoundKey, while the implementation of the decryption round requires their inverse operations.

ByteSub includes sixteen identical 8x8 S-boxes working in parallel. Each of these S-boxes can be implemented independently using a 256 x 8-bit look-up table. ShiftRow changes the order of bytes within a 16-byte (128-bit) word. This transformation involves only changing the order of signals and, therefore, it can be implemented through routing. The MixColumn transformation can be expressed as a matrix multiplication in the Galois field $GF(2^8)$. AddRoundKey is a bitwise XOR of two 128-bit words and can be implemented using one layer of 128 look-up tables, which translates to FPGA implementation of 64 CLB slices.

The goal of the FPGA is to compute the AES algorithm to generate the cipher text in the range of at least 11Mbps or higher. However, since the 802.11a emergence, it is desirable to have the ASIC/FPGA to run at 54 Mbps. Since CCM spends a lot of computational time on AES decryption/encryption, implementing the AES off chip on a separate FPGA would free the main processor.

Network protocols, particularly IEEE Std. 802.11i, are composed of several different layers. When a user wants to send information over the network, the information is separated into smaller individual packets. These packets are then sent down through the layers and additional information is added to them to ensure proper communication, such as source address and

destination address. The packet is then sent through the Ethernet and arrives at the destination where it travels up the layers and is reassembled for the receiver.

One of the layers that packets must go through in 802.11i, is the MAC layer. It is responsible for packing the packet and ensuring proper security. The security scheme that is being proposed is the CCM Mode Encryption that relies heavily on software block ciphers such as AES.

By including the use of a hardware block cipher, the processor can spend some of the computational time on other computations. The ideal goal would be to have this AES hardware implementation embedded on a Wireless Network Adapters for PCs and Laptops. Ultimately this could relieve enough time from the processor to enable faster communication rates and generally both speed up and secure wireless communications.

The AES implementation on the FPGA is a viable solution for improving the speed and processing power of CCM Mode Encryption. The implementation on a Xilinx Spartan II [5] that is partitioned into three modules (input interface, output interface and the AES block cipher engine) shows that a much larger FPGA or ASIC would be preferred, since both encryption and decryption could be implemented as well as some pipelining of processes.

3. HIGH PERFORMANCE IMPLEMENTATION

Most of the AES implementations have been oriented to achieve high performances [6], [7]. In [6] the AES implementation exploits the benefits of the development and application of a pipeline compiler. Several implementations of AES ciphers with keys of 128 bits with sub-key generation in parallel with data processing, have been prepared, analyzed and refined.

Figure 1 shows an example of the pipeline compilation process of a combinational block. The conversion flow starts with a fully unrolled version of the algorithm sequence to obtain a data-flow graph. This graph is built considering the time delays of every operator. The next step is to divide the graph in stages by register banks. The compiler decides the positions of these banks and inserts automatically the registers needed.

The AES algorithm can be implemented with three basic operations: XOR of two bit inputs, a tree of XORs which implements the *mixcolumns* function, and a substitution of bytes. The substitution boxes, SBOXs, are tables of 256 bytes whose

The Xilinx devices offer two alternatives to implement memory blocks: distributed memories and RamBlocks. The later option obtains better results for large memories. RamBlocks are specific memory elements of 18Kbits and can be used like single/dual port RAMs or ROMs. The Xilinx tool COREGEN generates directly this block from a bit map supplied by the designer. Figure 2 shows the VirtexII-4000 structure with 6 columns of 20 blocks, the selected manual placement and the flow of data. Every round uses 10 consecutive blocks of one column.

An additional optimization involving the connection of all registers to the reset of the system was also required. The reason for this is an optimization performed by the synthesis tool. The reset connections allow an increment of 15 MHz in the clock frequency, leading to a final frequency of 167.67 MHz and a throughput of 21.5 Gbps. The implementation uses 388 IOBs, 9,357 slices, 100 RAM blocks and 6,734,115 equivalent gates.

A pipelined architecture of AES 128-key without key scheduling is presented in [6]. This design uses 9196 slices and 80 RamBlocks providing a throughput of 16 Gbps versus the 21.5 Gbps obtained here, on a Xilinx Virtex v1000 device. The AES implementation offers excellent results. It proves the possibilities of using a pipeline compiler driving architectural design, allowing fast design exploration and supporting the analysis of several alternatives and refinements in reduced time.

4. LOW COST IMPLEMENTATION

Much of the research targets state-of-the-art technologies where the individual cost of those devices ranges in hundreds of US dollars. These implementations feature high speeds and high costs suitable for high-end applications only [8]. But the need for secure electronic data exchange will become increasingly more important to low-end customer products like wireless devices, thus the AES implementations must become very inexpensive.

Most of the low-end applications do not require high encryption speeds. Current

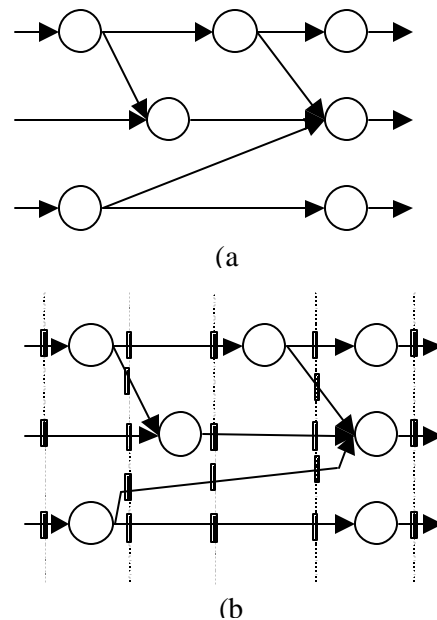


Figure 1: Example of conversion of a combinational graph (a) into a pipeline architecture (b)

wireless networks achieve speeds up to 60 Mbps. Implementing security protocols, even for those low network speeds, significantly increases the requirements for computational power. For example, the processing power requirements for AES encryption at the speed of 10 Mbps are at the level of 206,3 MIPS [9]. In contrast, a state-of-the-art handset processor is capable of delivering approximately 150 MIPS at 133 MHz, and 235 MIPS at 206 MHz.

Early AES designs were mostly straightforward implementations of various loop unrolled and pipelined architectures with limited number of architectural optimisations, which resulted in poor resource utilization. For example, AES 8x8 S-boxes were implemented on LUTs as huge tables left for synthesizers to optimise. Later FPGA implementations demonstrated better utilization of FPGA resources. Several architecture using dedicated on-chip memories implementing S-boxes and T-boxes were developed. Recent research has focused on fast pipelined implementations but most of them are too costly for practical applications [13].

There are few compact implementations of the AES algorithm in FPGAs. There exist

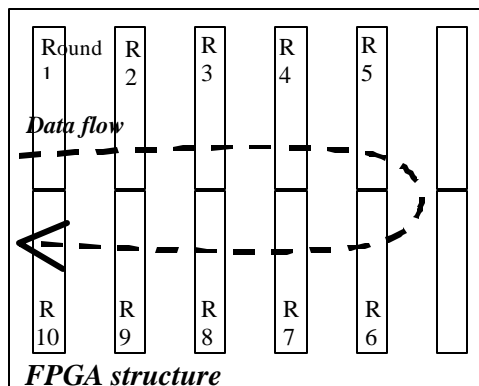


Figure 2: Manual placement of RAM-Blocks

commercial compact cores from Amhion [10] and Helion [11] companies. Both companies provide compact cores in encryption or decryption version only, and a 128-bit key schedule.

An analytical approach to compact the AES implementation resulted in the proposal of an AES S-box implementation based on composite fields [12]. Another approach to create a low-cost implementation of AES in the FPGA targeted for real life applications is to shift attention to older technologies and smaller devices. The implementation [13] fits in an inexpensive, off-the-shelf Xilinx Spartan II XC2S30 FPGA, with a cost that starts below \$10 per unit. Only 50% of the logic resources available in this device were utilized, leaving enough area for additional logic. This implementation can encrypt and decrypt data streams up to 166 Mbps. The

encrypted speed, functionality, and cost make this solution perfectly practical in the world of embedded systems and wireless communication.

5. LOW POWER CONSUMPTION

The IEEE 802.15.4 Low-Rate Wireless Personal Area Network (LR-WPAN) standard [14] also requires the use of the AES algorithm in CCM mode for wireless sensor networks where low power consumption is the priority. The networks should operate for many months from relatively small batteries, in sensing applications that range from industrial control and monitoring, security, and home automation to those found in health care, telematics, and intelligent agriculture. To attain low-power operation, it is necessary to minimize the device's power consumption while it is active for both the sensor (or actuator) and the communication transceiver. The sensor use is application specific, and its consumption can exceed that of the transceiver. Also the power consumed during standby mode can easily become the dominant component in the average power consumption calculation—becoming more significant than the power consumed during active operation.

The transmission of the unsecured packet with a first-generation transceiver requires 56 μJ assuming that the data payload is 16 bytes, that short (logical) addresses are used in the medium access control (MAC) header, and that the 2.4 GHz band is employed (which has a data rate of 250 Kbps) [15]. When CCM is applied, 13 additional bytes must be transmitted in the packet (8 for the MIC, 4 for the frame counter field, and 1 for the key sequence counter), thus transmission of the secured packet requires an additional 21 μJ .

If the AES security calculations are performed in software by a microcontroller, e.g., the Motorola MC9S08GB60, from its data sheet [16], the energy needed to perform the calculations is then: 35 μJ . So 56 μJ are needed to transmit this packet in an unsecured mode, while an additional 21 + 35 = 56 μJ are needed to transmit this packet in the CCM secure mode, using software to encrypt it.

Doubling the energy requirement to incorporate security may seem onerous, but recall that encryption functions are only performed when packets are transmitted or received and that IEEE 802.15.4 is by definition a low-data-rate network with relatively low throughput. Security functions are therefore rarely performed. As noted earlier, in these systems standby power consumption often dominates active power consumption in the battery life calculation.

Alternatively, a dedicated AES engine may be implemented in hardware to reduce power consumption further. The performance parameters of such a device are highly implementation specific; however, we can still make rough conservative estimates [15]. If we assume the engine draws 1 mA from 2 V and requires 100 cycles at 16 MHz to produce an output, the security calculation will require 87.5 nJ. This is significantly less power than that required by the software approach, but it comes at the cost of a dedicated circuit that must be purchased and there is still the need of 21 μ J to transmit the additional 13 bytes in the secure packet.

6. CONCLUSION

As 802.11i becomes a standard in the near future, it will replace the current 802.11 WEP security scheme and the AES algorithm in CCM mode will be the basis for data protection across the wireless medium. Since most 802.11i implementations of AES will be in hardware, there is strong need of thorough examination of each of the components of the AES algorithm for a perfect match into the architecture of the FPGA or ASIC. This paper considers the issues of high performance, low cost and low power design issues in the AES implementation and demonstrates obtained results as well as further research directions.

REFERENCES

- [1] R. Housley, Panel Discussion on the Differences and Similarities of Wired vs. Wireless Security, OpenSig 2003, 9 October 2003.
- [2] D. Whiting, R. Housley, and N. Ferguson: IEEE P802.11 wireless LANs: AES encryption & authentication using CTR mode & CBC-MAC. IEEE Tech. Rep. IEEE 802.11-02/001r2, May 2002.
- [3] N. Ferguson: AES Mode Choices: OCB vs. Counter Mode with CBC-MAC. IEEE 802.11-01/634r0, November 2001.
- [4] U.S. Dept. of Commerce, National Institute of Standards and Technology, Information Technology Laboratory, *Specification for the Advanced Encryption Standard (AES)*. Federal Information Processing Standard Publication (FIPS PUB) 197. Springfield, VA: National Technical Information Service, 26 Nov. 2001.
- [5] K. Vu, D. Zier: FPGA Implementation AES for CCM mode encryption using Xilinx Spartan II. ECE-679, Oregon State University, Spring 2003.
- [6] J. M. Díez, S. Bojanic, C. Carreras, O. Nieto-Taladriz: FPGA Implementation of Three IPsec Cryptographic Algorithms. WSEAS Trans. on Systems, Issue 1, Vol. 2, p. 229-234, 2003.
- [7] A. Hodjat and I. Verbauwhede: A 21.54 Gbits/s fully pipelined AES processor on FPGA, IEEE Symposium on Field-Programmable Custom Computing Machines, April 2004.
- [8] P. Chodowiec, K. Gaj, P. Bellows, and B. Schott, Experimental Testing of the Gigabit IPsec-Compliant Implementations of Rijndael and Triple DES Using SLAAC-1V FPGA Accelerator Board, Proc. Information Security Conference, Malaga, Spain, October 2001.
- [9] Ravi S., Raghunathan A., Potlapally N.: Securing Wireless Data: System Architecture Challenges, Symposium on System Synthesis, 2002.
- [10] Amphion: <http://www.amphion.com>.
- [11] Helion: <http://www.helion.com>.
- [12] Rijmen V.: Efficient implementation of the Rijndael S-box, available at: <http://www.esat.kuleuven.ac.be/rijmen/rijndael/sbox.pdf>.
- [13] P. Chodowiec, K. Gaj: Very compact FPGA implementation of the AES algorithm. LNCS 2779, pp. 319-333, 2003.
- [14] Institute of Electrical and Electronics Engineers, Inc., IEEE Std. 802.15.4-2003, New York: IEEE Press, 2003.
- [15] E. Callaway: Secure Low-Power Operation of Wireless Sensor Networks, Sensors, <http://www.sensormag.com/articles/0104/22/main.shtml>, January 2004.
- [16] Data sheet for Motorola MC9S08GB60 microcontroller.

Programming Language Concepts for Global Computing

Ulisses Ferreira

Abstract — *The present paper introduces a number of existing concepts about the programming language Plain that are useful for mobile agents and for the Internet programming, among other applications. This paper explores the ability to program with the unknown value, by preventing samples.*

Index Terms — *global computer, internet, programming language*

1. INTRODUCTION

For writing the present paper, the present author observed that, in order to provide a reasonable paradigm for general Internet programming, some combined and non-traditional programming language concepts need to be introduced aiming at specific applications over the Internet.

Code mobility is a relatively new field of research that has inspired intriguing ideas on programming techniques to improve related software[8]. New programming languages have been presented and discussed in workshops and conferences aiming at providing better standards and good examples for future language designs, commercially or otherwise.

In comparison with the state-of-the-art technology, to propose a better paradigm for programming mobile agents for the World Wide Web, for example, the language designer should consider that the underlying connections often fail or delay. A neutral state, which would represent the lack of result, could be assigned to the variable that takes part in the request in such a way that the program carries on running safely. Mobile agents need to be robust and make their own decisions remotely. Two very different and alternative approaches to agent are [25], [33].

Although the present programming language features are not necessarily specific for mobile code languages, the underlying environment can be the same. Here the present author explains how this additional value can be useful in programming on a global network[35] where the mobile agent paradigm and technology have become increasingly important. In logics, this value has been traditionally referred to as *uu* for providing an alternative value to *true (tt)* and *false (ff)*. Like many imperative features, the present ones also apply to functional languages.

Broadly, some pieces of work, such as [30], [31], have indicated similarities between technologies of code mobility and persistence, and some persistent languages are being explored at some universities. In the present approach, because the present author is looking for generality and efficiency, persistence is not provided directly by the language, but instead by programmable constructs. Because of that, this approach here is at least as applicable to mobile agents as persistent languages. As well as persistence, communication is another very active topic of research and much work has been done regarding fault tolerance and communication between mobile agents. However, relatively few satisfactory results have been achieved in terms of language facilities and abstractions.

Here we concentrate on the use of programming features in the context of a global environment and mobile agents programming.

The ability to represent and reason with partial information is well understood in the artificial intelligence and logic communities. However, very little of this work has been related to programming techniques. An exception is Extended Logic Programming, introduced by Gelfond and Lifschitz[16], [17], that can be used for the same purpose as that the present author is discussing here. Extended Logic Programming makes use of two forms of negation. In [23], the author

Manuscript received June 3, 2004. This piece of work corresponds to a chapter of the present author's Ph.D. Thesis (2001), at the University of Dublin, Dublin 2, Ireland. Updated author's e-mail: josefer@philosophers-fcs.org.

suggests that an important practical problem in Extended Logic Programming is how the programmer distinguishes whether a negative condition is to be interpreted as explicit negation, or as negation due to the absence of any clause in any closed world as an assumption.

The *unknown* value, *uu*, extends the semantics of other logics, such as classical and intuitionistic logic, according to the Lukasiewicz [38], [26] 3valued logic. Here the present author extends this value for each data type in programming for a global environment or for mobile agents. In arithmetic and relational expressions, *uu* as a resulting operand implies the] expression to result in *uu*. Accordingly, statements have to be adapted to make use of this value.

Most Expert System Shells made use of an *unknown* symbol to represent lack of information in Boolean variables, in a very restrictive way however. The present author generalizes this concept to programming languages in general, although he applies it here to mobile agents programming.

Agents have to be robust and, because of this, when connections fail or delay, programs should carry on running despite the lack of information. *uu* is a constant in programming languages that can be assigned to any variable of any datatype. This new constant guarantees both safety and robustness at the same time, because variables are never committed to any value that is not in the problem domain. An introduction on types for functional programming languages can be found in [34].

In section 2 the present author reviews some recent programming languages for such an environment, while section 3 is dedicated to an illustrative programming language. Section 4 introduces the concept of *unknown* together with other related concepts, and explains how they can be used to achieve the proposed goals. As a consequence, section 5 complements those concepts by stressing the importance of any form of lazy evaluation in programming, as well as timeouts, no matter the adopted paradigm. Section 6 contains other relevant features that are relevant enough to be mentioned. Section 7 contains the conclusion.

The examples in this paper are written in a programming language. In section 3, the present author discusses the syntax of the relevant subset of this language.

2. SOME CURRENT MOBILE CODE LANGUAGES

In the past few years researchers have seen the Internet as a popular environment for systems. Some of us would like to program and compute using this structure, i.e. to view parts of an Internet-like network as *global computers* [15]. Many companies, for example, are starting to have their own internal global computers for their specific purposes.

One of the most interesting ideas is not only to move code, as **Tcl**[32] and **Java**[3] do, but also computation (code along with context) over the network, that is, a computation which starts at some location may continue to execute at some other location. Synchronous connections to the original site may be set while a program is running remotely in such a way that any change in some variables transparently causes the value to be stored in the original site. Alternatively, new values of variables can be sent to the original site with no need for synchronous connections. Other paradigms of mobile computation already exist and they depend on the kind of entities that are transferred over the network, with respect to what is moved (code, data, connections, etc).

When the code is moved, what happens if the names it contains are bound to resources in the source virtual machine? This issue defines two classes of strategy, *replication* and *sharing*. The former strategy may be either static or dynamic. Concerning static replication strategy, constants, system variables and libraries, for example, are regarded as *ubiquitous resources*[27] and they can adopt such strategy, where bindings are deleted and set after arrival. As for dynamic replication strategy, the code migrates to another virtual machine along with bound resources and the original bindings are deleted. The original resource in the source virtual machine may be either deleted (*replication by move*) or kept (*replication by copy*). In the *sharing strategy*, the original resource is kept and remotely accessed through network references and connections.

In both strong and weak mobility[10], security[39],[40] is a very important matter. Locations must check for authorization and capabilities in order to prevent malicious software running. However, as long as that

is ensured by the system, a global network can be a very interesting and natural platform for computation. Thus, a new challenge emerges: how to provide these facilities and prevent the related problems?

Mobility should also be considered not only during the execution of programs but also during the elaboration of software. The emphasis in performance is no longer in the run-time code generated by compilers, but in the (dynamic) compilation process itself (when applicable), transmission and additional overheads to guarantee security and other requirements.

Acharya, Ranganathan and Saltz[1],[2] during the design of **Sumatra**, an extension of **Java**, consider three requirements for **individuals**: *awareness*, which is the need to monitor the level and quality of resources in their operating environment; *agility*, which is the ability to react to changes in resource availability, and *authority*, which is the ability to control the way that resources are used. Although they are important concerns, the present author thinks that these concerns should be treated at the application level, not at the language level.

Some programming languages for mobile computation are described and analyzed in [10] and other articles, and briefly described here:

- **Java**[3] is a strongly-typed object-oriented language. **Java** deals with security[12] and allows transmitting program byte-code to be interpreted by the **Java** Virtual Machine[28], but does not migrate computation. It supports weak mobility with dynamic linking. Security level is increased by the byte-code verifier at loading-time. Some security problems have been found[12]. It was shown that the ability to break **Java** type system leads to an attacker being able to run arbitrary machine code[11]. Static and dynamic type checking.

- **Telescript**[41] is an agent-based and object-oriented language that explicitly deals with locality, strong mobility, and finiteness of resources. There are two kinds of Execution Units: agents and places. Typically, when an agent is running on an interpreter, the *go* instruction causes the agent execution to be suspended, its code and current state are transmitted to a remote virtual machine and, there, the computation is resumed. However, agents do not maintain connections to remote agents. The **Telescript** run-time code is interpreted

without security checking since security is ensured at the language level. The replication strategy is dynamic, by move. Static scoping and name resolution. Static and dynamic type checking. In spite of the historical reasons for mentioning **Telescript** here, that technology was replaced by **Odyssey**[9], which is a **Java**-based version of **Telescript**, briefly speaking.

- **Tycoon**[29] provides thread migration like **Telescript**. It is a polymorphic, higher-order functional language with imperative features, which may support other paradigms indirectly, including object orientation. **Tycoon** provides strong mobility and support for persistent programming. All objects in this language have first-class status. Static and dynamic type checking, dynamic replication with strategy by copy, besides static replication strategy.

- **Agent Tcl**[18] provides strong mobility where the whole image of the interpreter can be transferred to a different site by executing a *jump* instruction. **Agent Tcl** also provides weak mobility by executing a *submit* instruction which allows transmission of procedures along with part of their global environment, to a remote interpreter. Typeless language therefore no type checking. Dynamic replication strategy, both by copy and by move. **Agent Tcl** is a Ph.D. Thesis[19].

- **Safe-Tcl**[5] supports active e-mail, where messages may include code to be executed when an interpreter reads the message after receiving it. However, **Safe-Tcl** does not support active e-mail code mobility at the language level but, instead, code mobility is achieved through a dynamic code loading mechanism. Typeless language therefore no type checking.

- **Obliq**[4] is an object-based language that encourages distribution and mobility. While a mobile object is migrating from one place to another, new connections are automatically open between source and destination places in order to guarantee that any change in the variables will update the state in the source place. Therefore, object references are transformed into network references. Although a simple language, there is some loss of efficiency and robustness due to some possibly very large number of connections in an unreliable environment. Dynamic type checking, sharing strategy.

- **Facile**[36] is a functional language, a superset of **ML** with primitives for

distribution, concurrency and communication. Mobile code programming was later added to this extension[27]. Static and dynamic type checking. Dynamic replication strategy by copy, besides static replication strategy.

- **TACOMA**[20], [21], the **Tcl** language plus primitives to allow a running **Tcl** script to send another script and initialization data to another host in order to execute the script remotely. Typeless language therefore no type checking. Dynamic replication by copy.

- **MO**[10][37] is a stack-based interpreted language which provides weak mobility and run-time type checking. Dynamic type checking, dynamic scoping rules, dynamic replication by copy.

Aglets Workbench, developed by IBM, is a mobile agent system based on **Java**. Like others, such as ObjectSpace **Voyager**, the system security and other issues depend on the **Java** system[24].

As mentioned before, in the present paper, the present author discusses some of the features of the programming language.

3. THE PRESENT PROGRAMMING LANGUAGE

The present sample programming language is a language that supports mobile agents, syntactically somewhat similar to **Java**. It supports strong mobility, as well as some forms of knowledge and belief representation, reasoning, and uncertainty treatment. As an on-going experimental project, security has not been a concern¹. The virtual machine interprets byte-code and the language provides both replication strategies by programmable handlers. BNF legend: boldface letters are keywords; italic words with initial capital letter are other terminal symbols; words in lower-case letters are non-terminal symbols; meta-symbols: | indicates alternative, e is the empty symbol of the grammar. Other terminal symbols: { (, ;) } are used in the grammar. The following BNF definition is of a very simple subset of the programming language in question, and where the first symbol denotes the starting symbol of the syntax:

¹ However, the first mobile agent in this language, referred to as **Se-Picou**, had perfectly run over the Internet: briefly, as a prototype, **Se-Picou** started running in the city of Edinburgh; went to a site in Brazil; made some calculation; sent partial results by e-mail; moved back to Edinburgh and finally resumed.

```

aprog ? classlist commandlist

classlist ? e | classdef classlist

type ? int | list

modifier ? private | public | e

onevardef ? id | assignment

idlist ? onevardef | onevardef ',' idlist

vardef ? modifier type idlist ','

handler ? evaluator | reactor

evaluator ? when id ',' do command

reactor ? when id ':=' do command

classdef ? class id '{' defs '}'

defs ? e | vardef defs | function defs | handler
defs

command ? assignment | '{' commandlist '}'
| functioncall | ifcommand | return |
return) expression

assignment ? id ':=' expression

ifcommand ? if expression then command |
if expression ',' command |
if expression ',' command ifnot command |
if expression ',' command else command |
if expression ',' command otherwise
command |
if expression ',' command
ifnot command
otherwise command

commandlist ? e | command ',' commandlist

```

where non-terminal symbols, namely *function*, *functioncall* and *expression* are as usual. In **Plain**, they are somewhat syntactically similar to **C++** or **Java**. The main difference is that the symbol \$ can be placed where a variable identifier is expected, as it will be explained below. There are other details that will be explained together with the examples. Both [13] and [14] formalize the semantics that will be explained.

4. UU IN GLOBAL COMPUTERS

As already stated, *uu* is a constant which stands for "unknown" or "undefined" and represents unavailable information.

More precisely, *uu* can be used in programming languages in accordance with the following description: For every data

type if any, the language designer can add a special value, namely *uu*, to represent lack of some *domain value*, i.e. some known value in the problem domain. Grammatically, *uu* or *unknown* is a constant. Variables either have *uu* or some domain value. In advance, besides other applications of *uu* in some programming language, *uu* can support fault tolerance over the Internet, and this will be clear while the subject is introduced.

Some languages adopt a default value as initial variable contents. Nonetheless, since there is now *uu*, any variable (at least in the **Plain** programming language) contains this value as the initial one. Programmers should certainly want to initially assign some values to some variables.

Handlers are provided for variables. For any variable, there can be one *evaluator* and/or one *reactor*, independently. As well as other purposes, one handler can protect a variable. The notion of evaluator is for permitting the programmer to write a piece of code for producing and providing some domain value to the corresponding variable, while the idea of reactor is for inspecting and protecting the variable against assignments. Thus, a reactor permits the programmer to write a piece of code to be executed whenever a value is meant to be stored in the corresponding variable. For instance,

```
int x, y;
when x, do { x := 3 * y; }
when x := do { x := $/2; }
```

is an example where two handlers are defined for the variable *x* (Remember, here a handler is a declaration **when**). At the first time that the value of *x* is being requested in an expression, the above evaluator is triggered, which in turn computes the triple of the value of the variable *y* assigning it to *x*. From the second time on, the computed value $3 * y$ is already available in *x* and, because of this, the evaluator is not triggered. This idea is not limited to *exception handling*, which in turn is a mechanism supported by some other languages.

An evaluator can contain **return** statement (similar to **C**) as an alternative to assigning a value to the requested variable. In the case of the **return** statement and no prior assignment in the evaluator, the evaluator is always triggered when the variable in question is being used, unless

some domain value has been assigned to that variable outside the evaluator.

Whenever a value is meant to be stored in *x*, the control is jumped to the corresponding reactor. Notice that the $\$$ symbol above is used in reactors to represent the value that, in other languages, would be stored unconditionally. In the above example, half the value is accepted.

Predicates can be provided to check whether a variable contains *uu*, for instance, **known** and **unknown**. For each of them, the value is accessed directly from the variable in question and, then, the corresponding binary (either true or false) result from the predicate expression is provided by the interpreter without evaluating the handler of this variable.

If the referring variable contains some value in the problem domain, the semantics is exactly the same as in imperative languages. However, if the variable contains *uu* instead, there are two semantic cases: If there is an evaluator, it is executed. Otherwise, i.e. when no evaluator exists, *uu* is used instead. Moreover, the semantics of an evaluator is not very similar to the semantics of any function call, for the latter is necessarily executed. As an example, in the case of Remote Procedure Call or Remote Method Evaluation, for being both remote, unconditional calls are probably left as the last resort.

In terms of programming language design, *uu* and handlers together replace exception handling used in other languages. This replacement tends to make any language more economic and hence much easier to use. More than this, handlers are very useful during program *testing* and *debugging* phases, because one might inspect what is being stored and, as mobile agents often escape from any of their users, *uu* together with handlers can be used in such a program for sending results to the person who is testing it. For example:

```
class mycl {
  public int x;
  private list queue := [ ];
  when x := do {
    x := $;
    queue := queue +
    [ [ #self + ``.x := `` + $ + `` at `` +
    LocalTime() ] ];
  }
}
```

```
mycl c; c.x := 10; c.x := 20; c.x := 30;
```

In the above class, or its subclasses, whenever *x* receives any value, this value is additionally stored in the queue together with the name of the object in question (which is referred to by the expression *#self*), the name of this field with one dot (it is *`.x`* in this case), and the current local time (accessed by *LocalTime()*). The '+' operator the concatenation of lists or strings, besides the arithmetic addition, as more usual. The square brackets are used to construct a list of values of any type. Here the programmer chose list of lists for programming reasons.

Because it is hard to implement debugging system for mobile agents in a relatively satisfactory way, the above piece of code can be written as long as handlers exist. More generally, when a mobile agent stops running, the local runtime system should provide a way of returning that agent to its home, when requested, in order to permit local debugging. Therefore, by using a small query language, the user can inspect the contents of the variables of his or her program. More than this, by writing the *trace* declaration in such a program, a mobile object support system can internally maintain debugging variables.

The difference from other paradigms might become decisive in language design. On the one hand, a variable in an evaluating expression may cause its value to be read from a data base or requested from a remote process, provided that its current value is *uu*. Further, a variable may work like a *cache* because in the subsequent uses, some domain value is available locally and the handler is not triggered. On the other hand, to assign a value to a variable may cause its value to be stored on a data base or sent to a remote host.

The following piece of code exemplifies a persistent field *p* and a remote field *r* that can live together in the same class:

```
class remoteandpersistentcl {
    public int p, r;

    public void ini(int i, int j) {
        inttodb(`p",i); p := i;
        inttourl(`www.aaa.bbb.ccc/cgi/server/r.txt",j);
        r := j;
    }

    when p := do { p := $; inttodb(`p",p); }
```

```
when p, do { return intfromdb(`p"); }
```

```
when r := do {
    r := $;
    inttourl(`www.aaa.bbb/cgi/server/r.txt",r
);
}
```

```
when r, do {
    r :=
intfromurl(`www.aaa.bbb/cgi/server/r.txt");
}
}
```

```
remoteandpersistentcl c;
c.p := 20; // also store 20 locally on data
base.
sendlocally(home,c.p); // send (c.p) home.
c.r := 30; // also update remotely.
sendlocally(home, c.r); // send (c.r) home.
```

Notice that, according to the above evaluator and program, while the *p* field is being retrieved from a data base (whenever its value is requested), the *r* field behaves as a cache over some global environment. The functions *inttourl*, *intfromurl* and *sendlocally* tell the underlying system to generate internally mobile agents to take part in the protocol. There has been a general criticism concerning mobile agents because they do not maintain connections. The present author agrees that agents should not see connections, but the mobile agent support system should provide remote communication in an appropriate way. This produces positive effects and abstractions in the programming language.

Here we concentrate on features for global computing and present some new aspects of *uu*.

5. LAZY EVALUATION AND TIMEOUTS

Lazy evaluation is one of the most interesting characteristics of programming, in particular in applications where time is regarded as important. In this paper, the present author is not regarding lazy evaluation as being only *call by need* of functional languages. If the language provides functions, lazy evaluation can also be very useful in the same platform, from the same point of view of the present section.

Programming for mobile agents on a global environment tends to be more personal. One of the reasons is that patience and

mood vary for different people as well as for the same person at different instants, and one of the purposes of agents is to represent users.

As an example of a situation, a mobile agent *ma* can communicate with a stationary agent *s* which in its turn can send a small agent remotely to *ma*'s home in order to return some piece of information to *s* which in turn can hand it to the mobile agent *ma*. To deal with faults and delays in communication, a timeout can be set, implicitly or explicitly, for every input operation. After that time, the result is *unknown* (*uu*) and the computation continues normally. Similarly, every output operation has a timeout.

In this way, the same statement can be executed at different locations[7], either sequentially or not. This situation happens often. Cache-like variables might produce a similar result as lazy evaluation. Computing with timeouts together with *uu* and handlers is not lazy evaluation, but it can give a somewhat similar impression of *impatience* and, because the resulting value in this case of exception is *uu*, variable values in programs are always sound and finally this scheme improves *agent robustness*.

Another way of dealing with faults and delays is to provide a standard semantics for basic operations such as arithmetic and relational. In particular, if the first operand is *uu*, the expression might result in *uu* without the evaluation of the second operand. This is a form of lazy evaluation.

6. OTHER FEATURES

Because generality is desirable, choices among various strategies for binding resources should normally be programmed. Handlers may be used to implement different strategies for variables that are resources, either local or remote.

During the compilation, in order to support higher-level communication between agents, names of objects in the source program can be written in the object code, which increases the agent size but it is still a good idea. A possibility is to generate only names defined in the dynamic part of the interface. If the language supports artificial intelligence techniques, perhaps it is even interesting to consider the idea of generating all names. Communication between agents can be set from a prefix in function calls containing the name of the destination agent. For example,

in $x := prov:func(params)$, the string variable *prov* is a name that indicates the agent which in turn might contain the *func* function definition. The *prov* value is an absolute (global) or relative (to the local host) address. If such a matching name of *func(params)* is undefined in that agent when the call is executed, *x* receives *uu*. In every function call (or method invocation) between two agents, a timeout can be attached. For example, in $x := prov:func(params) \text{ timeout } 3$, if the operation is not completed before 3 seconds, at that time it is interrupted and *x* receives *uu*.

The concepts of **home** and **Id** of agents ought to be key words in the programming language, in a similar way as exemplified above, outside the class *remoteandpersistentcl*.

Surprisingly, although *concurrent programming*[6] is an important technique that can help in certain applications, it is not a specific feature for mobile agent programming languages, as concurrency can be achieved at the operating system level.

However, *uu* permits a large number of parallel operations, not only parallel *and* and parallel *or*[14].

7. CONCLUSION

The presented concepts are harmonious with many approaches in software engineering such as in [22].

Local inefficiency is an issue of the features discussed in this paper. Assuming that, in practice, mobile agent support systems entail code interpretation, the interpreter has to check the presence of *uu* whenever a variable is being requested in an evaluating expression. However, as hardware is getting faster and larger, this is not considered a significant problem. Moreover, this problem can be compensated for the fact that mobility and remote accesses are the bottleneck in applications, and that variables can behave as cache and operations can be lazy. This combination is encouraged by the language.

ACKNOWLEDGMENT

My honest thanks to the reviewers who kindly collaborated with suggestions here.

REFERENCES

- [1] Acharya, A., Ranganathan, M., Saltz, J., "Dynamic linking for mobile programs", in *Mobile Object Systems: Towards the Programmable Internet*, Springer-Verlag, Apr. 1997, pp. 245--262, Lecture Notes in Computer Science, 1222.
- [2] -----, "Sumatra: A language for resource-aware mobile programs", in *Mobile Object Systems: Towards the Programmable Internet*, Springer-Verlag, Apr. 1997, pp. 111--130, Lecture Notes in Computer Science No. 1222.
- [3] Arnold, K., Goslin, J., "The Java Programming Language", Addison-Wesley Publishing Company, 1996.
- [4] Bharat, K., Cardelli, L., "Migratory applications", in *Mobile Object Systems: Towards the Programmable Internet*, Springer-Verlag, Apr. 1997, pp. 131--149, Lecture Notes in Computer Science No. 1222.
- [5] Borenstein, N. S., "Email with a mind of its own: The Safe-TCL language for enabled mail", First Virtual Holdings, Inc, Tech. Rep., 1994.
- [6] Burns, A., Wellings, A., "Concurrency in Ada", 2nd ed, Cambridge University Press, 1998.
- [7] Cardelli, L., "Global computation", *ACM, Computing Surveys*, vol. 28A, no. 4, 1996.
- [8] Cardelli, L., Davies, R. "Service Combinators for Web Computing.", <http://www.luca.demon.co.uk>, 1997.
- [9] G. M. Corp., "Odyssey White Paper", 1998.
- [10] Cugola, G., Ghezzi, C., Picco, G. P., Vigna, G., "Analyzing mobile code languages", in *Mobile Object Systems: Towards the Programmable Internet*, Springer, Apr. 1997, pp. 93--110, Lecture Notes in Computer Science No. 1222.
- [11] Dean, D., "The security of static typing with dynamic linking" in *Proceedings of the Fourth ACM Conference on Computer and Communications Security*, Zurich, Switzerland, April 1997.
- [12] Dean, D., Felten, E. W., Wallach, D. S., "Java Security: from Hotjava to Netscape and Beyond", in *Proceedings of the Symposium on Security and Privacy*, IEEE, Oakland, Cal., 1996, pp. 190--200.
- [13] Ferreira, U., "uu for programming languages", *ACM, SIGPLAN Notices*, August 2000, vol. 35, no. 8, pp. 20--30.
- [14] -----, "Programming languages features for some global computer", in *Proceedings of SSGRR 2003s International Conference on Advances in Infrastructure for e-Business, e-Education, e-Science, e-Medicine, and Mobile Technologies on the Internet*, IPSI, Scuola Superiore G. Reiss Romoli and Telecom Italia Learning Services, From 28 July to 3 August 2003.
- [15] Freeman, E., "Supercomputer earth: Massively parallel internet", Yale University, Tech. Rep., December 1993, supplement to the Yale Weekly Bulletin.
- [16] Gelfond, M., Lifschitz, V., "Logic programs with classical negation", in *Proceedings of 7th International Conference on Logic Programming*, The MIT Press, Cambridge MA, 1990, pp. 579--597.
- [17] -----, "Classical negation in logic programs and disjunctive databases", *Ohmsha Ltd and Spring-Verlag*, New Generation Computing, 1991, pp. 365--385.
- [18] Gray, R. S., "Agent TCL: a transportable agent system", in *Proceedings of the CIKM'95 Workshop on Intelligent Information Agent*, 1995.
- [19] -----, "Agent TCL: A flexible and secure mobile-agent system", Dartmouth College, Computer Science, Hanover, Tech. Rep. PCS-TR98-327, 1998, Ph.D. Thesis, 1997.
- [20] Johansen, D., "Mobile agent applicability", in *Mobile Agents: Second International Workshop, MA'98*, Lecture Notes in Computer Science, vol. 1477, Springer, 1998.
- [21] Johansen, D., van Renesse, R., Schneider, F. B., "An introduction to the TACOMA distributed system", Department of Computer Science, University of Tromsø, Tromsø, Norway, Tech. Rep. 95-23, June 1995.
- [22] Jourdan, G.-V., Zaguia, N., "Incremental software development strategy: A successful experience", in *Proceedings of The 2nd International Conference on Computer Science and its Applications*, Dey, P. P., Amin, M. N., Gatton, T. M., Eds., National University, San Diego, CA, USA, June 2004, pp. 100--104.
- [23] Kakas, A. C., Kowalski, R. A., Toni, F., "The Role of Abduction in Logic Programming", in *Handbook of Logic in Artificial Intelligence and Logic Programming*, Oxford University Press, 1998, vol. 5, pp. 235--324.
- [24] Karjoth, G., Lange, D. B., Oshima, M., "A security model for agents", *IEEE, Internet Computing*, vol. 1, no. 4, July/August 1997.
- [25] Kauffman, S., "Molecular autonomous agents", *Philosophical Transactions of The Royal Society*, June 2003, vol. 361, no. 1807, pp. 1089--1099.
- [26] Kleene, S. C., "Introduction of Metamathematics", D. Van Nostrand, Princeton, 1952.
- [27] Knabe, F. C., "Language support for mobile agents", Ph.D. dissertation, Carnegie Mellon University, Paitsburgh, Pa., Dec. 1995, also available as Carnegie Mellon School of Computer Science Technical Report CMU-CS-95-223 and European Computer Industry Centre Technical Report ECRC--95--36.
- [28] Lindholm, T., Yellin, F., "The Java Virtual Machine Specification", Addison-Wesley Publishing Company, Reading, Massachussets, 1997.
- [29] Mathiske, B., Matthes, F., Schmidt, J. W., "On migrating threads", *Fachbereich Informatik Universitat Hamburg*, Tech. Rep., 1994.
- [30] Da Silva, M. M., "Mobility and persistence", in *Mobile Object Systems: Towards the Programmable Internet*, Springer-Verlag, Apr. 1997, pp. 157--176, Lecture Notes in Computer Science No. 1222.
- [31] Da Silva, M. M., Atkinson, M., "Combining mobile agents with persistent systems: Opportunities and challenges", in *2nd ECOOP Workshop on Mobile Object Systems*, Linz, Austria, July 1996, pp. 36--40.
- [32] Ousterhout, J. K., "Tcl and the Tk Toolkit", Addison-Wesley, 1994.
- [33] Russel, S., Norvig, P., "Artificial Intelligence", 2nd ed., Prentice Hall, 2003, ch. Two.
- [34] Schmidt, D. A., "The Structure of Typed Programming Languages", The MIT Press, Foundations of Computing Series, 1994.
- [35] Tanenbaum, A. S., "Computer Networks", 4th ed, Prentice Hall PTR, 2003.
- [36] Thomsen, B., Leth, L., Prasad, S., Kuo, T.-M., Kramer, A., Knabe, F. C., Giacalone, A., "Facile antigua release programming guide", European Computer Industry Research Centre, Munich, Germany, Tech. Rep. ECRC--93--20, Dec. 1993.

- [37] Tschudin, C., "The messenger environment MO -- a condensed description", in *Mobile Object Systems: Towards the Programmable Internet*, Springer, Apr. 1997, pp. 149--156, Lecture Notes in Computer Science, 1222.
- [38] Lukasiewicz, J., "Jan Lukasiewicz Selected Works", *North-Holland Publishing Company and PWN - Polish Scientific Publishers*, Series on Studies in Logic and Foundations of Mathematics, 1970.
- [39] Vitek, J., Serrano, M., Thanos, D., "Security and communication in mobile object systems", in *Mobile Object Systems: Towards the Programmable Internet*. Springer-Verlag, Apr. 1997, pp. 177--200, Lecture Notes in Computer Science No. 1222.
- [40] Volpano, D., "Provably-secure programming languages for remote evaluation", *ACM Computing Surveys*, vol. 28A, Dec. 1996, participation statement for ACM Workshop on Strategic Directions in Computing Research.
- [41] White, J., "Telescript Technology: the Foundation for the Electronic Marketplace", *General Magic, Inc.*, 1994.

Information Extraction from Texts: Adapting a System for Summarization of News Reports to the Domain of Bioinformatics

Barbara Gawronska

Abstract — Natural language serves as important information source in all areas of human activity. The presence of a huge amount of texts on the Internet actualizes the problem of efficient information search; visual scanning of all the textual information is difficult and time-consuming. There is a need for efficient, high-quality systems that extract the relevant information from texts. The paper presents the architecture of an experimental system for automatic text understanding and information extraction, which has originally been developed for the domain of news reports. The possibility of adapting the methodology for the purpose of bioinformatics is discussed, and the similarities and differences between texts in the two different domains are discussed and exemplified.

Index Terms — bioinformatics, Information Extraction, Natural Language Processing, text understanding

1. INTRODUCTION

The field of Natural Language Processing (NLP) encompasses both theoretical work and practical applications aimed at transforming Natural Language (NL) input into less ambiguous data structures and/or generating NL output, formatted according to the user's need.

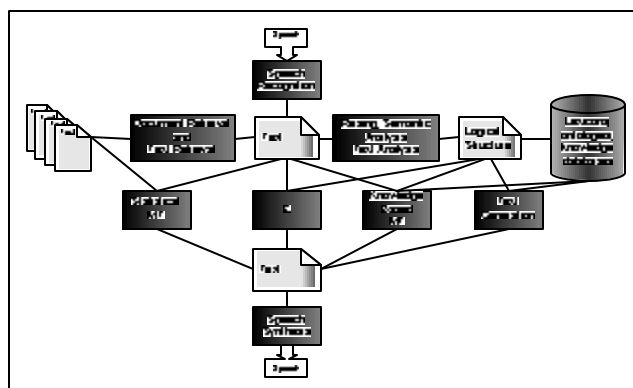
Figure 1 shows how large text corpora and lexical databases interact with language processing tools in order to achieve the desired transformations of the input. For example, spoken language may be

transformed to a text by speech recognition. The text may then be translated into another language, or it may be reduced into a shorter, more concise version by an Information extraction system. Other possible NLP-application can be inferred from the pathways outlined in Figure 1.

A huge amount of NL texts is nowadays available in electronic form and can to a certain extent be sorted and clustered automatically by search engines like Google, AltaVista and others. Despite this, an effective search for relevant information in NL texts requires a lot of human effort and is sometimes extremely time-consuming.

There is no field of international research or business activity where electronic texts written in natural language are not involved; therefore, tools for efficient and goal-oriented NLP are required.

The problem of finding relevant information in large number of texts has attracted the attention of more and more NLP-researchers during the last decade. This field of research is most commonly



Manuscript received May 4, 2004.

Barbara Gawronska is with the Department of Humanities and Informatics, University of Skövde, Sweden, (e-mail: barbara.gawronska@isp.his.se).

referred to as Information Retrieval (IR) and Information Extraction (IE), but, as often is the case with quickly emerging and growing research areas, there is no exact consensus as to the terminology. Certain terms are defined and used in different ways by computer scientists, computational linguists, and information researchers. Terms like Information Retrieval vs. Document/Text Retrieval, Information Extraction vs. Information Refinement, or Summarization vs. Abstracting, are in some contexts used as nearly synonymous, in other contexts – as related by subsumption.

This presentation adheres to Cowie's and Wilks' definition of IE, stating that IE "is the name given to any process which selectively structures and combines data which is found, explicitly stated or implied, in one or more texts" [1].

II. AN EXPERIMENTAL SYSTEM FOR EXTRACTING INFORMATION FROM NEWS REPORTS

During the years 1999-2002, an experimental system for summarization of news reports has been developed by the author's research group. The work has not been aimed on any commercial application. The research goals were to investigate the possibilities of automatic in-depth text understanding, based on the theoretical framework of Cognitive Linguistics, and a comparison between multilingual text summarization and traditional (sentence-by-sentence) machine translation.

Figure 2 shows the architecture of an experimental system for extracting information from news reports [5], [6], [7], [10],[12]).

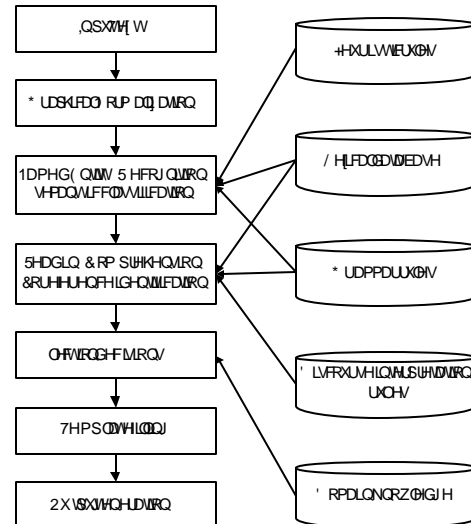


Figure 2. The architecture of the IE system for summarization of news reports.

A. The IE process step by step

The process of extracting important information from a text starts with graphical normalization: removing or replacing illegal characters, distinguishing the headline from the rest of the text, etc. The next step is Named Entity Recognition (NER) and semantic classification. This part of the system (implemented in the Delphi programming language) utilizes semantic and ontological knowledge stored in different databases (to be discussed more in detail in section III B).

The main goal of this component is to distinguish between:

- (a) Common nouns and proper names
- (b) Content words (nouns, adjectives, verbs) and function words (definite/indefinite articles, pronouns, prepositions, negation marks)
- (c) Semantic classes of the nouns (e.g. 'human', 'country', 'means of transportation'...)
- (d) Verbs referring to speech acts (say, confirm, deny, threat, propose...) and verbs referring to other types of actions and relations

Figure 3 contains a sample input text, and Figure 4 – a fragment of the output from the NER and semantic classification component.

Saudi Arabia denied it was seeking a Saudi man detained in Canada who claimed to be wanted in connection with a bombing in the kingdom that killed 19 U.S. soldiers.

The denial was made by Saudi Deputy Interior Minister Prince Ahmed bin Abdul-Aziz to the Saudi-owned London-based Asharq al-Awsat newspaper on Sunday. The man, Fahad al-Shehri, sought refugee status in Canada and said he was part of a mujahideen (holy warriors) group, though he denied involvement in the June bombing in the eastern Saudi town of Khobar, Canadian officials have said. Prince Ahmed told Asharq al-Awsat that Shehri was "not wanted in any case or

Figure 3. A sample input to the IE system (from Reuters corpus of news reports, used with the company's permission)

```
semcat('The',the,[det])
semcat(denial,deny,[speech_act(
    [fal,negative,[],[]]])
semcat(was,be,[cop])
semcat(made,make,[])
semcat(by,by,[prep])
semcat('Saudi Deputy Interior Minister Prince Ahmed bin Abdul - Aziz',[propername, human([high_status]])
semcat(to,to,[prep])
semcat(the,the,[det])
semcat('Saudi-owned London-based Asharq al -Awsat',
```

Figure 4. The result of semantic classification of the second sentence in the input text.

The next step in the process is labeled "Reading Comprehension and Coreference Identification". "Coreference Identification", [17] stands for discovering which words, or sequences of words, point out the same object (e.g. the sequence *Saudi Deputy Interior Minister Prince Ahmed bin Abdul-Aziz* is coreferent with *Prince Ahmed*). The process is based on partial sequence matching and/or matching semantic categories. A main difference between the Reading Comprehension (including Coreference Identification) process in the current system, and most traditional approaches to coreference resolution is the fact that the text is transformed from a linear sequence of sentences to a set of relations between objects, before Coreference Identification rules apply. These relations represent different events reported in the text, and/or different versions of events. An important feature is labeling every event as

- True according to a certain source,

as in *An official source said that X*

- False according to a certain source, as in *Another source denied that X*
- Future or hypothetical events, as in *The government received a warning for X*
- Actual events, as in *Today's attack did not cause any injuries*
- Background events, as in *During the last 30 years, 30 000 people died in ETA attacks.*

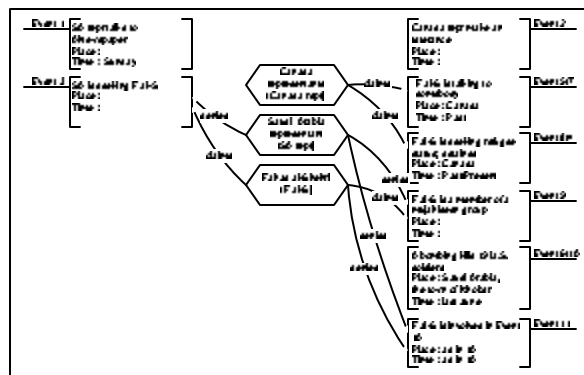
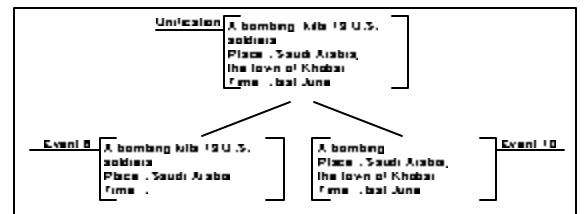
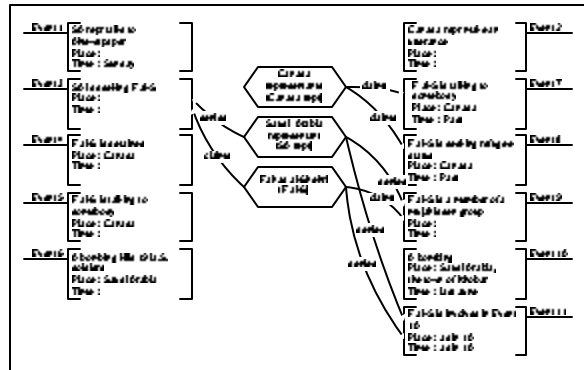


Figure 5 shows the result of the first step of text analysis. Event representations are related to the three main “actors”, or “sources” (the Saudi interior minister, the Saudi man detained in Canada, and Canada officials) by speech acts (claims and denials). Later on, event representations are unified (Figure 6), which leads to a compressed representation (Figure 7). The final step in the process is the generation of the output text in the target language (English, or Swedish). In text generation, the event templates function as an interlingua representation. The generation module is implemented in Definite Clause Grammar (DCG) with the filled template left of the arrow and language-specific English and Swedish syntactic structures on the right of the arrow. The Swedish module makes use of standardized ways of presenting certain events in Swedish news and, as a consequence, syntactic structures that are functionally, although not syntactically equivalent to certain English phrases. For example, the English sentence *There were no injuries* corresponds functionally to Swedish *Inga personskador rapporterades* (lit. *No person-injuries were reported*). The lexicon used in generation consists of phrases and words extracted from Swedish news reports and connected to interlingua codes.

A text is not the only possible output from an IE system. Another type of output could be an answer to a query (for a survey of query answering systems, see the <http://trec.nist.gov>). Yet another way to structure an IE-output is to let it remain a set of logical relations, and to use those relations in a more complex information system. In the following, we will discuss the different alternatives in the context of bioinformatics.

III. INFORMATION EXTRACTION FROM TEXTS IN BIOINFORMATICS

A. Major problems

One of the most serious problems that researchers within the field of bioinformatics have to deal with is the textual information overload [18], [19]. This is a generally acknowledged difficulty, and serious attempts to overcome it, or at least diminish

it, are in progress. For a review of the area, see [11].

The large medical literature databases MedLine and PubMed (<http://medlineplus.gov>, <http://www.pubmedcentral.gov>) provide access to medical lexicons and advanced document retrieval systems. Even a limited possibility of automatic query answering is available from MedLine’s homepage. Furthermore, text mining techniques are utilized to group the documents that concern similar genes/proteins/ diseases, and information about these is stored in ontological databases [14]. Despite all this – very valuable – work, researchers in the field of bioinformatics have to cope with several serious problems:

(a) The shortage of integration tools: specialized databases, dictionaries, and literature are available in electronic form, but the tools for synchronization of information coming from these sources are not sufficient.

There is also a need for improving existing ontologies, [26] and for developing an ontology of different kinds of relationship. The work of [23], [22], [24], and [26], is an important step towards this goal.

(b) Terminological discrepancies: information about the same gene/protein is stored under different names/codes in different databases. For example, [18] a gene with accession code “L36033” in a research database is mentioned as “HUMSDF1B” in the database Entrez [<http://www.ncbi.nlm.nih.gov/Entrez/>]; in other sources, it is sometimes shortened to “SDF1b”, or referred to by a full name: “Human pre-B cell stimulating factor homologue”. There are some resources for identification of synonymy within certain subdomains (the MedScan system [21] is one of the most ambitious and efficient systems), but the issue still remains a problem.

(c) Problems with coreference identification (as in *Loss of heterozygosity (LOH) at chromosome 9p21 is frequent in enteropathy lymphoma... this finding suggests that gene loss at this locus may play a role for...*). The importance of anaphora resolution is generally

acknowledged [22], [9]. However, only a few systems include this component. Hahn et al. [9] present an algorithm for coreference identification between hyper- and hyponyms (*the tumor – adenocarcinoma*), based on the Centering Theory [25]. The paper provides, however, no evaluation of the algorithm, nor resolution of nominal anaphora including coordination.

(d) Problems with identification of relations between objects (in this context, objects are chromosomes, genes, proteins, distortions...). Traditional text retrieval and text mining devices can inform the researcher that there seems to be some relation between e.g. a gene and a disease, but in most cases they do not specify what kind of relation it is. There is active research aimed at improving extraction of relations. [27] and [2] predict the existence of relationships between genes on the basis of textual co-occurrence statistics, while other researchers [20], [21], [4], [22] use rule- and template-based approaches for the same task. The work of [23] represents a hybrid approach (a stochastic word tagger is combined with rule-based sentence analysis). There is no doubt a progress towards identification of relations, but the existing systems process, as a rule, each single sentence separately, and do not include any coreference identification component.

(e) Problems with distinguishing between relations reported as being true, hypothetical, or absent, as in *It is beyond any doubt that the lack of gene X in chromosome Y causes disease Z*, vs. *The findings suggest that the loss of gene X may play a role in the development of disease Z*. vs. *Gene X is not involved in the development of disease Z*.

B. Lexical resources and domain knowledge for IE in bioinformatics

The problems enumerated above correspond to those that the general linguistic IE-system discussed above had to

deal with. What is needed is more sophisticated Named Entity Recognition (although there are good algorithms for identification of certain biological objects – [3]), better discourse interpretation (including modality recognition, i.e. discovering the distinctions between true, false, and hypothetical relations), and more reliable coreference resolution. The core technique of linguistic Information Extraction may thus be employed in bioinformatics, but an IE system aimed at this particular domain must have access to specialized databases as well as to general lexical databases.

A lexical database for English that is often used for Information Extraction is WordNet [16]. Especially the noun part of WordNet proved to be useful [5]. Lexical entries in WordNet are organized by synonymy, hyponymy (“is-a”) and meronymy (“has-a”) relations, which makes the database a combination of a lexicon and an ontology. The general problem concerning the use of WordNet in IE is, however, its frequently discussed and criticized fine-grainedness, that leads to extreme ambiguity [13], [15]. The original WordNet structure is thus too ambiguous to be utilized for IE in bioinformatics.

A way to overcome this problem would be to identify and annotate WordNet nodes with respect to different hierarchy types (since proteins may be classified either with respect to their structure, or their molecular functions, or their roles in biological processes), and to connect these annotated nodes to the information stored in bioinformatics databases, (like the Gene Ontology, [8], [28]) - possibly via specialized lexicons.

Figure 8 [5] outlines a possible scenario where WordNet and linguistic IE are integrated with specialized information sources.

C. Adapting a general linguistic IE system to the domain of biological texts

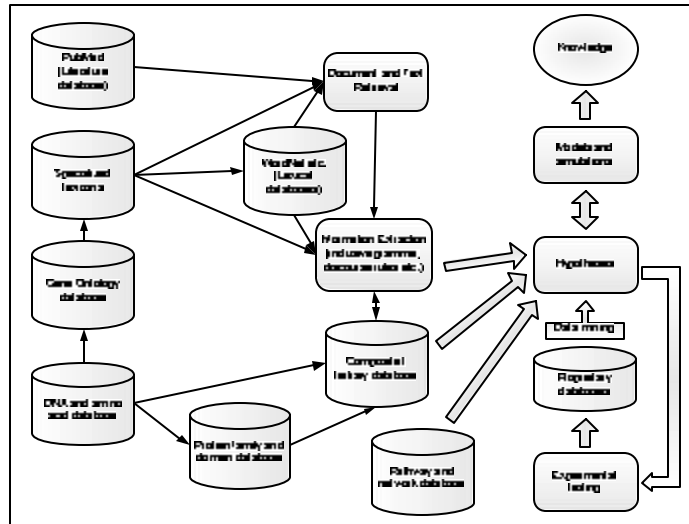


Figure 8. Information Fusion in bioinformatics [5]

An essential part of pre-processing a biological text is identification and classification of the main objects: genes, proteins, and diseases. This is in general equivalent to the Named Entity Recognition process in processing of news reports. However, the NER procedure required some additions, aimed at identification of acronym patterns and at linking acronyms with full names of the objects, like in e.g. *enteropathy T-cell lymphoma* (ETL).

Another modification concerns identification of relevant nodes in WordNet. A tool for re-grouping WordNet nodes has been developed for the purpose of semantic classification of objects in news reports [6]. For example, the first common node corresponding to the nouns *bomb* and *missile* in WordNet (version 1.6) is defined as “an artifact that is instrumental in accomplishing some end”. This definition is too general for the IE purpose, since it generates a lot of nouns that do not denote weapons. The interface developed within the project allows a convenient way of grouping relevant WordNet nodes under domain-specific labels chosen by the user. For example, the user can provide the nodes “bomb” and “missile” with a common label: “destruction as function”. The same interface can be utilized for a similar re-classification aimed at categorization of biological objects. Similarly, certain groups of verbs required a more consistent classification. For the purpose of processing news reports, a limited lexicon of speech act verbs (“Verblast”-[6]) has been constructed. The

“Verblast” used as a complement to WordNet is a database table organized around ca 60 lexical stems. The stems are connected to derivational rules (e.g. rules relating *confirmation* to *confirm*, or *denial* to *deny*), contextual patterns and interpretation rules. The latter deal with interpretation of events connected to a given lexical entry (e.g. *promise* or *suggest* normally introduces a hypothetical, future event, while *confirm* introduces an event that is true from the sender’s and somebody else’s perspective).

The “Verblast” had to be further developed in order to fit the biology domain. Speech act verbs like *say*, *deny*, *condemn*... are central for understanding news reports. In biology, however, the most frequent verbs specify either the relations between biological objects (*inhibit*, *suppress*, *activate*), or between researchers and their study objects (*examine*, *investigate*), or, finally, between study results and biological relations (*show*, *suggest*). The latter are ignored in most existing IE-tools, but they are essential to in-depth text comprehension. This group of relations may be successfully handled by the heuristics developed for the news reports domain. The system for IE from news reports achieved the following results, when tested on a corpus of ca 300 000 words [6]:

- Named Entity Recognition: recall 98%, precision 86%
- Named Entity Classification (i.e. ascribing semantic categories to strings recognized as proper names): recall 70%, precision 87%

- Identification and classification of speech acts: recall 97%, precision 86%

A pilot study performed on 30 biological abstract indicates that a similar performance can

be expected, provided incorporation of specialized databases (like Gene Ontology) into the system. This is a matter of further research. What can be reported hitherto, is the research performed on syntactic analysis of biomedical texts.

D. Parsing biomedical abstracts

In order to investigate which syntactic patterns are most frequent in biological texts, a linguistic analysis of biological abstracts from PubMed was performed. The texts consisted of totally 3500 words (after the elimination of the names of the authors and their affiliations). The analysis has shown that the number of words belonging to NPs was very high: 2559, i.e. 73% of the total number of words in the file. Personal pronouns were infrequent – only 28 occurrences, dominated by *we* (referring to the authors of the abstracts - 17 instances). This is the major difference the two text domains under discussion. In news reports the average number of personal pronouns is about 10 per one single text.

Totally 428 noun phrases were found in the biomedical abstracts. The average length of a non-pronominal noun phrase was 6.4 words. The longest NP in the material consisted of 28 words.

These quantitative results indicated that NPs should be treated carefully in the parsing process. The next step was the classification of syntactic patterns in noun phrases. The results showed that most NP-types were covered by the rules already implemented in the parser (based on Categorical Grammar and implemented in DCG) developed within the news reports project. We had to add two kind of rules: one dealing with long nominal compounds, like *candidate tumor suppressor genes*, and rules handling long appositional constructions with acronyms.

Ambiguous appositional constructions, like *the three important tumor suppressor genes located in this chromosome, CDKN2A, CDKN2B, and p14 ARF* require quite elaborated processing. To handle this kind of phrases, the parser makes use of the

following cues:

- The information from the Named Entity Classification component: CDKN2A, CDKN2B, and p14 ARF are names of genes.

- The information found in WordNet: the word *gene* is the top node in the hierarchy of genes

- A rule counting the acronyms and comparing the result with the quantifier in the first NP: three = 3.

As a result, the phrase is interpreted correctly: as referring to three (not six) genes. Figure 9 shows the output of parsing a similar example.

The parser has been tested on previously unseen texts (ca 1000 words). Its performance was satisfactory: 85% recall at 92% precision. The causes of problems are unforeseen gene and protein names, and ambiguous constructions including both *and* and *or* – for example: *Gamma irradiation, DNA-damaging drugs, expression of p14 (ARF) or oncogenic Ras, and replicative exhaustion.*

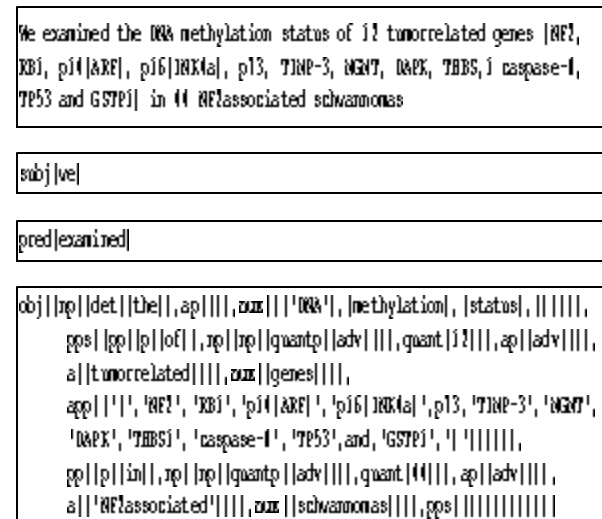


Figure 9. An example of a parse tree corresponding to a sentence with a long appositional construction

The parser performs – as an inherent part of the syntactic and semantic analysis – partial coreference identification within sentences. As mentioned, it identifies parts of appositional constructions as pointing out the same group of objects. It distinguishes predicative sentences (*A is a B*) from transitive sentences (*A causes B, A stimulates B etc.*). This is important for identification of referents: in a predicative

sentence, A and B are different descriptions of the same referent, while a transitive sentence denotes a relation between at least two referents.

Another feature of the parser is its next-to-perfect performance in analysis of relative clauses. The relative markers *that*, *whose*, *which* etc. are correctly identified with their syntactic correlates, which makes it possible to extract information not from main clauses only, but also from subordinated ones.

The functor 'nux', marked with bold style in Figure 9, identifies the kernels of possible candidates for anaphoric reference in the text. After parsing, the text is thus prepared for searching for coreferent noun phrases. In NLP, most efforts concerning anaphora resolution between sentences focus on pronominal anaphora. In the domain of bioinformatics, however, pronominal anaphora is of marginal interest. At the same time, biological abstracts are normally not comprehensible without resolution of anaphoric relations between nominal phrases: the authors refer to "this finding", "these genes", etc. It is not possible to understand (and automatically extract) the information about the reported results of biological experiments, if the actual referents of such general expressions cannot be identified in the preceding text.

For the time being, adapting the procedure for cross-sentential anaphora resolution to the new domain is under development. The preliminary results are encouraging, but the project requires a more elaborated semantic classification of the domain-specific vocabulary than the one that has been utilized for parsing. The planned incorporation of Gene Ontology may be a solution.

E. How to design the output?

Hitherto we have discussed the question of IE resources, word classification procedures, and syntactic analysis, but the question of the format of the final output also deserves some consideration. One possibility would be to structure the outputs as concise answers to queries. Another option would be to let the output have a structure that would be consistent with the hypotheses formulated on the basis of gene analysis and data mining techniques. This would give the researcher a possibility of

automated weighting of his/her hypotheses. Some human supervision would of course be required, since one can not expect an IE system to achieve 100% recall and precision, but the time-consuming process of visual scanning of literature would certainly be facilitated even at about 85% recall and precision.

IV. CONCLUSIONS

Researchers are constantly confronted by huge amounts of information. Manual and visual information search is extremely time-consuming, and often frustrating. There is an urgent need for the development of better tools for processing and structuring information coded in natural language, and of integrating such tools with domain-specific knowledge bases. Simple document retrieval, based on key words, is not sufficient. Texts should be searched for relevant information in a more elaborated and more user-oriented way. There is no doubt a resemblance and a considerable overlap between problems that have to be solved when processing an "everyday" text (e.g. a news report) and when processing a specialized scientific text. Both text kinds require Named Entity Recognition, though the entities to be recognized and classified belong to different semantic categories (persons, organizations and geographical names in news reports vs. genes, proteins, and diseases in biology). Attention should be paid to differences in name patterns. For example, appositional constructions are common in both kinds of texts, but the most frequent pattern in biology is a full name followed by an acronym, while in news texts, the pattern 'Proper name – function' (*XZ, the president of Y*) is the most widespread one.

Identification of phrases denoting relations and of the epistemic status of events (true, false, or hypothetical) is crucial to both types of texts. Again, the difference lies in the semantic categories of most frequent verbs. Coreference resolution is necessary in both domains. The general methodology for Information Extraction seems thus to be domain-independent. Domain adaptation requires a fusion of general language technology methods with domain-specific corpora, ontologies and knowledge databases.

REFERENCES

- [1] Cowie, J., Wilks, Y., (2000) 'Information Extraction'. Dale, R., Moisl, H., and Somers, H., (eds.) Handbook of Natural Language Processing. Marcel Dekker, New York.
- [2] Ding, J., Berleant, D., Nettleton, D., Wurtele, E., (2002) 'Mining MEDLINE: Abstracts, sentences or phrases?', Proceedings of PSB 2002, Hawaii, USA, pp. 326-337.
- [3] Fukuda, K., Tamura, A., Tsunoda, T., Takagi T., (1998) 'Toward information extraction: Identifying protein names from biological papers', Proceedings of PSB 1998, Hawaii, USA, pp. 707-718.
- [4] Friedman, C., Kra, P., Yu, H., Krauthammer, M., Rzhetsky, A., (2001) 'GENIES: A natural-language processing system for the extraction of molecular pathways from journal articles', Bioinformatics, Vol. 17, pp.
- [5] Gawronska, B., Olson, B., de-Vin, L. (2004) Natural language technology in multi-source information fusion, paper accepted for IPSI-2004, Kopaonik, Serbia, 1-4 April 2004
- [6] Gawronska, B., Torstensson, N., Erlendsson, B., (2004) Defining and Classifying Space Builders for Information Extraction, paper accepted for NLUCS-2004
- [7] Gawronska, B., (2002) 'Employing Cognitive Notions in Multilingual Summarization of News Reports', Proceedings of NLULP-02, Copenhagen, Denmark, pp. 103-119
- [8] Gene Ontology general documentation (2004) 'An Introduction to Gene Ontology', 18 March, <http://www.geneontology.org/GO.doc.html>
- [9] Hahn, U., Romacker, M., Schulz, S., (2002) 'Creating knowledge repositories from biomedical reports: The MEDSYNDIKATE text mining system', Pacific Symposium on Biocomputing 2002, Kauai, Hawaii, USA, pp. 338 - 349.
- [10] Harabagiu, S., Maiorano, S., and Pasca, M., (2003) 'Open-domain textual question answering techniques', Natural Language Engineering Vol. 9, Nr. 3, pp. 231 - 267
- [11] Hirschman, L., Park, J.C., Tsujii, J., Wong, L and Wu, C.H. (2002) 'Accomplishment and challenges in literature data mining for biology' Bioinformatics, Vol. 18, Nr. 12, pp. 1553-1561.
- [12] Hobbs, J., (1995) 'Sketch of an ontology underlying the way we talk about the world', International Journal of Human-Computer Studies, Vol. 43, pp. 819-830
- [13] Ide, N., and Véronis, J., (1998) 'Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art', Computational Linguistics Vol. 24, Nr. 1, pp. 1-40
- [14] Lin, D., (1998) 'Automatic retrieval and clustering of similar words', Proceedings of the Association of Computational Linguistics 1998, Montreal, Canada, pp. 768-774
- [15] Mendes, S., and Chaves, R. P., (2001) 'Enriching WordNet with Qualia Information'. Proceedings of Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations at NAACL 2001, Pittsburgh, U.S.A, pp. 107-112.
- [16] Miller, G.A., (1995) 'WordNet: An on-line lexical database of English', Communications of ACM, Vol. 38, Nr. 11, pp. 39 - 41
- [17] Mitkov, R., (2003) 'Anaphora Resolution', in Mitkov, R. (eds.), The Oxford Handbook of Computational Linguistics, Oxford University Press.
- [18] Narayanan, A., (2003) Document Technologies for Bioinformatics, Ms, Dept. of Computer Science, University of Exeter, GB.
- [19] Narayanan, A., Keedwell, E.C., Olsson, B., (2002) 'Artificial Intelligence techniques for bioinformatics', Applied Bioinformatics, Vol. 1 Nr. 4, pp. 191-222.
- [20] Ng, S.-K. and Wong, M., (1999) 'Toward Routine Automatic Pathway Discovery from On-line Scientific Text Abstracts', Genome Informatics, Vol. 10, pp. 104-112
- [21] Novichkova, S., Egorov, S., and Daraselia, N., (2003) 'MedScan, a natural language processing engine for MEDLINE abstracts', Bioinformatics, vol. 19:13, pp. 1699-1706.
- [22] Park, J.C. Kim, H.S., Kim, J.J. (2001) 'Bidirectional incremental parsing for automatic pathway identification with combinatory categorical grammar', Proceedings of PSB 2001, Hawaii, USA, pp. 396-407.
- [23] Putejovsky, J. and Castano, J., (2002) 'Robust relational parsing over biomedical literature: Extracting inhibit relations', Proceedings of PSB 2002, Hawaii, USA, pp. 362-373.
- [24] Rindfleisch, T., Tanabe, L., Weinstein, J. and Hunter, L., (2000) 'EDGAR: Extraction of drugs, genes, and relations from biomedical literature', Proceedings of PSB 2000, Hawaii, USA, pp. 517-528.
- [25] Sidner, C., (1983) 'Focusing in the comprehension of definite anaphora', Brandy, M. and Berwick, R. C. (eds.) Computational Models of Discourse, pp. 267-330. MIT Press, Cambridge.
- [26] Smith, B., Williams, J. and Schulze-Kremer, S., (2003) 'The Ontology of the Gene Ontology', Proceedings of AMIA Symposium 2003, Ottawa, Canada, pp. 609-613
- [27] Stapley, B. and Benoit, G., (2000) 'Bibliometrics: Information retrieval and visualization from co-occurrences of gene names in Medline abstracts', Proceedings of PSB 2000, Hawaii, USA, pp. 529-540.
- [28] The Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology, Nature Genet. 2000, Vol. 25, pp 25-29

easyTransport: An Interoperable and Secure e-ticketing Model based on Contactless Smart Cards

Antonio F. Gómez Skarmeta, Gregorio Martínez Pérez, and Carmen M. Yago Sánchez

Abstract—*The introduction and use of electronic ticketing (e-ticketing) is an element of key importance for the good deployment of the public and private transport area across the world. In fact, it not only produces benefits for passenger transport operators, but it also creates an infrastructure, especially in urban areas, which can be used easily by other sectors. The use of secure infrastructures to support the payment and management of such kind of e-services is a growing area of interest. This article provides an overview of the current state of the art regarding e-ticketing systems, mainly in Europe, and describes a new proposed architecture, named easyTransport, based on the use of contactless smart cards, and where important items for an e-ticketing schema are in place, such as, ease of use, interoperability, standard-based, and multi-application.*

Index Terms—*contactless smart card, e-payment, public and private transport, e-ticketing infrastructure, security*

1. INTRODUCTION AND MOTIVATION

MARKET globalization, the development of the Internet and the new information technologies have established a new scenario for business. Passenger transport environment is not an exception. Although passenger transport services are a key factor in the development of modern cities, transport operators have to deal with a complicated fare management, traffic congestion, and final customers getting more and more concerned about delays. The use of technological tools may alleviate these problems, namely an automated e-ticketing system based on contactless smart card technology. These systems have the following features:

- Quick operation. This feature is derived from contactless technology whose transaction time is about 150-400ms, much lower than magnetic stripe cards (2 or 3

seconds) [1]. This is a key factor for users' acceptance of e-ticketing systems because they usually have to wait in a queue.

- Ease of use. To operate the card it is only necessary to bring it near the card reader even without taking it out of the wallet.
- Security and Privacy. Contactless cards are more difficult to duplicate than magnetic stripe cards [2] and it is also possible to define access rights and encryption mechanisms so that information is only accessed by authorized entities and the operations made with the transport card.
- Customized fare media. It is possible to offer more flexible tariffs and diversify the kind of tickets and passed offered to final customers.
- Better management. An e-ticketing system leads to a better usage of data. Then, it is possible to have a better planning, more accurate clearing and loyalty programs.

Because of these advantages, all over the world, e-ticketing systems based on contactless smart cards are being deployed, most of them using proprietary technologies. European cities like Paris or London already have e-ticketing systems. In Europe their wide implantation is expected before 2008 [3].

In the European Union, the interest for these e-ticketing systems not only comes from transport operators but from institutions. There are initiatives at national level like ITSO (Integrated Transport Smartcard Organisation) [4] in UK to provide a platform and tool-box for the implementation of interoperable contactless smart card public transport ticketing and related services and at E.U. level like eESC (eEurope Smart Card) [5] a European Commission initiative to promote the use of smart cards including e-ticketing in transport applications.

Moreover, the standardization bodies are concerned about the necessity of developing standards at both technological and application level. ISO (International Standards Organization) has developed ISO 14443 [6][7][8][9], which standardizes the contactless technology and the

Manuscript received June 30, 2004. This work was supported in part by the Spanish Ministerio de Ciencia y Tecnología (MCYT) by means of the SAM project (TIC2002-04531-C04-04) and the Spanish Ministerio de Fomento, within its R&D 2002-2003 Project Framework in the Transport field.

T. C. Authors are with the Departamento de Ingeniería de la Información y las Comunicaciones, University of Murcia, Spain (e-mail: skarmeta, gregorio, carmen@dif.um.es).

European standardization body CEN (Comité Européen de Normalisation) is making an effort to obtain normative at application level.

In this context, we have considered interesting joining to the general effort with the development of a non-proprietary e-ticketing model, based on existent standards. This e-ticketing model, that we have called easyTransport [10] provides efficiency, reliability and security to fare collection processes without real changes in the infrastructure of transport operators, so migration has a low cost, as a real implementation has shown.

This model is suitable for any kind of passenger transport services, so it makes sense to offer the possibility of integrating several services in the same transport card. This is an added feature of our design: it allows transport operators from a geographical area to integrate their own prepaid fare media in a shared transport card.

This model also benefits from the advantages of a non-proprietary system: transport operators may analyze its specification to decide if it compatible with their current system, and any manufacturer may build components (e.g. smart cards) for the system; therefore operators are not tied to a proprietary technology or solution.

This article first describes the state of the art regarding e-ticketing systems. Then explains the easyTransport model proposed and then, how it was applied to a car park. Finally we conclude the article with our remarks.

2. STATE OF THE ART OF E-TICKETING SYSTEMS IN TRANSPORT SCENARIOS

Contactless smart card technology has led to the development of e-ticketing systems for public transportation based on it. These systems are substituting traditional ticketing systems (based on paper and magnetic stripe). Here, users introduce their tickets (and transport contracts) in their card and access the services through a touch and go system. The rapid growth and acceptance of these systems is not surprising because they present several advantages versus magnetic stripe cards based systems, as those commented above.

Because of these advantages, big scale implementations in Europe (Paris, London, Berlin, Rome or Moscow), America (San Francisco, Mexico D.F. or Santiago de Chile) or Asia (Seoul, Hong Kong, Singapore or Tokyo) are already working or will be in a short time.

But not only public transportation is implementing solutions for payment based on smart cards. The number of transport services using it is increasing, reaching for example touch

and go road tolling systems, payment parking systems or car-sharing services.

The card is usually issued by an operator or by an association of operators, and sometimes it allows the access to different services (for example in Bremen, car-sharing users may use their card to access public transport [11]) but it is because of an agreement between operators, not for the existence of interoperability. So, it is perfectly possible that users have to use one card for each transport service they want to use in their city.

In fact, from here to 2008 it is expected that e-ticketing schemes will be broadly introduced in Europe, and there is a high risk of market fragmentation [12] if research laboratories, companies and standardization fora do not consider interoperability as one of the main features in their designs.

A. Interoperability and Standardization

Interoperability means that users could access to different transport services owned by different operators using a compatible medium (e.g. smart card). Thus interoperability is a key feature to get integrated payment systems. But to achieve interoperability, both technological and application standards are required. The standardization process is still beginning, while the technology is partially standardized although there are not application standards yet.

These e-ticketing systems usually employ proximity cards. These cards communicate with the reader by radio frequency within a distance less than 10 cm. The card physical characteristics and card-reader communication protocol is standardized by ISO 14443 1-4. Card commands and memory structures are not standardized. Some manufacturers use ISO 7816-4, which is currently the most common format, but others utilize their own proprietary systems. Security protocols are normally proprietary [12].

Europe is conscious of the necessity of standards. The European Commission has launched the eESC (eEurope Smart Card) initiative to promote the use of smart cards including e-ticketing in transport applications. Moreover, the European standardization body CEN (Comité Européen de Normalisation) is making an effort to obtain normative at application level; for example CEN/TC 278 WG 11 is working on both PrEN 1545 (evolved from ENV-1545) whose scope is the coding of data elements that e-ticketing is using, and prENV IOPTA which defines how the data defined in PrEN 1545 is structured.

There are also national initiatives to develop interoperable ticketing specifications; for example Norwegian Public Roads Administration has

published a preliminary version for interoperable electronic ticketing system [13] and ITSO [4], a British association of transport operators supported by the government, is developing specifications to provide a platform and tool-box for the implementation of interoperable contactless smart card public transport ticketing and related services in the UK.

B. State of the Art Summary

There is an important migration from traditional ticketing systems (based on paper and/or magnetic stripe cards) to e-ticketing systems. Actually, in the next coming years, it is expected the massive implantation of these systems in Europe for public transportation. But these e-ticketing systems are also useful for other transport services like road-tolling or parking. So, giving the user the possibility of integrating these services in a single card is a good opportunity to provide an added value for both operators and users. It could be even the first step for integration of e-services belonging to different environments like e-government, e-health or access control.

At this moment, cards usually only contain fare media marketed by an operator or an association of operators. In this context, we consider that a step that facilitates the transition from a single-operator e-ticketing scheme to a multi-operator e-ticketing scheme with shared fare media is the development of an efficient, reliable and non-proprietary e-ticketing model permitting transport operators to share a single transport card in a secure way at the same time they maintain their independence (i.e. each transport operator may have its own fare media and management processes). This is the motivation for our proposal of an interoperable e-ticketing model, that we have called easyTransport.

3. THE EASYTRANSPORT E-TICKETING MODEL

In this section, we introduce easyTransport, our design proposal for a non-proprietary and interoperable e-ticketing system based on contactless smart card technology for transport services such as public transport, car parks or road tolling. This e-ticketing system allows transport operators to have a better fare collection management as well as to offer a better service to final customers. The easyTransport model has the advantages of e-ticketing systems described above such as quickness, flexibility, security or easiness of use. Moreover, our design offers also other interesting features:

- Non-proprietary. This feature avoids the dependency of operators from software

and hardware providers and manufacturers, and allows transport operators to have a complete knowledge of the easyTransport e-ticketing model.

- Oriented to passenger transport services.
- Shared transport card. This feature allows transport operators from a geographical area to collaborate to have a shared transport card, offering users the possibility of storing in a single transport card, prepaid fare media for different services defined by several operators.

The easyTransport model is also being developed considering that, in the future, it may be integrated in a multi-application city card, which might be used for the e-services offered in a city such as access control, health care or e-government.

A. Actors involved in the easyTransport Model

The different actors interacting in the model, whose relationships are depicted in Figure 1, are:

- Transport operators. Companies offering transport services. These services may be related to mass transport like urban transport, or to user-oriented services like car parks or road tolling. These operators group in alliances to implement the model in a geographical area and offer a shared transport card.
- Users. People who use the transport services offered by operators. They have a transport card, which stores their tickets.
- Retailers. Commercial establishments (for example kiosks), which have an agreement with transport operators to distribute their transport card or reload their prepaid fare media.
- Hardware and software providers. Companies, which provide to the operators hardware and software, which is compliant with the model.

In our model, there is also another special actor, the *easyTransport Authority*, an independent entity creating the non-proprietary specifications of the model. It is also in charge of admitting and supervising the alliances between different operators.

The easyTransport model is implemented by alliances of operators. Each operator deploys its own e-ticketing system but its transport card is valid for all the e-services offered by any operator member of its alliance. These alliances are independent one from another but the easyTransport authority supervises all of them in order to provide the general security of the alliance, and guaranteeing that cooperation is not a risk for them. Transport cards belonging to different alliances are not compatible.

B. The easyTransport e-Ticketing System

The easyTransport e-ticketing system follows a front-office/back-office structure (see Figure 2) where the front-office groups all elements which

interact with users, and the back-office processes and stores all the transactions made in the front-office. Both entities are explained in details in the next two sections.

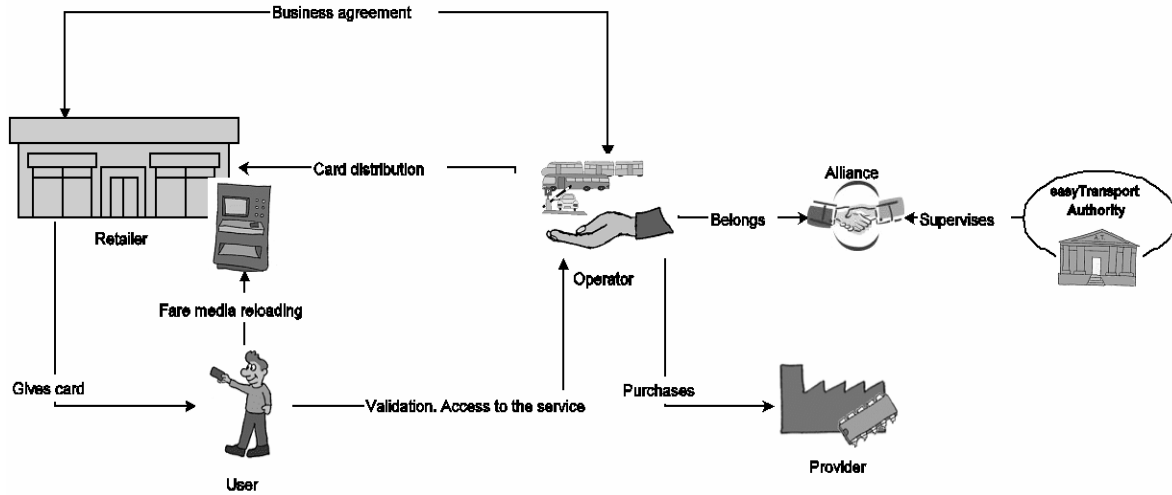


Figure 1 Actors involved in the easyTransport model and their relationships

1) Front-Office

The front-office makes all operations requiring interaction with users. Its two main components, because they are always present in every easyTransport implementation, are:

- Vending machine. It is used to introduce and reload prepaid fare media in the transport card and also to consult the state of the transport card.
- Validator. Users access the services by bringing the transport card near the validator, which selects the adequate fare media for the service, registers the transaction, and if it is the case, decrements a counter. This process is called validation. In certain services, like car parks, it is necessary to repeat the validation process to end the service (double validation).

In addition to these main components, there are two auxiliary components, which have human presence: the window and the inspector. The former has the mission of serving the users by giving them information about the system, answering their questions and solving any problems they could have. The latter is in charge of checking if the users are making a correct use of their transport card. He or she carries a device (for example a PDA) to interoperate with the transport card. These two auxiliary components can or cannot be present in the e-ticketing system depending on operator necessities.

Users interact with the front-office through their transport card, so a special attention has been paid to its design. This card is compliant with the ISO/IEC 14443 proximity contactless smart card standards and contains the prepaid fare media

and data structures required to manage and secure them. These data structures take in mechanisms for allocation of new products, selection of the right product, checking of data integrity (such as an electronic signature or cyclic redundancy codes), backup to recover the data from a failed transaction and security such as access key or access rights.

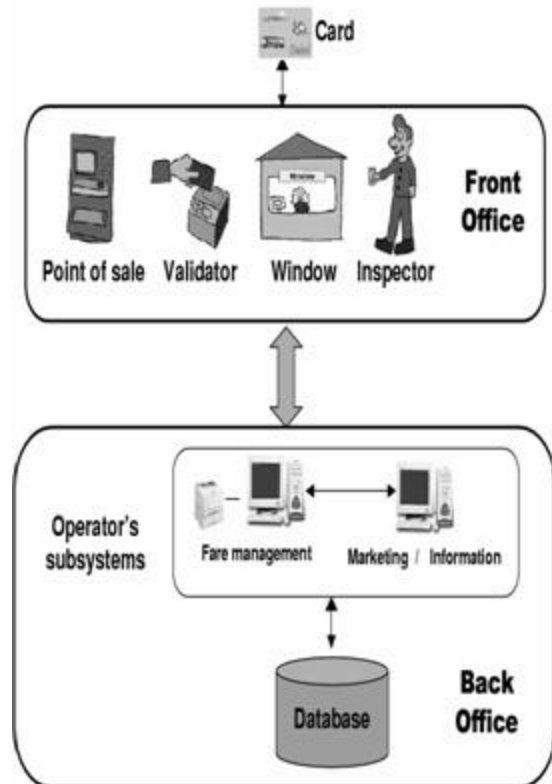


Figure 2 easyTransport e-Ticketing System

In order to provide the possibility of sharing the transport card, its memory is divided in common blocks shared by all operators of the alliance to store common information and private blocks specifics from each operator to store its fare media.

2) *Back-Office*

The easyTransport model does not define the way the back-office is organized because we consider that is strongly dependant from the business model and objectives of operators. Then, if we defined a fixed back-office, the model would not be able to fit the requisites of the operators or would suppose a disproportionate cost. However, in order to attend user petitions and for error recovery purposes, every back-office component in the easyTransport model has to store, at least, all transactions made by the front-office for either a year or the period fixed by the laws in case it should be higher.

The front-office/back-office communication may be done in an on-line or off-line way and using different media like GSM, the Internet or even magnetic support. However privacy and data integrity must be always assured. If the Internet is used, the channel must be secured and we recommend the use of public key cryptography, with similar experiences to those detailed at [14].

4. *A CASE OF STUDY: EASYTRANSPORT MODEL APPLIED TO A CAR PARK*

In this section we describe how the easyTransport e-ticketing system is currently implemented in a public car park, and what was the perception of final users and the car park managers and employees regarding the use of this new system.

The goal of this real implementation was, without supposing significant changes in the car park operation, to implant the system to provide higher transaction speed because of contactless technology, flexibility due to higher storage capacity of smart cards, better fare management system as a consequence of the registration of all operations, and security as a result of communication encryption, access rights and the use of a public key infrastructure.

The car park initially offered two kinds of passes: a purse-pass oriented to sporadic users, which used a closed electronic purse and a temporary-pass oriented to habitual users who paid for a limited period of time.

The real implementation of the system indicates that users are quite proactive to use these new set of contactless technologies, mainly because they have the possibility of having faster transactions and they can use the same smart card for several services inside the car park and in the near future outside it (i.e. busses, trains,

etc). Other considerations, such as, higher level of security are not very much of their interest, but are really sound to the company running the car park. However, the contactless manufacturer initially selected for the first real prototype provided some problems regarding working distances (no more than 7-8 centimeters); in fact, the driver in the car needed to be quite near to the validator to open the barrier, which supposed an important drawback for final acceptance of the system. It was solved with a new manufacturer, who provides higher smart card readers for the same contactless technology; thus, the same smart cards are used, but now with a working distance of around 20-25 centimeters, which is considered as acceptable by the final users and the car park managers and employees.

C. *e-Ticketing System Architecture*

The car park final implementation is shown in Figure 3 and it describes the LAN interconnecting the main elements (vending machines, and validators) and the window commented before. The main elements have the following functionality:

- Entry validators. There is an entry validator in any of the entrance gates of the car park (3 in the real implementation). They check the transport card looking for a valid product and register the entry time in both the card and the device. If one car driver has several cards (from several easyTransport implementations), the system selects its own card according to a product ID and a symmetric key. When these operations have been carried out, the entry barrier is opened.
- Exit validators. There is an exit validator in any of the exit gates of the car park. They register the exit time in both the card and the device, and if the user has a purse-pass they decrement the correspondent amount from it, and open the exit barrier.
- Vending machines. They are placed inside the car park and have an auto-explicative graphical interface in order to allow users to reload their passes and to consult their state (for example amount remaining or period of validity).

These operations have to be known in real time by a central element due to the necessity to know the state of the car park, i.e. the number of free places. In this architecture this function is performed by the window, which then it is not merely an informative element but becomes the core manager of the system. Then, besides being the adequate place to introduce the system to users and to solve their problems, the window

controls the car park state and orders the entry validators not to open the gates if the car park is full. To do this, the window needs to know the transactions made by the main elements. After this, the window sends all these data to the back-office, using SFTP with public key authentication as the main communication protocol.

The car-park security is oriented in two ways: the infrastructure must be protected from system failures, as well as from non-authorized access from the outside. To deal with system failures, the transaction registry is backed up on a secondary server, which provides redundancy to the architecture.

Administratively scoped multicast is used to ensure that both the primary and the backup servers receive the same information about the transactions being made. To prevent unauthorized access from the outside, the networked components of the architecture use private addressing (although the window has also a public address to communicate with the back-office).

These security levels and the flexibility of the final system were considered as the main advantages of the final implementation from the car park managers and employees perspective.

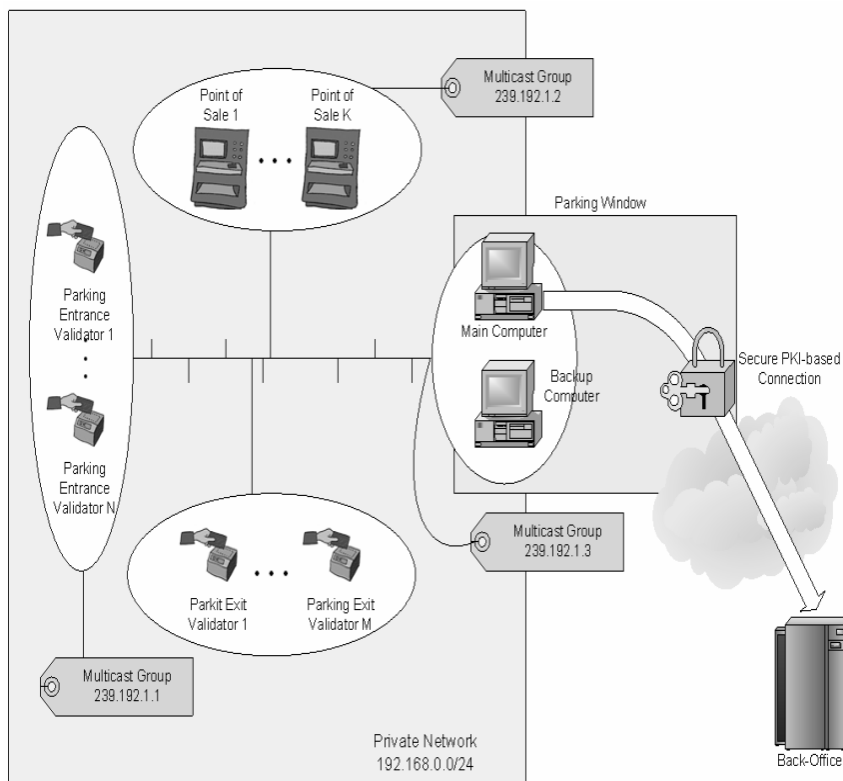


Figure 3 easyTransport car park infrastructure

5. CONCLUSION

In passenger transport environment, e-ticketing systems based on contactless smart cards are being deployed across the world. In particular, in the European Union, there is a special interest at both private and institutional sectors. Bodies of standardization are also involved in the development of normative for these e-ticketing systems: ISO has standardized the technology through ISO 14443 and CEN is working on application standards.

In this context we have proposed a non-proprietary interoperable e-ticketing system called easyTransport based on contactless smart cards and compliant with the existing standards. This system provides the e-ticketing advantages such as quickness, flexibility and security; moreover, it adds additional features: it is non-proprietary, can be implemented in different transport services (enabling intermodality), and allows transport cards to be shared by different operators.

The easyTransport e-ticketing systems are based on a front-office/back-office two-tier infrastructure whose implementation has been successfully tested in a car park.

Future work is oriented to integrate the easyTransport e-ticketing system in a city card multiapplication, which will provide besides e-ticketing several e-services such as e-health or e-government.

REFERENCES

- [1] Pérez, G., "Telemática un nuevo escenario para el transporte automotor," *CEPAL-United Nations*, 2001.
- [2] -, "Open smart card infrastructure for Europe," eEurope Smart Card, *eESC TB9 Public Transport*, 2003, Vol. 1, Part 3.
- [3] -, "CWA 14838-1. Facilitating Smart Card Technology for Electronic Ticketing and Seamless Travel - Part 1: EU Policy and User Requirements," *CEN, European Committee for Standardization*, 2003.
- [4] -, ITSO Integrated Transport Smartcard Organisation, <http://www.itso.org.uk>, 2004.
- [5] -, eESC eEurope Smart Card, <http://eeurope-smartcards.org>, 2004.
- [6] -, "ISO/IEC 14443-1 Identification Cards - Contactless integrated circuit(s) cards Proximity Cards Part 1: Physical characteristics," *ISO International Standards Organization*, 2000.
- [7] -, "ISO/IEC 14443-2 Identification Cards - Contactless integrated circuit(s) cards. Proximity Cards Part 2 Radio frequency power and signal interface," *ISO International Standards Organization*, 2001.
- [8] -, "ISO/IEC 14443-3 Identification Cards - Contactless integrated circuit(s) cards. Proximity Cards Part 3: Initialisation and anti-collision," *ISO International Standards Organization*, 2001.
- [9] -, "ISO/IEC. ISO/IEC 14443-4 Identification Cards - Contactless integrated circuit(s) cards. Proximity Cards Part 4: Transmission protocol," *ISO International Standards Organization*, 2001.
- [10] Yago Sanchez, C. M., "easyTransport: proposal of an e-ticketing model for passenger transport," Master's thesis, Computer Science Faculty, University of Murcia, 2003.
- [11] -, "Bremen Paper. Transport and Car-Sharing: together to the better," *UITP - Union Internationale des Transports Publics*, 2002.
- [12] -, "CWA14838-1. Facilitating Smart Card Technology for Electronic Ticketing and Seamless Travel - Part 2: Deployment of Smart Card Based Interoperable Ticketing Systems," *CEN, European Committee for Standardization*, 2003.
- [13] -, "Specification for Interoperable Electronic Ticketing System," *Norwegian Public Roads Administration*, Preliminary version, 2003.
- [14] Gomez Skarmeta, A. F., Martinez, G., Cánovas, O., Lopez, G., "PKI services for IPv6," *IEEE Internet Computing*, Vol. 7, No. 3, pp. 36-42.

Adaptive Collaboration: The Road Map Leading Telework to a More Advanced and Professional Working Format

Mayumi Hori and Masakazu Ohashi

Abstract - In this chapter, we would like to discuss the Adaptive Collaborative TeleWork as medium which connects individual to individuals, individual to organization, individual to community, individual to society, and individual to the world as large without the restriction of time or location. Information and Communication Technology (ICT) offers new ways to provide more flexibility to the working format. Especially the new method that integrates a number of different systems and applications into one system to enable the Adaptive Collaborative Telework has been generating much attention as it may meet the diverse and growing demands in the future of the Ubiquitous Society. In other words, in our Ubiquitous Society with advanced ICT infrastructure, tacit knowledge, the most difficult yet valuable type of knowledge to be transmitted which is gained through personal experiences and stored within individuals, has better chances to be transferred and stored on the network. Therefore, it is required to build a platform to make tacit knowledge more accessible and sharable in the society. The Next Generation Collaborative Studies Platform in Tokyo have been built since April 2003 Not only the experiment enabled people to share data through telecommunication, it also

demonstrated that utilizing the knowledge management systems in conjunction with the WDM facilitates an enhanced communication structure. In essence, the union of the two systems creates a real-time collaborative research environment by allowing users to share the processes and results of researches between the institutions regardless of their location. The purpose of this paper is to illustrate the road map to lead telework to a more advanced and professional working format by applying the Adaptive Collaboration. Also we would like to examine telework's effectiveness for human resources and how it would efficiently facilitate the Adaptive Collaboration environment, which may characteristically represent the working format for the 21st century.

Index Terms—collaborative system, telework, working environment.

Introduction

The progress and popularization of Information and Communication Technology (ICT) has placed a major impact on working environment. In addition to the development of ICT in Japan, other conspicuous factors contributing to the shifts in working environment are the aging of society, the falling birthrate, globalization of economy, and increasing concern over the global corporate competitiveness. All of these mega-trends are forcing corporations to review and revise their traditional

Manuscript received September 30, 2004.

Mayumi Hori is a professor at Graduate School & Faculty of Business Management, Hakuoh University, Tochigi, Japan Email:m.hori@hakuoh.ac.jp

Masakazu Ohashi is the Dean of Faculty of Policy Studies and a professor at Graduate School & Faculty of Policy Studies Chuo University, Japan. Email:ohashi@fps.chuo-u.ac.jp

employment practices and personnel management systems that had been considered as the strengths of Japanese corporations since the period of high economic growth during 1960's to 1970's.

Combined with an increased presence of women in the workforce, all of these trends such as the emergence of ICT and shifts in social and business environment are forcing corporations to alter their traditional practice that requires women to play only a supporting role. Many corporations are now seriously considering and viewing women workers as integral members of their team. ICT did not merely increase the number of jobs; it has been creating entirely new types of jobs. As an alternative working format, telework brings enormous potentials for revolutionizing traditional work and employment formats.

In recent years, telework, which has not yet commonly practiced in Japan, has attracted a great deal of attention from Japanese, especially from women workers. Looking at the types of telework in Japan, most of them are for outsourcing and are not for unskilled, low-paid work.

We would like to emphasize that the Adaptive Collaborative Telework, which transcends the boundaries of traditional telework, may play a significant role in realizing a true collaborative society. The philosophy of the Adaptive Collaboration is not simply to computerize all of the operations in corporations and administrative agencies, but more so to build a society where diversity is embraced and creativity appreciated, thereby allowing workers to pursue their mission in a coordinated manner.

1. A Proposal for the Adaptive Collaborative Telework and Experimental Pilot Study

1-1. The Concept of AC

Expansion of the versatility of ICT has facilitated many corporations and administrative agencies to merge and collaborate with each other and enabled them to enter into new business schemes. On the other hand, it has become extremely difficult to maintain the competitive advantage in the present market as the

culture of sharing and collaboration prevailed. Furthermore, many corporations and administrative agencies have been urged to meet the diverse needs of the people while improving economic efficiency. In accordance with these situations, we would like to propose the Adaptive Collaboration (AC) as an essential concept for the new paradigm of knowledge integration and collaboration in the Ubiquitous Society.

The Adaptive Collaborative (AC) is defined as a system that efficiently relates, shares, and utilizes data, information, and knowledge in the Ubiquitous Society where the amount of information created grows at an accelerated pace. This AC system would also allow entities of different ontological level to be linked laterally therefore making it easier for people in the organization to appreciate each other's expertise and the know-how, which essentially encourages further development and innovation. Likewise, the system breaks the conventional relationships within and between organizations.

1-2. Experimental Pilot Study on the AC

As we believe the AC is an essential component for the new working style of the 21st century, the Next Generation Collaboration Studies Platform in Tokyo have been built since last April (2003) supported by the Ministry of Post and Telecommunications. In order to examine the effectiveness of collaborative work through telecommunication, an experiment utilizing the knowledge management systems and the Wavelength Division Multiplexing (WDM) was conducted in cooperation with several universities and research institute in Japan(The Report of Society for the Advance Study on e-Society).

The purpose of this pilot study is to examine the effectiveness of applying the XML Web Services into the ACW essentially from the perspective of users. The XML Web Services automatically sorts and relates different application systems and computing resources required to process the queries

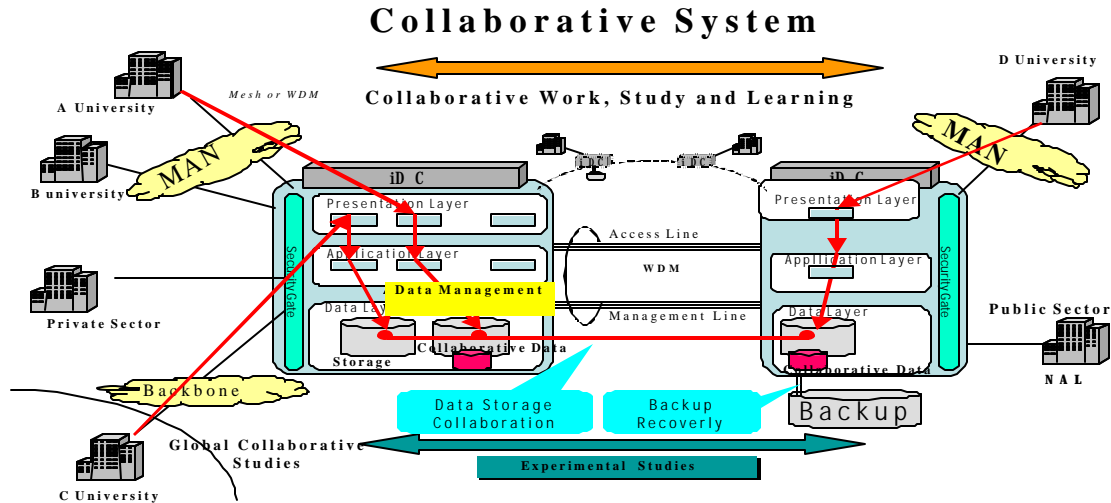


Figure1: Collaborative Work System

on the Internet regardless of the program format. Therefore, we believe that utilizing this technology, users may be able to execute their process without having to worry about unfamiliar technologies and complicated procedures.

As the ACW environment involves complex and entangled elements, it is hard to put a simple desk plan into practice. Therefore, for our pilot study we created a test-bed system simulating the real collaborative work and research environment where several corporations, universities, research institutes conducted experiments on their basic technologies and performance, and studied the efficiency of the system. Our aim was to develop a system, based on the empirical evidences, that organically manages data and storage. The system searches, retrieves, extracts, provides, and utilizes data and data structure that are independent of the application types, and allows users from different location and organization to share the storage which is suited for managing large amount of data.

The experiment proved the Adaptive Collaborative Telework to be very effective. Beyond merely sharing data though telecommunication, the experiment demonstrated that utilizing the knowledge management systems in conjunction with the WDM facilitates an enhanced communication structure. In essence, the union of the two systems creates a real-time collaborative research environment by

allowing users to share the processes and results of researches between the institutions regardless of their location.

This demonstration experiment was conducted in the following settings: 1) a collaborative research and experiment system in an integrated distributed environment, and 2) Adaptive Collaboration for visual media contents production. These trials were performed using the database and applications in the collaborative workspace located in the external storage within the iDC (Internet Data Center). The knowledge management systems enabled users to control the access to files and data stored in the iDC. Not only allowing authorized individuals to customize the settings to obtain access to specific organizations, workgroups, and workplaces, the experiment also performed detailed control over the visibility and invisibility of the intelligent folders, bulletin boards, and emails. An application that intelligently relates the databases

in the external storage to each organization was also introduced, allowing users to voluntarily create and/or alter the settings of the collaborative workspace. The AC is also autonomous-distributed collaborative work system with an iDC centralized, layered datatype sharing model, which enables the clients distributed on the network to autonomously provide services to their consumers (figure1).

2. Results of the Experiment.

2-1. Performance Tests

The series of performance tests clarified the effectiveness of the MAN environment composed of the multiple iDCs connected with WDM. The following are the details of what we found in the performance tests: 1) there was a performance degradation of network access among [FC-SW]-[WDM]-[FC-SW] due to the interconnection problem between WDM and FC-SW as well as the lack of tuning of FC-SW configuration. Improving the tuning of FC-SW configuration is the future assignment. 2) The IP connections such as NFS going through WDM were successfully established within the multiple iDCs without any stress, which correctly enabled data sharing via multiple file servers located in remote areas. 3) Acceleration of the data transfer rate was found to be possible by reducing the server load with the dispersed file servers within the remote iDCs. 4) It is required to improve the MAN environment which ensures the redundancy by doubling the lines running within the multiple iDCs.

2-2. The Potentials of Applying the Experimental Results and Effectiveness of Collaborative Work

The experiment proved the effectiveness of the MAN environment composed of multiple iDCs connected with WDM. Furthermore, there is a good chance of building the enhanced MAN environment which enables the sharing of sever storages and applications resources on the iDCs seamlessly while ensuring the security and reliability. Moreover, although the test bed environment was built on the two iDCs for the experiment, it is highly possible to expand the network operating area. Hence, we can formulate a vision for the future utilization of the system as a collaborative work environment for interregional businesses. Further, as for telework, the main subject of this paper, the experiment proved the Adaptive Collaborative Telework to be very effective.

Beyond merely sharing data though telecommunication, the experiment demonstrated that utilizing the knowledge

management systems in conjunction with the WDM facilitates an enhanced communication structure. In essence, the union of the two systems creates a real-time collaborative research environment by allowing users to share the processes and results of researches between the institutions regardless of their location.

Utilizing the knowledge management system allowed people with different expertise or in different fields –those who do not share the same terminology or methodology-to work collaboratively on the same project. Thus, this experiment proved the effectiveness of collaborative work through telecommunications by not only highlighting the inherent nature of telework as an alternative working style, but also by suggesting its clear potency for efficiently yielding significant results through real-time collaboration.

During the collaboration process on a research project, a variety of information including documents and numeric data was collected in the storage. Information submitted by geographically

-dispersed researchers provided a diverse data structure associated with their writing style, wording, and specific terms they used. The knowledge management system analyzed and clarified the relationships among terms, sentence structure, numeric data, and semantic data. As a result, the system clarified the relationships and commonalities even when the relationships or structures of the data submitted by researchers were different from each other.

3. How Does the AC Promote Business Benefits?

3-1. Effective Use of Human Resources with the AC

This all contributed to making it easier for the researchers to effectively work together, and to compare each other's findings. Consequently, employing this system not only enables collaboration among researchers in the same field, but also promotes a group effort among those in different fields, and allows greater opportunities for obtaining references from a third party. Accordingly, analysis on data structure that informs the relationships between data and words only served to enrich and deepen the collaboration process.

The Adaptive Collaboration generates innovative ways to make effective use of human resources of both in-house and outside staffing with ICT. For in-house teleworkers, a collaborative, group-work environment may help them maintain favorable working conditions as well as achieve good results in their work.

The AC requires one to work toward common goals with other members of the group who have perspectives other than one's own. It can also assist group members in creating a shared new value and understanding. Although collaboration requires harmony, it does not suppress or discard different perspectives and values. Sharing of common goals encourages each member to assume a responsibility and commitment for creating new knowledge that in turn benefits the group as a whole. Therefore, instilling this culture of "sharing" is critical for the successful AC.

3-2. New Services linking Knowledge and Expertise between Public and Private sectors

Another main purpose of introducing the AC is to make a difference in the nature of the work of public officials as it would change the way they carry out their job at home by aggrandizing the definition of telework. In other words, unlike the conventional unilateral services offered by the government, the AC promotes new lateral services that link knowledge and expertise between public and private sectors of different ontological levels. It is also different from the Telework Centers which only change the location of work while the quality of work remains the same. Alternatively, the AC divides and distributes work by its nature and quality such as research and writing papers that can be conducted outside of the conventional office. For example, many software engineers value the working style of the Adaptive Collaborative Telework because the nature of their work allows them to work independently while applying ICT to permit high-quality collaboration so that they can remotely and continuously check their system's integrity with other engineers.

Although telework has been considered merely as a means of outsourcing in Japan, ICT has increased the potential for its expansion, resulting in the conception of the Adaptive Collaborative Telework. Consequently, as the global economy continues to move further from a manufacturing base more towards a service base, the demand for the Adaptive Collaborative Telework will grow. Thus, the critical factor becomes a process of determining how to divide and distribute the work by both its nature and quality.

4 . Changes in Employment Practice in Japan and the AC as a Personnel Strategy

4-1. The Increased Presence of Women in the Workplace

With the fundamental structural change in the Japanese economy and society, a broader diversification of the employment structure has emerged. The most common pattern is the increase of part-time workers and dispatched workers (workers supplied by temporary help agencies). The rapid entry of women into the workforce is ample evidence of a change in occupational attitude among workers. Overall in the 25 years since 1975, women workers have increased by almost 36%, compared to about 18% for men: 18.78 million(1965), 19.53 million(1975), 23.04 million(1985), 24.74 million(1989), 26.56 million(1998). More women in the workplace will inevitably lead to the collapse of the rigid role consciousness—the idea of "woman's place is at home, man's is at work." This attitude has long been a central value in male-female relations in Japan. Today, with advanced levels of education, more women are seeking positions to exercise their special abilities, which would also entail them the independence and allow them to continue working without dropping-out.

Moreover, traditionally Japanese women staying at home engaging only in household chores and childrearing had a limited social circle. That is, they interacted mostly with family members and people in the same community. People traditionally grow personal relationship and strong bounds of trust through face-to-face communication. Likewise in business field, most valuable knowledge transmitted and shared among

group is called “water-fountain knowledge.” Japanese business people very much value this type of knowledge and often grow personal level relationship outside of office through dinner and drinks. Therefore, for teleworkers to instill the culture of sharing, innovative and ingenious contrivances utilizing ICT are necessary.

4-2.The Possibilities of Telework for Women

Telework brings women not only the opportunity to exercise their expertise as a professional, but also the interactions with others that were once impossible or at least very difficult. This social interactions itself encourage women at home to maintain better mental health. As a result, telework has a potential to entail Japanese women an independency as a professional which leads a better self-confidence, better mental health through more active social interactions, and finding a better balance between traditional chores and work therefore improving the quality of their lives.

There are many women, including the elderly and physically-challenged, whose abilities are not fully exercised in our society. The AC would apparently promote employment opportunities for them so that they can make the most of their abilities.

Furthermore, till today after the high economic growth period from 1960's to 1970's, Japanese corporations and administrative agencies had built their organizational foundation based on the entrenched practices of the personnel and wage system such as permanent employment, seniority-based pay and promotion, and uniform wage system.

However, today's economic environment has shown us that these entrenched practices in personnel and wage systems are no longer effective but hinder the sustainable and continual development of organizations, and even endanger their survival. Along with a number of profound transformations in the society such as globalization, these Japanese traditional practices that only seek for stability but do not require competition among employees need drastic changes. That is, organizations are required to instill more competitive strategies which enable individual employees to maximize their

potential, share and utilize knowledge and competence that organization possesses, collaborate with other organizations and research institutes to share the know-how and expertise.

Organizations are also required to operate the “core-competence management” to concentrate their management resources on their strengths and efficiently utilize outsourcing. The AC facilitates the effective utilization of human resources by enabling the distributed operation and encouraging women workers to exercise their full potential. Therefore, the AC is one of the essential personnel strategies that powerfully assist organizations.

Conclusion

By experimenting with the test, we could perform technical examinations for storage operation, the switching technology for network failure, and knowledge management for collaborative work environment. Additionally, we could also examine the network and the collaboration process of visual image contents creation from the perspectives of both users and providers in the collaborative work environment.

With a conventional business model, individual organization builds an information system that only fits the organization. However, in the era that requires innovative management, organizations are strongly urged to collaborate and cooperate with a variety of organizations to create more value and competitiveness as seen in the XML Web Services, rather than trying to compete against each other. That is, organizations will no longer need to stick to specific data or applications but to be flexible enough to adopt appropriate objects according to each business model and project. Consequently, the AC, we believe, can bridge between the traditional top-down, hierarchical organization and the horizontal business models.

The AC also acts as a contact point between analogue organization and digitalized system through innovative Internet businesses based on the network such as those operated by the iDC. This digitalized system enabling the AC is the XMLWeb

Services. Therefore, the AC is composed of the two main components: iDC and the XML Web services.

The AC is a promising business model and is effective for collaboration for utilizing knowledge and expertise of autonomously distributed clients. It is especially useful when launching projects that transcend the boundaries of organizations. Hence, we believe that the AC is essential for the advanced and sophisticated application of distributed work such as telework we discussed in this paper.

References

- [1] Masakazu Ohashi edi, The Report of the Advanced Studies for the Social Capital of e-Society, The Society of the Basis for the eCommunity,2003,The Report of Society for the Advance Study on e-Society, The Society of the Basis for e-Community, Tokyo,Japan,2004
- [2] Masakazu Ohashi, Kaori Sasaki & Mayumi Hori, On the study of knowledge Structuralizing and Adaptive Processes Based on Project Based Learning, The Annual Bulletin of The Institute of Policy and Cultural Studies, Vol.7,Chuo University,Tokyo,Japan,2004
- [3] Masakazu Ohashi, Public iDC and c-Society, Kogaku Tosho, Tokyo, Japan, 2003.
- [4] Masakazu Ohashi , Time Business, NTT Publication, Tokyo, Japan,2003.
- [5] Masakazu Ohashi, edi., Knowledge-Based Collaborative Work, Report of Supplementary Budget Project of the Ministry of Post and Telecommunications,2003
- [6] Applying XML Web Services into Health Care Management, Proceedings of the Thirty-Eighth Annual Hawaii International Conference on System Sciences,Hawaii,2005
- [7] Mayumi Hori & Masakazu Ohashi, Adaptive Collaborative Telework in Public Administration, eChallenge e2004, Vienna, Austria,2004
- [8] Mayumi Hori & Masakazu Ohashi, Telework Changes Working Style for Japanese Women, Proceeding of AWEEB2004, International Workshop on Advanced Web Engineering for EBusiness, Frankfurt, Germany, 2004.
- [9] Mayumi Hori& Masakazu Ohashi, Telework and Mental Health-Collaborative Work to Maintain and Manage the Mental Health, Proceedings of the Thirty-Seventh Annual Hawaii International Conference on System Sciences, Hawaii, 2004
- [10] Mayumi Hori, Society of Telework and Working for Women, Publishers of Chuo University, Tokyo, Japan, 2003.

[11] Mayumi Hori, The Development of IT and a New Work Format for Women in Japan, Proceedings of the world 2001, Edi. Reima Suomi, Ministry of Labour Finland, 2001.

[12] Mayumi Hori & Masakazu Ohashi, Information Technology and The Possibility of Women's Work: A New Work Format for Women in Japan, The 6th International ITF Workshop and Business Conference "Working in the New Economy", Amsterdam, 2001.

Dr. Mayumi Hori

Mayumi Hori received the B.E. and M.E. degrees in economics from Rikkyo University, Japan in 1976,1993, respectively and Dr. Policy Studies degree from Chuo University, Japan, in 2003.Now she is a professor at Graduate School & Faculty of Business Management, Hakuoh University,Tochigi,Japan. She is a director of the Web Services Initiatives in Japan, which synthesize and integrate the needs of the users, aiming to build the systems from the perspectives of users. And she is the vice-chairperson of The Institute for the Studies of an Information Society,a member of The Society of the basis for e-Community,and The Society of the Basis for the Information Community in Japan.
Email:m.hori@hakuoh.ac.jp

Dr.Masakazu Ohashi

Masakazu Ohashi received the B.E.,B.S.,M.E. and Dr.Eng.degrees from Chuo University,Japan, in 1976,1978,1980,1983, respectively. Now he is the Dean of Faculty of Policy Studies and a professor at Graduate School & Faculty of Policy Studies Chuo University, Japan. Dr.Ohashi is the chairperson of Time Business Forum, iDC Initiatives,The Institute for the Studies of an Information Society, The Society of the basis for e-Community,The Society of the Basis for the Information Community in Japan and a member of UN/CEFACT TBG6.
Email:ohashi@fps.chuo-u.ac.jp

Molecular Processes as a Basis for Autonomous Networking

Bettina Krüger and Falko Dressler

Abstract— *Autonomous networking has become the buzzword for attempts of building high-scalable network architectures, which are self-organizing, self-maintaining and self-healing. Few of these approaches were successful and none has shown to provide all the promised functions. We try to study the processes in computer networks using molecular processes as the paradigm. This novel approach shows many similarities between computer networking and cellular mechanisms. In this paper, we focus on the area of network security as one research area with high demand for high-scalable mechanisms providing the needed functionality. After identifying similarities between nature and technology, we discuss potential research domains, which are high potentials for learning directly from molecular biology using the example of security threats in communication networks. We see the proposed mechanism as a generic approach for autonomous networking. The countermeasures against attacks in computer networks are only a special example to introduce the mechanisms.*

Index Terms— *Autonomous Networking, Bio-Inspired Networking, Network Modeling, Next Generation Networks*

1. INTRODUCTION

WE investigated the possibilities to utilize the infinite experience of the nature to address questions in computer science, focused on problems in networking. Among other things, this approach was motivated by a book from Nobel Laureate Prof. Manfred Eigen [7]. Unlike most

activities in bioinformatics, where computerized methods are employed to study natural processes in more detail, we discovered that at least some of our nowadays problems might be easier to understand and even to be solved if we directly learn from natural mechanisms. In this paper, we focus on mechanisms known from molecular biology, which we can adopt to improve internet technology.

During the last couple of years, great progress was made to make computer networks more stable, more efficient, and more secure. Nevertheless, we also experience that there are still many open issues, especially in terms of network security. Distributed denial of service (DDoS) attacks, worms, and viruses are getting more aggressive and much harder to prevent [9]. Autonomous networking should help to solve these problems. We address these issues by applying mechanisms learned from natural processes and show potential solutions.

In simple configurations of low-speed networks, well-known mechanisms can be employed to examine all the network traffic, to filter unknown or suspicious data packets, and to program firewall rules blocking typical attacks [3, 11]. Unfortunately, these mechanisms fail in high-speed backbone networks due to restrictions in CPU capacity and free memory for processes and queues. Therefore, it is not possible to run all kinds of required monitor processes at once directly on the network components.

Our idea is to limit the number of active processes to those which are really required in the current situation, e.g. if a particular worm is being distributed, only a prevention scheme for this event is needed. The solution is provided by achievements of molecular biology. In every single cell of an

Manuscript received August 10, 2004; revised September 7, 2004.

B. Krüger is with the Medical Clinic IV, Nephrology, University of Erlangen-Nuremberg, Germany.

F. Dressler (contact) is with the Department for Computer Sciences, University of Erlangen-Nuremberg, Germany (e-mail: bettina.krueger@fd42.de; dressler@informatik.uni-erlangen.de).

organism, the program for all reactions and mechanism is coded in the DNA and is activated exactly in the situation when it is needed. Every signal from the extracellular environment is recognized specifically and results in a specific signal transduction cascade. Signal transduction normally initiates the translation of the necessary genes which finally leads to a cellular answer, e.g. to maintain the cellular function, but also to react to pathological situations, e.g. any kind of inflammation, etc.

The same procedure seems to be adequate for reactions in computer networks. If we are able to describe solutions for typical problems in form of processes and their course of actions, we are able to build flexible network components with low resource requirements and a high efficiency in terms of network operations. Fortunately, we can employ methods from software engineering for such descriptions. Monitor components act as the receptors and try to identify the behavior of the network. Due to the availability of sampling methodologies and statistical methods, such monitoring can be done even in high-speed network environments [6, 13]. We already addressed similar scalability issues in other areas such as network measurements [5].

Based on the recognized behavior of the network, new processes can be created and activated using a description of the available mechanisms. Therefore, only these processes exist in memory and are activated which are required in the current situation.

We see the proposed mechanisms as a generic approach for autonomous networking. The detection of security threats and the activation of countermeasures are only special examples to introduce the mechanisms.

In this paper, we show the similarities between the signaling mechanisms in cellular systems and networking entities responsible for packet forwarding, intrusion detection, and firewalling.

The rest of the paper is organized as follows: in section 2, the cellular mechanisms are described followed by a description of a general networking architecture in section 3. Possible research

issues resulting from an analysis of the similarities between both the cellular systems and computer networks are discussed in section 4. A section describing related work and some conclusions complete this paper.

2. CELLULAR MECHANISMS

All organisms share one common information system which is mostly the DNA and which codes for the organization of the whole organism. This organization is a highly regulated process from the single cell up to complex organs of the body. The hierarchy in the organism is very high. Every process, e.g. movement, metabolism, communication, etc. is organized by interactions of several organs. Organs represent an assembly of one or more tissues, which fulfill a common function. One tissue is build by different cell types. One cell type consists of identical cells, which are associated and communicate with each other to fulfill a common function within the tissue. Single cells communicate with each other as well as with cells of other tissues by sending signals to which the target cells respond by specific gene expression. In this way, a signal can be carried out and influences the function of higher units such as the organs.

In this paper we want to focus on mechanisms how cells interact with their environment in general and, secondly, on intracellular processes which are initiated by extracellular signals and result in the specific cellular response out of the pool of information given by the DNA.

D. Cellular Signaling

The functionality of an eukaryotic cell relies on the complex network of biochemical processes. Within these processes, single reactions take place in a coordinated fashion. They can take place simultaneously and successively. Thus, these processes must be highly regulated and controlled. This also means that these mechanisms are very specific for the given result.

The main goals of cellular processes are to regulate the intracellular metabolism and to communicate with their environment.

Physical or chemical attractions from the environment are signals for the cell to change intracellular processes. Chemical attractions can be low-molecular metabolites, hormones, or ions. They can be sent by other cells of the same tissue or by cells from other tissues. Furthermore, physical parameters such as heat, pressure, or electrical signals can induce cellular reactions.

E. Intercellular Signaling

Communication between cells can occur by different processes. First, cells can release soluble molecules such as hormones etc. that are transported via the blood (long distance in the organism, e.g. hormones) to the target cell. Other soluble factors are released into the extracellular space to reach the neighboring cells in a short distance (Fig 1 A).

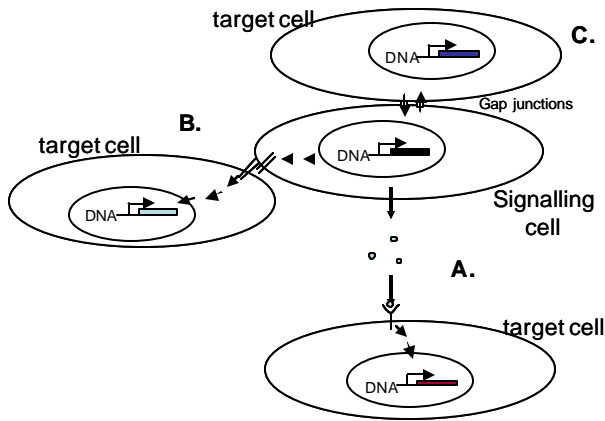


Fig 1. Intercellular communication. One cell can communicate with the neighboring cells via A. soluble factors, B. receptor interactions, C. direct contact, e.g. "gap junctions". The target cells react on the current stimulus by gene transcription.

These molecules are recognized by the target cells and induce a specific biochemical answer. By the release of soluble factors many cells can be activated simultaneously which results in a coordinated reaction of the organ or organism.

Secondly, cells can also communicate via cell surface molecules. In this process, a surface molecule of one cell or even a soluble molecule, which is released by one cell, directly binds to a specific receptor molecule on another cell (Fig 1 B). Thirdly,

communication between cells occurs via direct connections between two neighboring cells, which allow a direct exchange of metabolites (Fig 1 C).

F. Intracellular Communication

In either case, the signal from the extracellular source is transferred through the cell membrane. Inside of the target cell, complex signaling cascades are involved in the information transfer (signal transduction), which finally result in gene expression or an alteration in enzyme activity and, therefore, define the cellular response.

Because of the great variety of signal transduction pathways, only one example for receptor-mediated signal transduction is presented here which might have a great relevance for computer networks: The MAPK signaling pathway is a major pathway in eukaryotic cells, which is activated by different types of receptors, e.g. receptor-tyrosine-kinases or G-protein-coupled receptors [15].

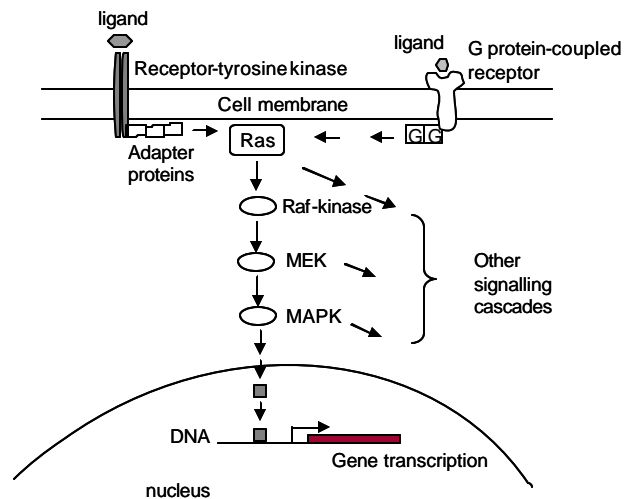


Fig 2. Intracellular signal transduction, shown on the example of MAPKinase pathway. Different ligands can activate different receptor types which result in MAP kinase activation.

Upon binding of the signaling factor (ligand), the receptor is activated. This activation is mostly attended by a phosphorylation or conformational change of the receptor, which make it possible for adapter proteins inside of the cell to bind to

the receptor. Similarly, these adapter proteins are activated.

The signal is carried on to a signaling molecule called Ras. As shown in Fig 2, the activation of different types of receptors comes together in the activation of Ras. The signal is carried on by protein kinases. Protein kinases phosphorylate other proteins. Phosphorylated protein kinases are able to convey the phosphorylation to the next kinase. Finally, a transcription factor is activated by this signaling cascade which move into the cell nucleus and bind to the DNA.

The binding of a specific transcription factor to a specific binding site on the DNA result in gene transcription, which finally induce a specific cellular response. This example shows only one straight-forward signaling pathway. For example, the same protein kinases Ras, MEK and MAPK are also involved in other signaling cascades. Thus, signaling cascades are often highly networked, but at least result in a very specific gene transcription and, therefore, result in a very specific cellular response [12].

3. NETWORK ARCHITECTURE

We describe the relevant elements of general network architectures starting at a high abstraction level, the structure of the global Internet, and dig deeper until we reach the internal operation of a single networking node. The focus is on the characterization of the information and data paths in the network concerning packet forwarding, monitoring, and firewalling as well as management functions including intrusion detection mechanisms.

G. Internetworking Structure

From a high-level point of view, an Internet consists of a multitude of individual networks. Each of these networks, which are called domains, hides its internal structure from the outside world. Such a scenario is shown in Fig 3. In reality, all these domains are managed from different service providers. Thus, the interaction between the domains is limited, due to unequal configurations and mistrust.

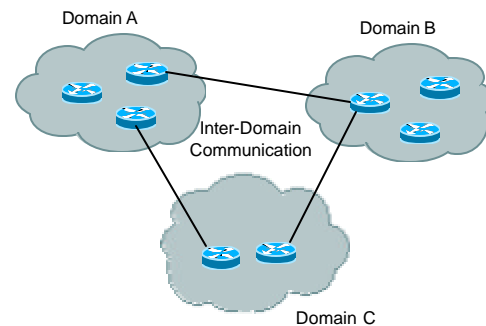


Fig 3. Internetworking Structure. Large internets consist of multiple network domains

Looking at security issues, this is a serious problem. Denial of service attacks can only be detected if knowledge about the network traffic in multiple domains is available. Inter-domain signaling and interaction is required, but there are no or, at least, few of such mechanisms developed or even deployed. The interaction between the domains is currently restricted to the exchange of routing information.

H. Intra-Domain Mechanisms

If we look a little closer and examine the components of single domains (still with the focus on network security), there are, among other entities, routers, monitors, firewalls, and intrusion detection systems. These entities and their data and signaling paths are shown in Fig 4.

It is important to understand that this is only a logical point of view. In reality, some or even all of these functions can be implemented in a single box.

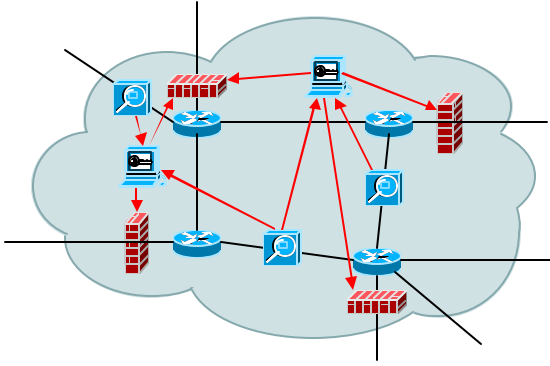


Fig 4. Communication in a single network. Typical entities are routers (packet forwarding), monitoring probes, firewalls, and controlling IDS systems

The routers are responsible for the raw packet forwarding. Sometimes, statistical information is sent to the management or IDS systems. Network monitors are employed to analyze the traffic in the network in order to detect suspicious data flows or unusual network behavior. Based on such information, examinations that are more precise can be initiated, e.g. based on attack signatures, and countermeasures become possible. The intrusion detection systems signal signatures of violent hosts to the firewall systems. At this place, rules are installed which prevent any further attacks from the identified intruders.

Today, there are some intrusion detection systems available, a few of them as open source software. Unfortunately, they do not work in highly distributed configurations and their interaction is limited. Research needs to be done to improve these mechanisms, especially the interoperation between different kinds of networking components.

I. Single Node

In the last step, we zoom into a single network node and analyze the primary components in order to compare them later with corresponding parts in cellular environments. Fig 5 provides a schematic overview. Shown are only those parts, which we think are directly improvable by studying cellular mechanisms.

First, each network node has interfaces connecting it to other network nodes. Logically, each interface has two sub-interfaces. One for inter-node

communication and one for the raw data transfer. Internally, each interface consists of input and output queues with fixed, but configurable size and behavior, e.g. the selected queuing algorithm.

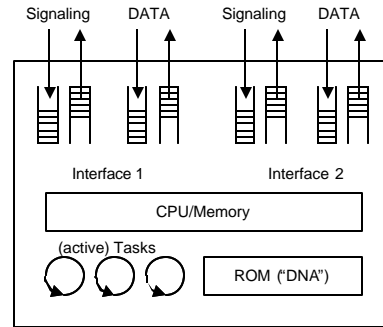


Fig 5. Schematic overview of the resources in a single network node those are required for processing and signaling.

Secondly, limited resources such as CPU capacity and memory are available for sharing among numerous tasks. Some handle the routing tables and algorithms, others coordinate the packet forwarding and still others perform security checks and inter-node communication. Tasks “exist” per default, i.e. they are started when the system is booted. Other tasks can be created on demand allowing a high flexibility.

In terms of network security, it would be desirable to have tasks searching for security threats while at the same time forwarding data packets at very high data rates. Typically, this is not possible because the main resources such as CPU capacity and memory are very limited.

Therefore, the primary question is which task to run at which time. Furthermore, not all tasks and algorithms can exist in memory waiting to get started. Other solutions are required when and how new tasks are created from saved high-level descriptions of their composition and behavior.

4. SIMILARITY ANALYSIS AND RESEARCH ISSUES

We discuss potential research issues in this section by analyzing similarities in both worlds.

J. Composition of the Components

A first obvious result is the cognition that cellular systems consist of similar units (single cells), which react on changes in the environment to survive, and transfer information to similar units, respectively. They are directly comparable with network nodes.

Furthermore, the interaction of different cell types in a tissue is comparable to networking domains. If we zoom into both systems, the same similarities appear as tissues are made of different cells, each focused on its very specific purpose. Network nodes, on the other hand, are also very specific for their particular tasks such as packet forwarding or network monitoring.

Currently, we only assume the possibility that some of the cellular information transfers might be useful and can be copied to advance the specificity of operations in network technology.

In computer science, great effort was made in simulating mechanisms based on constructive building blocks. We intend to create object-oriented programming modules representing network nodes, their internal behavior, and their external communication mechanisms. This is the basis for representative simulations of the new mechanisms. The novel approach is to adopt and to transfer mechanisms how organs, tissues, and single cells communicate to computer networks to enhance their specificity and velocity of responses in every situation.

K. External Signaling Pathways

The primary difference in the external signaling pathways between communication networks and cellular systems is the presence of data traffic in computer networks sharing the same infrastructure that is used for inter-node communication. Besides this fact, the concepts of inter-node communication and cellular signaling are similar. A typical problem in communication networks is the scalability of mechanisms, e.g. routing or configuration tasks.

We believe that there are high potentials in examining the mechanisms for signal transduction in cellular systems and applying them to communication systems.

The most impressive issues are given by the autonomic behavior of individual cells by reacting to signals from the direct environment on the one hand, and refer the answer to other tissues or "other networks". One example is given here: During inflammation, the cells of the affected tissue send signals to recruit cells of the immune system from the blood which means that a local problem in the body is referred to those cells which are able to initiate a response, namely the cells of the immune system. The signals that are sent by the cells of the inflamed tissue reach the target cells, the blood vessel cells (endothelial cells) outside of the cell. The signal which is mostly a small protein (in this specific example it is called IL-1 β), very specifically recognizes one receptor on the surface of the target cell. This recognition results in the gene transcription and translation of a specific protein, called E-selectin. Furthermore, this E-selectin protein is one of the first proteins that are involved in the recruitment of cells of the immune system in the blood via the endothelial cells which are found on the luminal site of all blood vessels and represent the connection between the blood and the surrounding tissue on their way to the center of the current inflammation. This example shows a two step process in which a specific signal is sent from the signaling cell (in the inflamed tissue) to a first target cell (the endothelial cells) in the environment. Secondly, these target cells recognize the problem and are able find an answer to recruit help from far away (the cells of the immune system).

Similar mechanisms help to solve problems in communication networks. An intrusion detection system is working all the time on analyzing network traffic. If suspicious traffic is detected, the system might, as usual today, contact its management system or an attached firewall to enforce countermeasures. This methodology has many drawbacks, e.g. the necessity of the possible and undisturbed communication between the IDS and the management/ firewall node and the dependency on the proper function and availability of the management system.

Looking at network security, one can

imagine a similar two step information transfer as described for the biological systems.

L. Internal Behavior of Individual Components

In this section, we focus on one single cell. This cell receives information from outside and processes this information inside and gene activation is the answer. But gene activation as a result of the intracellular signaling cascade is often influenced by the presence of intracellular inhibitor and effector molecules which are, in turn, regulated by other control mechanisms in the cell [2, 8]. These control mechanisms appropriate the specificity of the gene translation and, therefore, the cellular answer. This means that a signal transduction cascade is often not a straight-line cascade, but can be networked with other signaling cascades and all these processes succeed highly coordinated and regulated. A lot of relationships between signaling cascades are known already and it turned out to be the importance of timing and combination how they work together to reach a very specific answer.

The same autonomic behavior applied to network nodes would help us making communication networks more efficient. There are a number of topics to discuss. Much pressure is on high scalable and effective intrusion detection with automatically involved countermeasures. We are working on such mechanisms allowing monitoring, analyzing, and processing of high-speed traffic. Processing power of individual nodes is first allocated for the supervision of ongoing attacks and the effectiveness of taken countermeasures. Then, the remaining capacities are directed to the analysis of unknown traffic and the detection of new attack traffic. Considering this platform, a lot of similarities between the biological processes of signal transduction and communication networks can be expected. The knowledge on the specificity of receptor-ligand interactions as well as the specificity of intracellular signaling pathways and the cellular answer that results from the particular pathway might give impressions how the mechanisms in computer networks can be improved.

5. RELATED WORK

The first approaches to identify mechanisms in nature to address technological problems, especially in computer science date back to the mid 1970s. Later, the human immune system was used for investigations on computer viruses and their detection [4]. This is still the best-known example of so-called bio-inspired computing [10].

The group of Prof. Suda is investigating an architecture, which they named the bio-networking architecture [16]. The basis of their project is a middleware platform [14], which aims to incorporate mechanisms known from swarms of bees and ant colonies in order to achieve a high level of autonomy and reliability.

Especially in the field of computer networking, we believe that there is still much work to be done employing bio-inspired mechanisms. The relevance is also demonstrated, for example, by the current research-funding programs of the NSF [1] and the European Union.

Other research areas in computer science and engineering do also profit from bio-inspired research. Examples are pattern recognition, robot control, and cognitive sciences.

6. CONCLUSIONS AND FURTHER WORK

In conclusion it can be said that we were able to show many similarities in the signaling pathways known from molecular biology and from computer networks.

The attempt to adopt mechanisms from the basis of the organism, the single cell, looks promising for further studies because the cell has to maintain its own assembly and thus react specifically on changes of their direct environment without the necessity to know about all interactions in the organism. The efficient communication system of cells can be copied to computer networks. The next step will be to analyze processes in network entities which might react in a more "cellular" or "biological" manner to improve their effectiveness. Our primary goals are the development of self-configuring, self-maintaining, and self-healing network systems proving a better protection against security threats.

In conclusion, actual problems of computer networks have to be discussed on the model that each part in the network corresponds to an adequate structure in the organism. Looking at the defined adequate structure in the organism, the mechanisms of interaction with other parts of the system can be analyzed in detail and assigned for computer networks. We are going on in our research activities creating models representing the signaling pathways in communication networks and allowing us to incorporate the studied mechanisms from biology. From such a model, simulative analyses can be done to show the advantages of our new algorithms.

REFERENCES

- [1] "National Science Foundation: Summary of FY 2003 Budget Request to Congress," NSF, 2003.
- [2] N. Abdel Wahab and R. M. Mason, "Connective tissue growth factor and renal diseases: some answers, more questions," *Curr Opin Nephrol Hypertens*, vol. 13, pp. 53-8, Jan 2004.
- [3] R. K. C. Chang, "Defending against Flooding-Based Distributed Denial-of-Service Attacks: A Tutorial," *IEEE Communications Magazine*, vol. 10, pp. 42-51, October 2002.
- [4] P. D'haeseleer, S. Forrest, and P. Helman, "An Immunological Approach to Change Detection: Algorithms, Analysis and Implications," Proceedings of IEEE Symposium on Security and Privacy, Oakland, CA, USA, May 1996, pp. 110-119.
- [5] F. Dressler, "A Scalable Environment for Quality of Service Measurements in the Internet," Proceedings of 2nd IASTED International Conference on Communications, Internet, and Information Technology (CIIT 2003), Scottsdale, AZ, USA, November 2003, pp. 307-312.
- [6] N. Duffield, "A Framework for Packet Selection and Reporting," Internet-Draft, draft-ietf-psamp-framework-08.txt, September 2004.
- [7] M. Eigen and R. Winkler, *Laws of the Game*, Princeton Univ Pr, 1993.
- [8] A. Hahn, J. Heusinger-Ribeiro, T. Lanz, S. Zenkel, and M. Goppelt-Struebe, "Induction of connective tissue growth factor by activation of heptahelical receptors. Modulation by Rho proteins and the actin cytoskeleton," *J Biol Chem*, vol. 275, pp. 37429-35, Dec 1 2000.
- [9] M. Handley, "Internet Denial of Service Considerations," draft-iab-dos-00.txt, January 2004.
- [10] S. Hofmeyer, "Who says biology need be destiny?," in *CNET News.com*, April 13, 2004 2004.
- [11] R. Kemmerer and G. Vigna, "Intrusion Detection: A Brief History and Overview," *IEEE Computer*, pp. 27-30, April 2002.
- [12] T. Pawson, "Protein modules and signalling networks," *Nature*, vol. 373, pp. 573-80, Feb 16 1995.
- [13] J. Quittek, T. Zseby, B. Claise, and S. Zander, "Requirements for IP Flow Information Export," draft-ietf-ipfix-reqs-16.txt, June 2004.
- [14] J. Suzuki and T. Suda, "Design and Implementation of a Scalable Infrastructure for Autonomous Adaptive Agents," Proceedings of 15th IASTED International Conference on Parallel and Distributed Computing and Systems, Marina del Rey, CA, USA, November 2003.
- [15] T. van Biesen, B. E. Hawes, D. K. Luttrell, K. M. Krueger, K. Touhara, E. Porfiri, M. Sakaue, L. M. Luttrell, and R. J. Lefkowitz, "Receptor-tyrosine-kinase- and G beta gamma-mediated MAP kinase activation by a common signalling pathway," *Nature*, vol. 376, pp. 781-4, Aug 31 1995.
- [16] M. Wang and T. Suda, "The Bio-Networking Architecture: A Biologically Inspired Approach to the Design of Scalable, Adaptive, and Survivable/Avilable Network Applications," Proceedings of 1st IEEE Symposium on Applications and the Internet (SAINT), San Diego, CA, USA, January 2001.

Bettina Krüger (bettina.krueger@fd42.de) received her M.Sc. in biology from the University of Goettingen, Germany and her Ph.D. from the University of Erlangen, Germany in 1998 and 2003, respectively. She is currently a PostDoc at the Medial Clinic IV, Nephrology at the University of Erlangen-Nuremberg, Germany.

Falko Dressler (dressler@informatik.uni-erlangen.de) received his M.Sc. and Ph.D. in computer science from the University of Erlangen in 1998 and 2003, respectively. Currently, he is an assistant professor of computer science at the University of Erlangen-Nuremberg, Germany.

H1B Visa and the Policy Vignette for 21st Century Dislocated Workers in the United States

Henry O'Lawrence

Abstract - An H1-B visa is considered an option for skilled professionals who are not U.S. Citizens or green card holders to get permission to work in the United States; a situation that is constantly affecting both individuals (called dislocated workers) and the U.S. companies. This reflects a huge gap between supply and demand in the American labor market in the area of high tech industry. This article discussed both the issues facing the H1B visa policy and the solution and the alternative solution. The entire issue of the H1B visa rests on one of control. Who will control the future of the US labor market? The high tech industry, like any other industry is in need of high tech workers, but does not want to pay an unprofitable price for them, or labor unions, who have historically exerted much leverage over the manufacturing industry in America, but now are losing in their struggle to maintain historically high U.S. wage levels in a global market.

Index Terms— economy, labor market, workforce

1. INTRODUCTION

Many economic studies have detailed the negative earnings effects of a job displacement in the United States. As a result, many dislocated workers suffer annually at the range from 20 to 40 percent in earnings losses (Stevens, 2002). Also, income stagnation and wealth inequality has

become persistent characteristics of the economy. Many had lost their jobs while many with jobs continue to feel a growing sense of insecurity, believing that they may be the next to lose their jobs regardless of how long they have held them.

Many permanent and full-time jobs have been replaced by contingent workforce, and major problems was attributed to be: (1) increased foreign trade and business competition; (2) technological changes such as the influx of computers into work places; (3) declining membership in labor unions; (4) shrinking employment in manufacturing; (5) corporate consolidation, new systems for organizing work; and (6) immigration.

Major theory regarding contingency workforce was further discussed that the growth in the number of temporary help jobs has allowed firms to respond to market stimuli on both the demand and supply side. According to Lane (2003) on the demand side, firms have developed alternative work arrangements because technological advances and the consequent job specialization that make it possible for firms to hire employees for specialized tasks rather than relying on employees with broad, generalized job descriptions. On the supply side, the increased number of women and young people in the workforce has increased the total number of workers in the labor force available for flexible employment. This paper discussed the issue facing our workforce and the policy issues behind the obstacles against the dislocated workforce in America.

Manuscript received August 19, 2004.

Henry O'Lawrence is with California State University, Long Beach USA (e-mail: holawren@csulb.edu).

2. *THE ISSUE FACING AMERICAN WORKFORCE*

An H1-B visa is considered an option for skilled professionals who are not U.S. Citizens or green card holders to get permission to work in the United States; a situation that is constantly affecting both individuals (called dislocated workers) and the U.S. companies. This reflects a huge gap between supply and demand in the American labor market in the area of high tech industry. According to the associated press (Oct. 3, 2000), the Senate passed a legislation that will provide nearly another 600,000 new visa over the next three years for foreigners who want to work for U.S. Industry. The disquiet here is either this would really keep the economy strong or hard with the problems of illegal immigrants or political refugees that are demanding to remain in the United States.

Are we to believe that United State dependent on high tech workers from another country or lack competitiveness with other nations unless those high tech worker allowed to work in U.S., or does that mean that United States institutions were not capable of training our youths, dislocated worker who will rather preferred to be less paid in salary than not having an income to feed their family? The country is currently divided among technology companies who contended that 300,000 jobs are going unfilled for lack of qualified workers; while labor unions are arguing that the companies want more immigrants to put downward pressure on the wages of Americans holding the same Jobs.

An agreement with Rep. Lamar Smith (R-Texas) chief sponsor of the Judiciary panel's measure was well taken that H1-B visa increase in the number of foreign high-tech workers without protecting American workers. When we look at the proposed number, it does not include foreign students or others who are well qualified and are already in the United States. Unfortunately, these groups of people are considered to be in the job market already, while by, honestly speaking, they are not.

Understanding that the fee for this visa currently is \$500, and raising it to \$1,000 allow the government to raise within three years about \$450 million from this kind of

visas and then use the money to provide scholarship to 60,000 U.S. students in this fields that currently need foreign skilled power. It is expected that in this way, U.S. students will in future fill the current vacuum, and the money will also go to the training of 150,000 U.S. workers to fill such vacancies in the future. It seemed like is the love of money.

- *POLICY ISSUES*

The U.S. economy is in a constant state of change, resulting in the dislocation of large numbers of workers, even during times of economic expansion and recovery from recession. The United States Bureau of Labor Statistics estimates that "approximately 10.8 million workers lost their jobs because of business closures and permanent layoffs in the five years between 1981-1986 (Blong & Shallots, 1990, p. 28). The most important issues addressed were based on three questions: (1) how many workers have been dislocated by business closures and permanent layoffs? (2) What are the reasons for closures and layoffs? (3) to what extent does the federal government assist dislocated workers?

According to Bureau of Labor Statistics data estimation on average, there are about 2.3 million workers dislocated annually between 1979 and 1984. Despite the economic recovery between 1983 and 1984, over 16,000 establishments with 50 or more employees experienced a closure or permanent layoff affecting about 1.4 million workers in this two-year period. The reason for closures and layoff was attributed to problems of competition in the marketplace, with some of the competition come from imports.

About 70 percent of the businesses indicated that reduced product demand and/or increased competition were major factors influencing their decision to close or lay off workers. Over 35 percent of the closures and layoffs for establishments with 100 or more employees were in industries in which the Department of Labor has certified workers as being adversely impacted by international trade. These closures and layoffs affected about 436,000 workers.

To what extent do employers assist dislocated workers? To reemploy dislocated workers has been or much concern and the

important major factors affecting their reemployment include the availability of both financial and placement assistance. Most establishments that experienced a closure or a permanent layoff provided some assistance to their dislocated workers. Slightly more than half of the businesses experiencing a closure or a permanent layoff offered their workers severance pay, about a third offered placement assistance and 3 percent offered occupational training. Finally, 37 percent of the establishments offered their employees both financial and placement assistance; however, 36 percent offered neither.

The dislocated worker experiences both social and psychological problems that are a result of separation from long-term employment (Kates, Grieff, & Hagen, 1990). The frustrations of entering the labor market unprepared and the humiliation of unsuccessful job search are common anxieties experienced by the displaced worker. Classroom anxiety and reluctance to participate in educational opportunities are also constraints for dislocated workers. For this special population, "...education is not their priority; getting another job is their most critical need" (Blong & Shultz, 1990, p. 28). The trauma of job loss and immediate loss of income coupled with the competition for jobs by hundreds of workers with the same skills and experience, create overwhelming problems for the displaced worker. Education and training should provide outreach to this special population by improving existing services to guarantee a better-prepared work force to meet the challenges of the twenty-first century.

The debates of the high tech labor shortage has prompted the issue of the H1B visa to bring high tech workers to the United States due to a shortage of skilled workers and anticipated future shortages. The H1B visas derive their name from a definition found in the **Byzantine** Immigration and Nationality Act at 8 USC & 1101(a) (15) (l) (b). Basically, H1B is a category of visa application for those "who come temporarily to the United States to perform services in a specialty occupation that requires theoretical and practical application of a body of highly specialized knowledge and attainment of a bachelor's or higher degree in that specialty.

4. *The Desired Outcome*

In 1988, the Congress enacted the Economic Dislocation and Worker Adjustment Assistance Act to improve assistance provided to dislocated workers. Key provisions of this legislation are the establishment of (1) state rapid response teams to offer workers assistance before they are laid off and (2) labor-management committees to facilitate this assistance. However, relatively few states have experience with dislocated worker project involving labor-management groups. Closed plants, lost jobs, lives in disarray-all this and more portray the failure, for many, of the "American dream," and the emergence of a special adult population referred to interchangeably as dislocated or displaced worker.

Congress analyzes various proposals to expand or modify programs assisting dislocated workers, certain lessons learned and in three areas: allowing states the flexibility to select sponsors from a variety of operating organizations, as opposed to channeling all funds through service delivery areas, has proven successful, and having project staff with expert labor market knowledge and extensive contacts with local employers was a key ingredient to success.

5 *The Reported Unemployment Rate*

The national unemployment rate was indicated to be less than 5 percent in 1997, and more than 12 million jobs created since 1993, for many U.S. workers particularly minority, the labor market outcomes have stagnated or worsened. What is also creating employment barriers for minority is the flexible production of the new economy in increased contingent work. However, the decreasing unemployment has not quiet resulted in commensurate reductions in poverty. It was indicated that family poverty rates of 14 percent or more have existed for the past 15 years; the rate was reported to be 30 percent for African American (GAO, 1998).

In the fall of 1995, America's official unemployment rate was 5.7 percent. Those officially unemployed are just a small part of the total number of workers looking for more work. Adding together the officially unemployed (about 7.5 million), those who say they want work but do not meet one or

the other of the tests for being actively in the labor force and therefore are not officially counted as unemployed (another 5 to 6 million), and the involuntary part-timers who want full-time work (approximately 4.5 million) yields an effective unemployment rate approaching 14 percent (Thurow, 1996, p. 165).

Each year, millions of American workers are displaced from their jobs as a result of permanent and structural changes in the economy (Moore, 1990). The demand by business and industry for workers with up-to-date skills who can perform successfully in a more complex and demanding workforce represents the response of the American economy to the onset of the technological age. One result of this has been a shift "...in job opportunities from the hands of American workers to their minds" (Seitchik & Zomitsky, 1989 p. v).

A review of the literature also shows that workers in the goods-producing industries (e.g., mining, construction, and manufacturing) were more at risk for losing their jobs than workers in the service-producing industries (Fullerton, 1987). Researchers agree that although differences in worker tenure or productivity affect displacement (Herz, 1991), discrimination based on age, sex, or race may also lead to the targeting of particular groups of workers (Wise, 1989).

For the most part, worker displacement is a reflection of industry conditions. As foreign competition increase and the American economy responds to rapid advances in automation and computerization, clearly the need for greater job retraining and education is necessary to empower dislocated workers with the skills needed to compete in an ever-increasing, technologically advanced society.

The dilemma of the dislocated worker has received attention both in professional journals and periodicals as well as popular reading. Because unemployment represents both social and economic issues, the popularity of this subject can be seen in the numerous articles that address this special population. With thousands of workers looking for jobs, and facing the economic uncertainties of lost pensions and medical insurance, the media presents a dim yet accurate portrayal of relevant issues.

The dislocated workers reflect the changing economic climate of our country. In recent years, we have seen the emergence of a changing labor force that presents new challenges for training and employment. Our workforce had become more technologically advanced. Jobs that were once secure have become obsolete and replaced with employment opportunities requiring higher skills. Today's displaced worker appears to require a level of assistance beyond that provided by federal funding.

Our workers of the future will also need assistance to adapt to further economic change and restructuring. Clearly, the need for a joint partnership between business, government, education and the individual workers is in order to keep our working class working and to ensure a competitive and sound economy for the future.

6 CONCLUSIONS AND DISCUSSION

A Clearly the need for education and retraining as well as an increase in federal outreach is necessary to address the needs of the displaced worker. My proposed policy entitled, Job Opportunities through Business, Federal Outreach, Retraining, Counseling and Education (Job Force) is designed to keep the working class working, and to strengthen the American economy as we prepare for the year 2000.

The Economic Dislocation and Worker Adjustment Act (EDWAA) replaced title III of the Job Training Partnership Act (JTPA). Although EDWAA represents the government's answer to the issue of unemployment, clearly this is an insufficient response. My policy mandates an increase in federal funding responsive to the level of unemployment. As unemployment increases, Congress should direct more money toward federally assisted training and retraining activities. Research points to the fact that the United States budget figures represent outlays of funds rather than expenditures and that the latter, "might be significantly less than the amount of money available" (Wise, 1989, p. 20).

Increasing federal expenditures, however, is not enough. My policy calls for a proactive stance by business, education and government to provide and assist in training

and skill development of workers before displacement occurs. Under my policy, federal regulations will require business to create a framework for continued on-the-job training and education. Federal funds will be used to supplement training activities in addition to company and employee and employee funds.

The training programs will stress individual personal development and job-specific training and will be scheduled to take place during the workday. Career counselors will be recruited to assist the employee in researching employment opportunities' facilitate stress management workshops; provide counseling services; and, disseminate information on educational opportunities. The entire issue of the H1B visa rests on one of control. Who will control the future of the US labor market? The high tech industry, like any other industry is in need of high tech workers, but does not want to pay an unprofitable price for them, or labor unions, who have historically exerted much leverage over the manufacturing industry in America, but now are losing in their struggle to maintain historically high U.S. wage levels in a global market.

While H1B visas now assist the technology companies by replenishing their workers at a relatively low cost, the issue is: should this be done at the expense or future of the US citizen? The crux of the matter is that of a real shortage of technology workers, vs. a shortage of wage motivators that companies can play against US citizens getting involved in the technology sector. The rewards and compensation must be in line with the tasks or the field will in fact experience a dearth of professionals. Is this really the case though? The high tech sector claims this is so, as it can be seen in numerous Congressional testimonials. The labor industry claims this is not the case; their argument is that there are in fact enough high tech workers in the US, but they are pursuing other fields due to low compensation in this field (Such as MBA's, medical, and law degrees.) Both sides have presented countless testimony to congress bolstering their points, among some of the arguments.

Perhaps the best non-biased study concerning the whole industry is the March 20, 1998 report put out by the General

Accounting Office (GAO), titled: Information Technology: Assessment of the Department of Commerce's Report on Workforce Demand and Supply. This report conclusively demonstrates that the demand for IT workers is in fact critical, but it criticizes the Department of Commerce conclusions regarding supply. That being, Commerce's methodology in analyzing the future pools of IT workers was flawed. While it is true that the number of IT workers receiving Bachelor's degree's in computer science and information science falls well short of projected industry demand, the report fails to acknowledge *other* sources of IT workers, such as those with associates degrees, Ph. D's, degrees in other fields, and employees who can or have received industry wide training. While a shortage of IT workers currently exists, future supply, and even demand, has not been demonstrated conclusively. Indeed, the whole issue of supply needs to be explored more thoroughly in the future, since it remains unclear.

Thus said the issue is clearly over labor costs, and not the physical presence of someone being in the US. Work can be completed in a local office on computer and sent as an email attachment to the home office in a matter of seconds. Since this work is brainpower work and not manual labor, the electronic means of transmission and distance design takes a new meaning. The key driver here is labor costs and control. Even if they have to retrain an electrical engineer for their designs, in say Bangalore, India, it is incredibly cheaper than to retrain US engineer for the same tasks. Likewise, many of these countries have NO Western style labor laws or even minimum wages, but they do have a plethora of people, thereby creating a buyers market for labor costs.

The focus of this should not be underestimated, while the US will remain a Mecca for sales and administration in the high tech industries, it will spell doom for the future of high tech workers, as Asian countries are becoming more accessible and on-line. The only current obstructions to a wholesale transfer of design to developing countries are political instability and educational access to the workers there, meaning that the Asian thinking is not Western thinking, and can hinder the cultural

mores of doing business and expected outcomes. This is a transitional period and will not last long. However, already there is a strong preference for Western styles and concepts in many Asian countries, as the older population dies off, indigenous cultures will be compromised or go by the wayside entirely. When that happens, we will truly have a “global” market, complete with global thinking.

REFERENCES

- [1] Blong, J. T. & Shultz, R. M. (1990). The dislocated worker: When training is not enough. *Community, Technical & Junior College Journal*, 60, 28-32.
- [2] Cetron, M. & Davies, O. (1989). *American renaissance: Our life at the turn of the 21st century*. New York: St. Martin's Press.
- [3] Fullerton, H. N. (1987). Labor force projections: 1986 to 2000. *Monthly Labor Review*, 110(2), 19.
- [4] GAO Report (March, 1998). *Information Technology: Assessment of the Department of Commerce* Report on Workforce Demand and Supply. GAO/HEHS – 98-106 Information Technology Workers.
- [5] Herz, D. E. (1991). Worker displacement still common in the late 1980's. *Monthly Labor Review*, 114(5), 3-7.
- [6] Kates, N., Greiff, B., & Hagen, D. (1990). *The psychological impact of job loss*. Washington: American Psychiatric Press, Inc.
- [7] Mangum, G. & Mangum, S. (1996). *Of heart and mind: Social policy essays in honor of Sar A. Levitan*.
- [8] Moore, T.S. (1990). The nature and unequal incidence of job displacement costs. *Social Problems*, 37(2), 230-235.
- [9] Seitchik, A. & Zornitsky, J. (1989). *From one job to the next: Worker adjustment in a changing labor market*. W. E Upjohn Institute for Employment Research.
- [10] Thurow, L. C. (1996). *The future of capitalism: How today's economic forces shape tomorrow's world*.
- [11] Wise, L. R. (1989). *Labor market policies and employment patterns in the United States*. Boulder: Westview Press. Pp. 18-25, 70-89.
- [12] Stephens Jr., M. (2002). Worker Displacement and the Added Worker Effect. *Journal of Labor Economics*. The University of Chicago, vol. 20, no. 3.
- [13] *Tech Law Journal*. (1998). H1-B visa Bills, S 1723, S 1823, HR 3736. (On-line). Available: <http://www.techlawjournal.com/congress/s1723h1b/default.htm> (Retrieved: 8/1/04)
- [14] Lane, J. (2003). *Pathways to Work for Low-Income Workers: The Effect of Work in the Temporary Help Industry*. The Urban Institute, Washington DC. (Retrieved on: 7/23/04)

Factors affecting supplier adoption of governmental reverse auctions: an exploratory case study

Abirami Radhakrishnan, Steve Davis, and Robert B. Davis

Abstract—Governmental reverse auctions have not been as widely adopted by suppliers as buyers expected. There is no prior study of this issue. Using an embedded case study, we found the following factors affect willingness to adopt governmental reverse auctions: a system factor (difficulty), perceived external factors (price based competition intensity, customer power and information intensity) and organizational readiness factors (financial readiness and technological readiness). Also we provided practitioners with suggestions to improve interaction with governmental reverse auctions.

Index Terms— Case study, E-procurement, Governmental reverse auctions

I. INTRODUCTION

Although United States Federal and State Governments spend about \$2.8 trillion on procurement annually [29], only about 0.4% is handled by governmental reverse auctions [13]. Governmental reverse auction systems are inter-

organizational systems that facilitate electronic procurement by United States Federal and State Governments. An auction is a market institution with an explicit set of rules determining resource allocation and prices on the basis of bids from market participants. Auctions play a prominent role because they remain one of the simplest and most familiar means of price determination in the absence of intermediate market makers. Auctions are called “reverse” if suppliers bid instead of buyers, and the prices are bid down instead of up [15, 18, 19].

Although some procurement such as service contracts emphasizing “best value”, may not be appropriate for reverse auctions, a wide variety of commodity products are entirely appropriate for procurement by reverse auction. We are not aware of any federal regulations restricting reverse auctions. On the contrary, the Federal Acquisition Regulation mandates competitive procurement, fair and open competition and a mechanism to establish fair market price [1]. Reverse auctions fulfill these requirements for most situations. The low percentage involvement with reverse auctions is affected by the inability of government buyers to motivate enough suppliers to adopt reverse auction systems. Even though governmental buyers are heavily committed to reverse auctions [25], suppliers are not euphoric about participating in governmental reverse auctions. Indeed, the actual adoption of governmental reverse auctions by suppliers has been far below experts’ expectations. Several marketplaces that began with highly optimistic initial projections have already failed and many are struggling to survive [17, 30, 31]. There seems to be an

Manuscript received Aug 31, 2004. The authors appreciate financial support from The SEAMS Association and assistance from and The Hosiery Association.

Abirami Radhakrishnan is with the Department of Management, Clemson University (e-mail: aradhak@clemson.edu); contact person Steve Davis is with the Department of Management, Clemson University, Clemson, SC 29634-1305, USA (e-mail: davis@clemson.edu); Robert B. Davis is with the Department of Management, Clemson University (e-mail: bdgot@yahoo.com)

unexplained discrepancy between the promise and the reality of governmental reverse auctions. Government agencies apparently believe their reverse auction systems are fine and should be appealing to suppliers, yet many suppliers are reluctant to adopt. Because of the importance of this issue, this reluctance of suppliers to adopt deserves serious attention.

Studies on adoption of reverse auctions in private industry from suppliers' perspective [8, 9, 10, 11, 12] show two major concerns that may hinder their adoption of reverse auctions. Suppliers widely perceive online reverse auctions as a divisive tool used by corporate buyers to drive down prices without adequate consideration given to other measures of performance resulting in increasing price based competition intensity. This perception is not always accurate. Buyers satisfied with "non-contractible" aspects of suppliers such as quality, supplier innovativeness, information sharing, responsiveness, trust, and flexibility might stay with a dedicated supplier base even when their prices are not lowest [24]. Suppliers perceive these auctions tip the balance of power toward corporate buyers.

The aforementioned supplier concerns may cause them to distrust corporate buyers, impacting them adversely [12, 32]. Suppliers tend to distrust corporate buyers if they believe they use online reverse auctions not with a real intention of choosing the best supplier but instead are attempting to drive down unit prices. The implications of studies of private industry do not necessarily apply to governmental reverse auctions because [25]:

- (1) governments are not usually subject to same kind of aggressive competition that private industry faces,
- (2) governmental agencies create laborious procedures and check points to insure that procurement procedures are fair and to avoid fraudulent practices, and

- (3) managers in governmental agencies do not have the same incentives to lower costs as managers in private industry.

We have not found any study that has explored suppliers' adoption of governmental reverse auctions. Therefore, this article investigates the major factors that influence suppliers' willingness to adopt governmental reverse auctions. We conducted an exploratory embedded case study guided by theory from literature on adoption of other related inter-organizational systems. The next section summarizes this theoretical background followed by our research methodology, case analysis and findings, and conclusions.

2. THEORETICAL BACKGROUND

Since a governmental reverse auction system is a type of inter-organizational system, we reviewed the inter-organizational systems adoption literature to find critical factors affecting adoption of systems similar to reverse auctions [6, 14, 15, 16, 18]. From this literature, we employed system factor – difficulty, compatibility, and perceived benefits (Table 1). System difficulty refers to the perceived difficulty in using reverse auction Websites [6]. Compatibility refers to the perceived compatibility with the existing technical infrastructure [14]. Perceived benefits include reduced cost to find appropriate marketing opportunities and reduced paper work [14, 15, 16, 18]. High perceived benefits and compatibility may encourage companies to adopt reverse auctions. Low system difficulty may encourage companies to adopt reverse auctions.

Inter-organizational systems literature suggested another factor – organizational readiness (Table 2), which includes financial readiness (to meet ongoing expenses during IT resources usage) and technological readiness (expertise to use reverse auction Websites) [16]. High financial readiness and technical readiness may encourage companies to adopt reverse auctions.

Table 1. Perceived reverse auction system factors

Factor	Definition	Proposed relationship willingness to adopt governmental auctions
Compatibility	Perceived compatibility with the existing technical infrastructure [14].	High compatibility could encourage firms to adopt governmental auctions.
Difficulty	Perceived difficulty in using governmental reverse Web sites [6].	High difficulty could discourage firms from adopting reverse auctions.
Perceived benefits	Perceived benefits from adopting governmental auctions [14,15,16,18] * Paperwork reduction. * Reduced costs to find appropriate marketing opportunities	If suppliers perceive benefit they may be willing to

Table 2. Organizational readiness factors

Factor	Definition	Proposed relationship with willingness to adopt governmental reverse auctions
Financial readiness	Financial resources to pay for installation costs and ongoing expenses during usage, including salary of specialists [16].	Difficulty of paying for necessary equipment or services could discourage adoption.
Technological readiness	Level of preparedness to use modern IT resources [16].	Readiness to use modern IT resources could encourage adoption.

B2B reverse auctions literature suggested perceived external factors -- price based competition intensity, customer power, and information intensity (Table 3) [8, 9, 10, 11,

12, 17, 30, 31, 32]. We call these “perceived external factors” because they cannot be controlled directly by supplier firms. Low levels of these factors may encourage suppliers to adopt governmental reverse auctions.

Table 3. Perceived external factors

Factor	Definition	Proposed relationship with willingness to adopt governmental reverse auctions
Price based competition intensity	Perceived intensity of competition among suppliers based on lower price [8, 9, 14].	The negative perception that reverse auctions are used as a divisive purchasing tool designed to drive down prices could discourage adoption.
Customer power	Perceived bargaining power enjoyed by customers [14].	High customer power could discourage adoption.
Information intensity	Perception on the extent of exchange of production, scheduling and capacity information [8, 15].	High information intensity could discourage adoption.

Auction theory [23] and Transaction Cost Economics (TCE) [33] could explain some of these perceived external factors. Auction theory acknowledges the importance of information (e.g. bid price) and its potential exploitation in open-bid auctions like English, Dutch, and online reverse auctions [20]. Having imperfect information and knowing that only the lowest bid wins in a reverse auction often cause suppliers to bid less than the worth, which is known as the “winner’s curse” [21]. According to TCE, unequal access to information in market contracting, such as reverse auctioning, may enable one party (e.g. the buyer) to act opportunistically and exploit information advantages [27]. We present our research model in Figure 1, where a “+” or “-” sign indicates whether we propose a positive or negative relationship between factors and willingness to adopt.

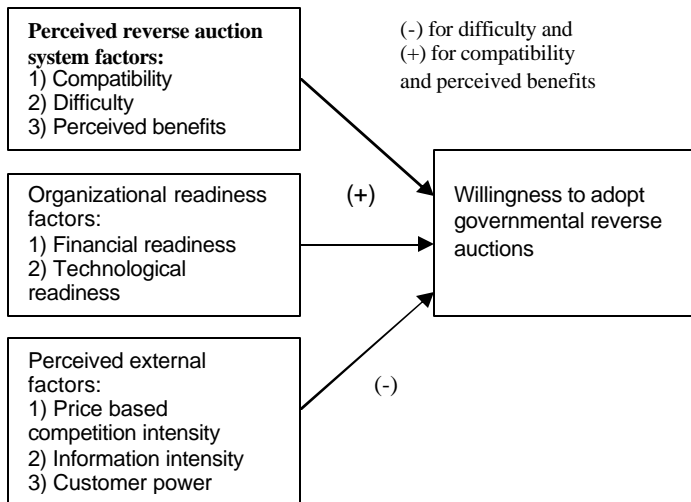


Figure 1. Proposed model for willingness to adopt governmental reverse auctions

3. RESEARCH METHODOLOGY

A. Nature of Our Study

We conducted an exploratory case study of the effects of the three groups of factors (system related, organizational readiness, external) on the willingness to adopt governmental reverse auctions. A case study is appropriate for an investigation like this that is one of the first to explore a phenomenon [34]. We employed a case study design (two industry trade associations each representing multiple companies). Using two associations allowed comparisons between them for more general research results [5, 34].

B. Organizations Considered

We wanted to know the issues from the perception of industry associations having many firms who are still deciding whether to adopt governmental reverse auctions. We chose two industry associations having many such members, referred to as “HA” and “SA” in this study. These associations voice to government agencies concerns of member firms. Each association has a senior full time employee having expertise in governmental procurement regulations and policies. These experts routinely provide advice to member firms and are very knowledgeable about the issues that influence member firms’ adoption of reverse auctions. In effect these experts have

already polled members, and knew perceptions and experiences. Associations have also hired persons (as full time employees) who review daily postings of governmental reverse auctions. They notify member firms of opportunities appearing to match their capabilities and they assist member firms in participating in auctions.

We present the demographics of the member firms of the two associations below to highlight the nature of member firms under each association. SA has 168 firms in the sewn products industry, of which 144 are small firms (less than 100 employees), 16 firms are medium firms (100-250 employees) and 8 firms are large firms (more than 250 employees). Most of the firms under SA association manufacture woven, knitted, or sewn fashion apparel and sportswear. HA has 541 firms in the hosiery industry, of which, 370 firms are small firms, 130 firms are medium firms and 41 firms are large firms. Most of the firms under HA association manufacture or distribute hosiery items like socks and tights.

C. Data Collection Method

We primarily collected and triangulated qualitative data from multiple sources. We conducted structured telephonic interviews with an expert from each of the two associations. We relied heavily on the support of these two experts. We used a structured interview guide for our telephonic interviews (see Appendix). This interview guide had both closed ended questions and open ended probing questions. Findings from closed ended responses were compared with answers to the open-ended questions. We verified answers by having interviewees review our notes on their responses.

We also conducted several face – to – face meetings with representatives from many firms under each association who were willing to participate in our project. We discussed reverse auctions problems and issues during four formal meetings, one with 10 and another with 14 industry representatives of the SA association and two others each with 8 industry representatives of the HA association. Almost all of these representatives were upper level managers (general managers, marketing managers and plant managers)

having good knowledge of company involvement in reverse auctions.

We also reviewed reports from secondary sources such as trade journals. We also conducted content analysis of several governmental reverse auction web sites. We considered data associated with the period 2002 - 2004. Krippendorf (1980) defines content analysis as a technique wherein a researcher searches for structures and regularities in the content and draws inferences on the basis of these regularities [22]. Here content refers to Web site functional capability. For consistency, the same researcher did content analysis of the various Web sites.

4. CASE ANALYSIS AND FINDINGS

We did comparative analysis of the telephonic interview data collected from the experts of the two associations and face-to-face meetings with representatives of several member firms to identify the issues that influenced suppliers' willingness to adopt governmental reverse auctions (Table 4).

Organization	Perceived auction system factors			Outcome
	Difficulty	Compat- ibility	Perceived benefits	Willingness to adopt reverse auctions
SA	High	High	High	Low
HA	High	Low	Low	Low

Organization	Org. readiness factors		Outcome
	Financial readiness	Technological readiness	Willingness to adopt reverse auctions
SA	Low	Low	Low
HA	Low	Low	Low

Organi- zation	Perceived external factors			Outcome
	Customer power	Price based competition intensity	Information intensity	Willingness to adopt reverse auctions
SA	High	High	High	Low
HA	High	High	High	Low

A. Perceived Reverse Auction System Factor - Difficulty

Experts from both associations did not agree on other system factors, and we had no other evidence of their importance, but system difficulty stood out as a major system factor that dissuaded suppliers from adopting reverse auctions.

Experts opined that firms who were reluctant to adopt perceived that governmental reverse auctions were very complex to use, highly arduous and time consuming. The Federal Government recently has begun to consolidate procurement sites into a single Web site called "Federal Business Opportunities". Even though the law mandates that all Federal Government agencies post every bid above \$25,000 on this site, suppliers still need to search many federal, states, local agency, and private party Web sites for bids less than \$25,000. Each Government department has its own reverse auction Web site.

Some states have multiple procurement departments having different Web sites. To find some kinds of bids may require searching more than 200 reverse auction Web sites, whose interfaces have different formats and standards. Because some Web sites are difficult to navigate without training, many suppliers need outside help and assistance to locate suitable bids. They perceive that many of these Web sites have inadequate capabilities to match suppliers' offerings with customers' requirements. The expert of the SA association remarked "... Now, there are many different ways of posting information and getting responses from bidders. At least, there should be a standard for systems within each state (for government offices). I spent 8 hours trying to find all information needed to submit a bid and finally I gave up..."

The expert on the HA association remarked "...takes too much time for small manufacturers to participate in these reverse auctions. They need to be simpler, and one stop shopping (consolidated) is needed."

Our content analysis (analysis of functional capability) of 177 governmental reverse auction Web sites (60 Federal, 52 State, and 65 local agency Web sites) provides more evidence of the system difficulty issue. From our consultation with experts from the two associations and with

representatives from member firms, we developed a list of important functional capabilities, including: availability of vendor guidance, ability to view bid guidelines, ability to download bid guidelines, availability of procurement history, and email notification. We compared the functional capabilities of different categories of governmental reverse auction Web sites (Figure 2). We found that nearly 70% of the local agencies run Web sites do not have many of the functional capabilities. Nearly 85% of the federal Web sites do not provide procurement history. Nearly 50% of the federal Web sites do not provide bid guidelines. Nearly 80% of the federal sites do not have email notification. Nearly 70% of the state Web sites do not provide procurement history and nearly 48% of the state sites do not have an email notification facility.

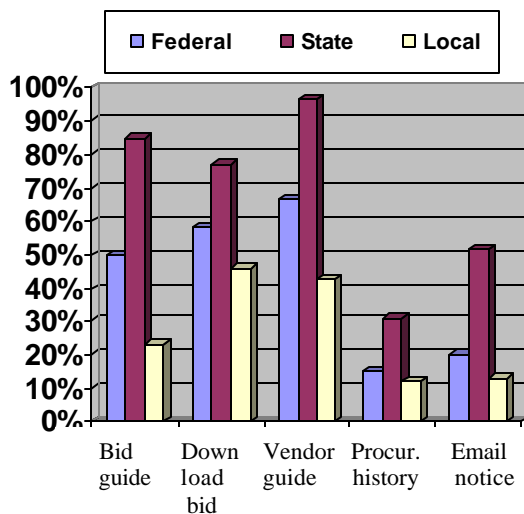


Figure 2. Comparison of functional capabilities of governmental reverse auction Web sites.

B. Perceived External Factors

We found all three external factors – price based competition intensity, customer power, and information intensity play an important role in affecting the willingness to adopt governmental reverse auctions.

Price Based Competition Intensity.

Consistent with B2B reverse auctions literature [8, 9, 10, 11, 12, 17, 26, 30, 31], experts from both associations opined that member firms, especially those dealing with high margin, low volume products such as

fashion apparel, sportswear and special tights perceived high competition intensity.

Since governmental reverse auctions encourage suppliers from various countries to bid and since online reverse auctions facilitate this process, suppliers from many different countries may participate. Many suppliers have to compete with players from other lesser-developed countries where costs are much lower on account of much lower wages, raw material costs and overheads. (Sometimes this competition is indirect when local firms outsource.) This makes them unable to qualify as lowest bidder. For example, an expert from HA association remarked, "... wage requirements in United States do not allow competition with low foreign wages. Sometimes foreign wages are as low as 19 cents per hour. Similarly, the power charges and the facility expenses are very low in some countries. Another problem is that raw material manufacturers within the United States cannot compete with the open market where regulations by the Environmental Protection Agency and the Occupational Safety and Health Administration regulations do not govern health and safety issues..."

Mithas et al. (2004) found that buyers concerned about "non-contractible" aspects of suppliers such as quality, supplier innovativeness, information sharing, responsiveness, trust, and flexibility generally prefer to work with a dedicated supplier base instead of using reverse auctions [24]. However, some government buyers in industry sectors where non-contractible issues are important unwisely employ reverse auctions. This invokes supplier perception that buyers use price as a divisive tool to promote competition and it dissuades suppliers from adopting reverse auctions.

Customer Power. Experts from both associations opined that the member firms perceived that switching costs for Governmental procurement offices are low. Because customers (buyers) can easily switch between suppliers of these kinds of products, they have higher bargaining power that dissuades suppliers from adopting reverse auctions. Open-bid auctions increase supplier's bargaining costs, making the process disagreeable to suppliers. The open-bid format forces price concessions

from the supplier thereby increasing customer (buyer) power. Our finding is consistent with applications of Transaction Cost Economics theory to strategic information systems [2, 3, 4, 7]. Bakos (1986) [2] (by applying TCE in the context of strategic IT systems) asserts when switching costs are low, one enjoys higher bargaining power. Government can easily switch when conducting reverse auctions. Our finding is also consistent with the practitioner literature on reverse auctions in private industry [9, 11]. Reverse auctions practitioner literature highlights that higher customer power in aerospace and automobile sectors dissuades sellers from participating in reverse auctions.

Information Intensity. Both the associations' experts said that customers sometimes expect suppliers to share information on capacity, production, and scheduling. Providing this information is difficult for suppliers because they are unable to commit capacities for uncertain future business volumes. For example, an expert from SA association remarked "...I did not win one Federal solicitation because it asked for projection of capacity 6 months in the future; I could not do it due to the uncertainty of future business; this is a common problem because of long lead times....."

Suppliers perceive the requirement to share production and scheduling information as a pressure tactic to increase manufacturing flexibility.

C. Organizational Readiness Factors

We found that financial readiness and technological readiness encourage adoption of governmental reverse auctions, consistent with inter-organizational systems literature [14, 16, 28].

Financial Readiness. Both the associations' experts perceived that many member firms are not willing to invest to support governmental reverse auctions, especially the smaller firms. The most expensive requirement is trained personnel who could search effectively for new bid opportunities among governmental reverse auction Web sites.

Technological Readiness. Both the associations' experts perceived that many member firms do not have adequate infrastructure (such as dedicated computers with web browsers and high speed internet connection) to carry out reverse auctions processes. They also said there is a lack of competent people having training in dealing with reverse auctions. When asked for factors that would enable member firms to effectively adopt reverse auction sites, an expert of SA association pointed out that "...Some companies have a contract person on staff. Working with the governmental reverse auction sites could be full time job. In some cases, a person would need to specialize in certain reverse auction sites or products...."

Poor financial and technological readiness of the member firms of the two associations dissuaded them from adopting governmental reverse auction sites, consistent with inter-organizational systems adoption literature [14, 16, 28].

5. CONCLUSION

To our knowledge, this study is the first of its kind to build a theoretical framework on factors affecting suppliers' willingness to adopt governmental reverse auctions. An embedded multiple case study supported this framework. The following factors could affect willingness to adopt governmental reverse auctions: a system factor (difficulty), perceived external factors (price based competition intensity, customer power and information intensity) and organizational readiness factors (financial readiness and technological readiness). These findings broaden the current theory on B2B reverse auctions by adding new factors, system difficulty and organizational readiness.

There are several implications for practitioners. Participating in reverse auctions may be worthwhile for suppliers in cases where their value proposition is appropriate. For example, companies that deal with high volume, low price commodity kind of products may derive benefits by participating in governmental reverse auctions because they offer an opportunity in addition to other marketing channels. Even if the profit margin may be less than with other channels there may be

economies of scale or the opportunity to employ capacity that would otherwise be wasted.

Participating firms need to invest in personnel who are well trained to handle governmental procurement procedures and Web sites. Industry associations could help such member firms by employing such professionals because benefits and costs would be shared by all such member firms. Possibly trained professionals could be augmented or replaced by intelligent agents, knowledge-based systems that could help search and bidding processes.

Second, representatives of industry associations could lobby with the government to make reverse auction Web sites more user friendly and to make the reverse auction bidding processes less cumbersome. They need to convince government that addressing the system difficulty issue would likely encourage more suppliers to adopt governmental reverse auctions. Also, representatives should recommend that a one stop governmental Web site should be developed, instead of each government department running its own reverse auction Web site and forcing suppliers to search several Web sites to find new opportunities. Until such a site is developed, it would help if all governmental reverse auction Web sites supported consistent data formats.

Third, managers of federal and local government reverse auction Web sites should improve functional capabilities. These steps will encourage adoption by suppliers.

There are several implications for academic researchers. It would be worthwhile to conduct research to evaluate the framework developed in this exploratory study. For our study, our unit of analysis was an association. So we could not compare individual member firms. Also, our analysis was primarily based on qualitative data. It could be worthwhile to collect large-scale empirical quantitative data by means of a structured survey instrument (questionnaire developed after assessing validity and reliability of constructs) and involve suppliers from several industries. Alternatively, multiple case studies with firms from several different industries could help evaluate the framework.

Since system difficulty of reverse auction Web sites appears such a critical factor affecting adoption, researchers could investigate this issue in more detail by conducting usability studies [6] to develop specific recommendations for reducing cumbersomeness and complexity.

APPENDIX

The guide for our telephonic interviews with experts from the two associations can be found at

<http://people.clemson.edu/~davis/Appendix.doc>

ACKNOWLEDGMENT

We thank Dean Allred and Elroy Pierce for providing valuable information about their experiences in governmental reverse auctions. Sarah Friedman and Maggie Parker provided helpful data, advice and support.

REFERENCES

- [1] Acquisition Regulation. <http://www.arnet.gov/far/>
- [2] Bakos, J.Y., Treacy, M.E., "Information Technology and corporate strategy: a research perspective", Management Information Systems Research Center, University of Minnesota, MIS Quarterly, 1986, pp. 369-385.
- [3] Bakos, Y.J., "A strategic analysis of electronic marketplaces", Management Information Systems Research Center, University of Minnesota, MIS Quarterly, 1991, pp. 295-310.
- [4] Bakos, Y.J., "Reducing buyer search costs: implications for electronic marketplaces", *INFORMS*, Management Science, 1997, pp. 1676-1708.
- [5] Benbasat, I., Goldstein, D. K., Mead, M., "The case research strategy in studies of Information Systems", Management Information Systems Research Center, University of Minnesota, MIS Quarterly, 1987, pp. 369-385.
- [6] Benbunan-Fich, R., "Using protocol analysis to evaluate the usability of a commercial web site", Elsevier, Information and Management, 2001, pp. 151-163.
- [7] Choudhury, V., "Strategic choices in the development of inter-organizational information systems", *INFORMS*, Information Systems Research, 1997, pp. 1- 24.
- [8] Emiliani, M. L., "Business-to-Business Online Auctions: key issues for purchasing process improvement", Emerald Group Publishing Limited, International Journal of Supply Chain Management, 2000, pp. 176 - 186.
- [9] Emiliani, M.L., Stec, D.J., "Squaring online reverse auctions with the Caux Round Table Principles for Business", Emerald Group Publishing Limited, International Journal of Supply Chain Management, 2002, pp. 92 - 100.

- [10] Emiliani, M.L., Stec, D.J., "Realizing savings from online reverse auctions", Emerald Group Publishing Limited, Supply Chain Management: An International Journal, 2002a, pp.12 – 23.
- [11] Emiliani, M.L., Stec, D.J., "Aerospace Parts Suppliers' Reaction to Online Reverse Auctions", Emerald Group Publishing Limited, Supply Chain Management: An International Journal, 2004, pp. 139-153.
- [12] Emiliani, M.L., "Sourcing in the Global Aerospace Supply Chain Using Online Reverse Auctions", Invited paper, North Holland, Industrial Marketing Management, Special Issue on global supply chain management, 2004, pp. 65-72.
- [13] "E-Procurement: finally ready for prime time", *AberdeenGroup*, 2001, pp. 1- 12.
- [14] Grover, V. "An empirically derived model for adoption of customer based inter-organizational systems", Blackwell Publications, Decision Sciences, 1993, pp. 603- 637.
- [15] Gebauer, J., Segev, A., "Emerging technologies to support indirect procurement: two case studies from the petroleum industry", Kluwer Publications, Information Technology and Management, 2000, pp. 107 – 128.
- [16] Iacovou C.L., Benbasat, I., Dexter, A.S., "Electronic Data Interchange and small organizations: adoption and impact of technology", Management Information Systems Research Center, University of Minnesota, MIS Quarterly, 1995, pp. 465 – 486.
- [17] "Reverse Auctions fail to deliver on their promise", Information Week, 2003, Sept 8, p.10.
- [18] Jap, S.D., "Online reverse auctions: issues, themes, and prospects for the future", Sage Publication, Journal of the Academy of Marketing Science, 2002, pp. 506-525.
- [19] Jap, S.D., "An exploratory study of the introduction of online reverse auctions", American Marketing Association, Journal of Marketing, 2003, pp. 96 – 110.
- [20] Kambil, A, Van Heck, E., "Reengineering the Dutch flower auctions: a framework for analyzing exchange organizations", INFORMS, Information Systems Research, 1998, pp. 1 - 19.
- [21] Klemperer, P., "Auction theory: a guide to the literature" Journal of Economic Surveys, 1999, pp. 227-286.
- [22] Krippendorff, K., "Content Analysis: An introduction to its Methodology", Sage Publications, Beverly Hills, CA, 1980.
- [23] Milgrom, P., "Auctions and bidding: a primer", Journal of Economic Perspectives, 1989, pp. 3-22.
- [24] Mithas, S. Jones, J.L, Mitchell, W. Non-contractibility and asset specificity in reverse auctions: An Empirical Analysis University of Michigan Business School Working Paper, http://faculty.fuqua.duke.edu/~willm/bio/cv/working_papers/2004_07_ReverseAuctions.pdf, 31, July, 2004
- [25] National Institute of Governmental Purchasing, "U.S. Federal Government Procurement: structure, process and current issues", http://www.nigp.org/research/Papers/041003USG_ovt.pdf 12, February, 2004.
- [26] "Reverse auctions have a disastrous impact", Packaging Magazine, April 10, 2003, p.8.
- [27] Parker, D., Hartley, K., "The economics of partnership sourcing versus adversarial bidding: a critique", European Journal of Purchasing and Supply Management, 1997, pp.115-125.
- [28] Premkumar, G., Ramamurthy, K., "The role of Inter-Organizational Systems and organizational factors on decision mode for adoption of IOS", Blackwell Publications, Decision Sciences, 1995, pp. 303 – 336.
- [29] "Fiscal year 2001-2005 strategic plan", Procurement Executive Council, 2001, p. 29.
- [30] "Online reverse auctions create two procurement camps. (E-Procurement Strategies)", Purchasing, March 21, 2002, pp.36S6 – 36S8.
- [31] "Mixed Message", Traffic World, June 23, 2003, pp. 15-16.
- [32] Tulder, R., Mol, M., "Reverse auctions or auctions reversed: first experiment by Phillips", Elsevier, European Management Journal, 2002, pp. 447 – 456.
- [33] Williamson, O.E., "The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting", Free Press, New York, NY, 1985.
- [34] Yin, R. K., "Case Study Research, Design and Methods", (2nd ed.), Sage Publications, Beverly Hills, CA, 1994.

Abirami Radhakrishnan is a Ph.D. candidate in Industrial Management in Clemson University. Her research interests include inter-organizational systems for supply chain management, supply chain integration, e-commerce systems and performance measurement systems.

Steve Davis has a Ph.D. from Georgia Institute of Technology. He is a Professor in the Department of Management with a joint appointment in Computer Science at Clemson University. His research interests include software engineering and information systems for supply chain management.

Robert B. Davis is a student in the Department of Management at Clemson University. His research interests include e-procurement systems.

Quantifying the Learning Efficiency of Programming Learning and Its Implications

Yuh-Huei Shyu

Abstract—*With the advent of distance learning, educators are having a hard time deciding whether to adopt this kind of education model or not. We need some metric systems to judge the superiority of different education models. To simplify the problem, our discussion will be restricted to programming language courses. Focusing on the issue of “learning efficiency,” in this paper, we will define some quantifiable parameters and measuring rules. These parameter values can reflect the features of the underlying education models. Derived parameter values not only reveal bottlenecks of the current programming learning processes, but also provide guidelines for designing better education models. The salient results are: (1). the learning inefficiency is in fact an intrinsic property of the traditional model itself, and (2). “Active Teaching and Passive Learning for Novices.” is a better teaching strategy for novices. Finally, combining all results we obtained, we propose the “CD-type learning model.” This model might provide a very efficient learning environment for novice programmers.*

Index Terms—*learning efficiency, Distance learning, Active learning, Passive teaching, Passive learning, Active teaching, CD-type learning*

1. INTRODUCTION

WITH advances of Internet technologies, hundreds of distance learning systems have been designed, implemented, and adopted. Distance learning systems provide

environments that break temporal and spatial limitations [2,3,4]. Some researches indicate that distance learning is or can be as effective as traditional education [4,9]. While the latest report given in SIGCSE'04 indicates that online students are less successful than on ground students [11]. These reports bring the following question to the forefront:

“Is distance learning superior to the traditional education system?”

This question is very hard to answer. The major difficulty comes from the fact that the meaning of “superiority” is not clearly defined. Current methods for judging the superiority of different education models are mainly based on learners’ final scores or questionnaires. This kind of comparison seems to be very subjective. Thus, we need some metric systems to measure the performances of education models.

Typical education activities involve not only the *instructors* and *learners*, but also different *underlying education models*. We found that the underlying education model plays a key role on learners’ learning efficiency. Once an instructor adopts an education model, he also sets the minimum time requirement for learners to master the designated course. We will discuss this issue later in the paper.

In this paper, the traditional education model will be defined as teaching activities that using no Internet technologies or mailing mechanisms; while the distance learning model is defined as teaching activities that using Internet technologies. In the traditional education model, instructors adopt different types of teaching methods such as lectures, discussions, and hands-

Manuscript received July 24, 2004.

Y. Shyu is with the Department of Information Engineering, Tamkang University, Taiwan, R.O.C., (e-mail: yuhhuei@mail.tku.edu.tw).

on. From this point forward, we will use the words “traditional education models” to represent different types of the traditional education model. Analogously, we will use the words “distance learning models” to represent different types of the distance learning model. First, let’s point out the following facts.

1. A course teaching activity involves at least three elements: the instructor, the learners, and the underlying education model.
2. Elements of the instructor and the learners are changeable.
3. Most learners’ final scores highly depend on the instructor’s teaching quality.

Using learners’ scores to judge the superiority of an education model is very subjective and has the following shortcomings.

1. Little or no consideration of the learner’s background, talent, and motivation.
2. Little or no consideration of the instructor’s teaching quality.
3. Little or no consideration of the influence of the underlying education model.

We need a clear definition of the meaning of “superiority”. To achieve that, we need to decide parameters and measuring rules so that selected parameters will reflect the superiority of the designated education model as precisely as possible. In addition, these parameter values should be quantifiable. In this paper, we will focus only on programming courses without further specification. These parameter values can reflect the features of the underlying education models. Derived parameter values not only reveal bottlenecks of the current programming learning processes, but also provide guidelines for designing better education models. The salient results are: (1). the learning inefficiency is in fact an intrinsic property of the traditional model itself, and (2). “Active Teaching and Passive Learning for Novices.” is a better teaching strategy for novices. Finally, combining all results we obtained, we propose the “CD-type learning model.” This model might provide a very efficient learning environment for novice programmers.

The remainder of this paper is organized as follows. In section 2, we will discuss the concept of “learning efficiency” and define some measuring parameters and rules. In section 3, we will evaluate parameters for different models. In section 4, we will discuss issues related to comparison of education models. In section 5, we will propose a new education model called “CD-Type learning model.” In section 6, we will provide conclusions and recommendations for future research.

2. LEARNING EFFICIENCY

A. Definitions

There are many analogous definitions for the words “effective” and “efficient” in dictionaries and reports, we will use the definitions given in [1]:

1. Effective means doing things successfully.
2. Efficient means using resources wisely and without unnecessary waste.

Note that whenever we talk about effectiveness we often use some subjective criteria. While when we talk about efficiency, the parameter “time” is considered explicitly. Now, we will extend the above definitions to course teaching and learning.

1. “Teaching effectively” means teaching a course successfully.
2. “Learning effectively” means learning course content successfully.
3. “Teaching efficiently” means using teaching resources available wisely and without wasting unnecessary time.
4. “Learning efficiently” means using learning resources available wisely and without wasting unnecessary time.

To judge the superiority of an education model, it would be more convincing to judge its efficiency, rather than its effectiveness, since learners’ grades only reflect the teaching/learning effectiveness under a pre-scheduled syllabus with a designated instructor. These data should not be used to determine the superiority of the underlying education model.

B. Parameters of Learning Efficiency

In general, a course includes many students and one instructor. Therefore, instead of considering “teaching efficiency,” we shall consider the “learning efficiency” of a course. We will use the time required to reflect the status of learning efficiency. Learning efficiency is influenced by three major factors: the instructor, the learners, and the education model used. The first two factors are changeable (subjective) parameters; therefore, we shall ignore these two factors. The third factor is what we will try to quantify.

The concept of efficiency takes two values to compare: the base value and the target value. We will use the concept of span of total time scheduled as the base value, and the concept of span of total time actually used as the target value. For example, we define one semester as the base value for a given course. The total time that a learner needs to “finish” this course is defined as the target value. Here, we will let the meaning of “finish” be open. From the instructor’s aspect, it may mean when a learner passes the course. While from a learner’s aspect, it may mean when he is confident with the course content.

The learning efficiency depends on both the base value and the target value. These values have the following interpretations.

1. Base value depends on the education model.
2. A smaller base value means that the course is using a more efficient model.
3. Target value depends on both the learner and the instructor.
4. A smaller target value means that the teaching is more effective.

We will now provide assumptions and definitions that will be used in later discussions.

Assumption 1:

Self-study of a course takes more time than being guided by a qualified instructor.

Definition 1: Course Content Delivered Time (T_{ccd})

The “Course Content Delivered Time” of a given course is defined as the time interval between the beginning of the first scheduled lecture and the end of the last

scheduled lecture. In the traditional education model, the T_{ccd} of a given course is about a semester or a quarter.

Definition 2: Maximum Lectures Interval Time (T_{mli})

The “Maximum Lectures Interval Time” of a given course is defined as the maximum time interval between two adjacent scheduled lectures.

Definition 3: Course Mastering Time (T_{cm})

The “Course Mastering Time” for a learner in a given course is defined as the time interval from when he starts the course to the time when he “thinks” that he has mastered the course content.

We shall leave the meaning of “thinks” stated above open. Intuitively, the lower value of T_{cm} means that the learning is more efficient.

Assumption 2:

Under the same definitions given above, we will leave it for related people to decide if the learning is efficient.

Assumption 3:

Assume that most students study the course according to the pace of the instructor’s lectures or course syllabus.

The whole idea here is trying to transfer the abstract concept of learning efficiency to some measurable quantities. These kinds of parameters may not be as precise as those used in other research. However, they are quite enough to reveal bottlenecks of learning efficiency. We have noticed following important results:

1. T_{ccd} sets a maximum time limit for delivering a course.
2. T_{ccd} also sets a minimum time limit for mastering course content.
3. T_{ccd} and T_{mli} values depend on the education model used and have nothing to do with the instructor’s teaching quality.
4. T_{cm} depends on both T_{ccd} and the instructor’s teaching quality.
5. Under the above assumptions, for a given learner and a course, $T_{cm} ? T_{ccd}$

C. Implications of T_{ccd} , T_{cm} and T_{mli} Values

There may be many reasons for why a learner cannot learn efficiently. Most people

would agree that the quality of an instructor plays a dominant factor in learners' learning efficiency in a given course. Nevertheless, previous results indicate that T_{cd} plays a part in determining the minimum time required to master the course content. Since that T_{cd} value has nothing to do with the teaching qualities of the course instructor, the following result is surprising:

“Learning inefficiency is in fact an intrinsic property of the traditional education model.”

Assigning a well-qualified instructor to a course is very important. However, since the value of T_{cd} is bound, it seems that the traditional education system has no workable way to bypass this barrier. Also, notice that if we want to design a more efficient learning model, then we have to minimize both the T_{cd} and T_{cm} values.

Some learners learn things fast, while others learn slowly. Remember that the third assumption states that most students study the course according to the pace of their instructor. Therefore, the T_{mli} value is a factor in the speed of learning. In general, a bigger T_{mli} value means that the learning pace is slower, and this will inevitably result in the slowing down of the progress of a fast learner. Therefore, we need a new learning environment that could provide a “true” self-pace learning.

To develop such a new learning environment, we need new concepts, new pedagogy theories, and new models. An easier way to think about the solution of this problem is trying to reduce the T_{cd} and T_{mli} values while still providing good content.

3. T_{CD} , T_{MLI} AND T_{CM} VALUES OF DIFFERENT MODELS

A. Traditional Education Model

In traditional programming language courses, values of T_{cd} and T_{mli} are very easy to compute. Since the value of T_{cm} depends on the individual (educator or learner), it is not as easy to get its exact value. However, the minimum value for average learners is very easy to get (e.g., using the criterion of passing the course).

B Distance Learning Model

Distance learning models do not provide a uniform learning pattern. Therefore, its T_{cd} , T_{mli} and T_{cm} values are highly depend on the course arrangement and the educator's teaching style.

While most distance learning courses provide text-only content, some courses do provide video content. Video content often leads to lengthy delay. Since that the distance learning model is very sensitive to multimedia data's delay, therefore, above parameters are not adequate to reflect its efficiency. More researches need to be carried out.

4. ISSUES RELATED TO MODELS' COMPARISONS

A. Comparing Different Education Models

The education models just provide bare platforms to carry out the teaching activities. This leaves instructors a great flexibility to decide what kind of strategy is to be used. Associating different courses and teaching strategies with different education models results in a large number of possible combinations.

Different courses carry different characteristics. Therefore, instructors use different models and strategies to carry out the teaching activities, such as lecture, discussion, or self-study. Quantifiable parameters introduced above reflect some time limits of the underlying models. Yet, we still don't have the measuring rules to measure those changeable factors. Therefore, any kind of comparison will be criticized as incomplete and subjective.

B Appreciation Instead of Comparison

In order to avoid the incomplete comparisons, we use the method of “appreciation,” instead of the method of “direct comparison.” This method goes like this: we first raise some important issues and then discuss them. Through discussion, we will have a deeper understanding about the problems involved. *At the end of this section, no conclusions will be made. Instead, we will leave readers to judge which model is better.*

C. For a Novice, Which One Is More Important: Direct Interaction or Demo Together With Imitation?

“In the beginning of the newborn baby’s learning process, which one is more important, direct interaction or demo together with imitation?”

Let’s start by looking at the typical growing process of a child. After a baby is born, he spends lots of his time listening and observing Mom’s demonstrations (this is a one way communication). In addition, he tries very hard to imitate Mom’s voice, gestures, and movements. It takes quite a long time to have a “meaningful” interaction with Mom and other people. During these years, he will build up his fundamental linguistic and physical capabilities.

“Do newborn babies ask meaningful questions?” The answer is definitely “No”; the baby just keeps on observing and imitating. These facts lead us to reconsider the necessity of providing “direct interaction” for programming novices. Based on the above observations, it seems that the quest for direct interaction is not always necessary

D. Behind Learners’ Questions

When a learner raises a question, it may not mean that he needs immediate help. For example, a learner may propose a question because he wants to give the instructor a good impression and receive a higher score. On the other hand, the fact that no questions are raised in a class doesn’t mean that everyone understands the content. For example, a learner having trouble with all of the content may not know what to ask.

Therefore, the over emphasis on the total number of questions raised may seriously distract learners’ focusing.

E. Classify Questions Raised

We may roughly classify learners’ questions into two types: trivial and nontrivial. The trivial question is defined informally as those questions that can be answered quickly by a simple demonstration, simple example, or short explanation.

Frequent interactions put a heavy burden on the instructor and novices [6]. Better teaching/learning quality and grading policy will dramatically reduce the number of trivial questions. In programming language course, we ask:

“For novices, is direct interaction more important than imitation and repeated practices?”

F. ALPT: Active Learning and Passive Teaching

In active learning, instructors reserve some course content and pose some questions to their students. Through self-exploration or teams, students answer the questions and understand the course content. The whole learning process involves students’ active learning and frequent interactions. We characterize this kind of teaching strategy as ALPT (“Active Learning and Passive Teaching”).

The ALPT strategy is still widely adopted. Unfortunately, one feature of ALPT strategy is that it often takes much longer than expected. Therefore, in a time of knowledge explosion, we may not be able to train enough qualified person in time using this kind of strategy.

G. BCL and ATPL Concepts

Skilled people and unskilled people look at things differently. A trivial thing for a skilled person may be a big theory for an unskilled person. *In a time of knowledge explosion, if one does not have enough solid background on some specific fields, then “self-investigation” is synonymous with “waste of precious time.”*

While dealing with the knowledge-explosion problem, improving “learning efficiency” is a more constructive strategy. We propose the following BCL (Baby-Centered Learning) concept:

*When facing a new course (topics), let the instructor “thoroughly” study it first. After fully absorbing the course content, the instructor will try to transfer the content to learners in the **shortest time possible**.*

We characterize this new teaching strategy as ATPL (“Active Teaching and Passive Learning”). This strategy emphasizes the instructor’s (thorough) demonstration and explanation and the learner’s repeated imitation. This is very different from the ALPT strategy that emphasizes the learner’s self-exploration.

“Programming learning” is a special kind of “language learning” that requires 100% of correctness. Therefore, any premature question will take up lots of time. We do not neglect the fact that active learning is very important. Instead, we only want to emphasize the fact that in order to learn faster, learners should be equipped with sufficient background prior to starting active learning

H. Active Teaching Leads to Huge Data Size

The adoption of the ATPL strategy in traditional education puts a rather heavy load on the instructor, especially for a large class in which the students’ backgrounds differ widely. Thorough demonstration implies large number of images and clear explanation implies lengthy audio data. Therefore, the adoption of the ATPL strategy leads to a huge volume of course content. In Table 1, some multimedia data sizes of an implementation of C language using the ATPL strategy are provided [5]. Putting such content (Table 1) on a web site would result in lengthy download for even just one topic. Therefore, it is not a good idea.

I. Teaching/Learning Cost

Teaching and learning cost is a crucial issue; yet, it is often omitted in academic discussions. Distance learning is very expensive. Before a distance learning environment can be provided, costly infrastructure needs to be established. In addition, better local equipment is often required in order to get a better connection to the Internet. In distance learning, the total cost for a learner is the cost charged by the content server and the cost of connecting to the web (e.g., telephone bill). For those in poor countries, the expensive infrastructure makes distance learning a dream.

Therefore, in designing a new education

model, the factor of cost should be considered seriously

5. SEARCHING FOR A NEW EDUCATION MODEL

A. Proposing the CD-Type Learning Model

We need a new learning model with the following features:

1. Lower T_{ccd} value.
2. Lower T_{mli} value, say, equal to 0, if possible.
3. Good course content.
4. Low learning cost.

Based on these four features, a new education model, named “CD-type learning model,” is proposed here. It is defined as a software system that satisfies the following conditions:

1. Contains complete pre-recorded and at least one semester’s course content.
2. Course content is well ordered and organized in units, and can be displayed in any order.
3. Course content may not contain video; however, it should contain clear images and audio.
4. Everything is stored on CDs and can be executed without using any browser.
5. Provides the function of replaying.

Condition (1) guarantees that the T_{ccd} value is low. Condition (2) will result in a smooth learning curve.

In the traditional class meeting, learners often look attentively at the blackboard and listen carefully to what the instructor says. If there were nothing on the board, then they would watch the instructor. Therefore, in the class meeting, the content (could be transparencies) and the audio explanations are the most important elements. Therefore, condition (3) would minimize the differences between a real class meeting and a software education environment. Clear images and audio will help novices’ understanding.

Condition (4) makes sure that the learning process isn’t interfered with due to content delay, and also provides a portable learning environment with affordable cost. The differences between pre-recorded content on CDs and downloadable content from a web site are the learners’ affordability (cost

of the infrastructure) and the time needed to access the content (complete the whole downloading). Condition (5) provides a self-pace learning environment.

Since the CD-type learning model contains complete course content, the T_{cd} value depends on the instructor's lecture speed. In general, it is less than 48 hours for a 3-credit hours' course. Moreover, since the $T_{mii} = 0$, the T_{cm} value depends on the quality of course content and the individual learning pace.

One thing we would like to point out here is that the name "CD-type" was coined at the very last stage of the implementation of this new model. The adoption of the CDs simplifies everything and provides better performance.

The major differences between the CD-type model and the conventional used CAI (Computer Assisted Instruction) system are that CAI systems typically only focus on some particular topics and are used as a supplementary learning tool. Whereby, the CD-type model is designed as an alternative to current learning models.

B. Implementation

The proposed CD-type learning model focuses on the issues of high learning efficiency and low learning cost. MPTE system (Multimedia Programming Training Environment) is one of possible implementations based on CD-type model and the ATPL strategy. MPTE integrates the roles played by the instructor, textbooks, and the computer. It provides the following functions:

1. A learning tree is used to give learners an overall view of the course content.
2. Each node of the tree corresponds to a special topic and is associated with a multimedia file, or a source program file together with a comment file.
3. The multimedia file corresponds to instructor's lecture and can be replayed at any time.
4. MPTE is a learning platform. We may link it with a designated compiler or interpreter. If we do so, then MPTE becomes a particular learning environment for that

language.

5. Learners can do more practices within this environment by modifying demo examples and performing more drill problems provided in the system.
6. The node constructed by the course content designer is protected and cannot be modified.

MPTE provides a very fast learning cycle and the cost to a learner is just the price of those CDs'. Therefore, even a low-end PC system will provide an excellent learning environment.

C. Current Status of the MPTE System

Currently, the MPTE system serves as a programming course learning platform running under Windows 95/98/ME systems. The whole system includes the MPTE system and a set of content designing tools. We used Borland Builder C++ to write the source code. It is about 22,000 lines of code. The complete course content is shown Table 2.

6. CONCLUSIONS

Using students' questionnaires/scores to judge education models is too subjective. We need some metric systems to judge the superiority of different education models. Learning efficiency is one criterion to judge the superiority. In this paper, we transform abstract concept of "learning efficiency" to quantifiable parameters. These parameter values reveal some properties of education models. In distance learning models, parameters defined in this paper are not adequate to tell us about their superiority. Yet, measured values lead us to identify learning bottlenecks of the traditional education model, and we are very surprised to notice that the learning inefficiency is in fact an intrinsic property of the traditional education model itself. Parameter values also provided us guidelines for developing better education models.

Constant failures diminish ones' learning momentum very quickly. Therefore, we suggest the strategy of ATPL for novices. *These results might completely contradict our old beliefs. We emphasize the fact that*

learners will learn better if they are equipped with sufficient background and skills before starting active, interactive learning. It is up to the instructor to decide when to start the traditional ALPT strategy. ATPL strategy contradicts conventional pedagogy theories about active learning, and the implementation of the ATPL strategy will result in a huge data size for course content.

All of these results lead to the proposal of the CD-type learning model. Based on this new model, a possible implementation system called MPTE was designed, implemented and tested. MPTE provides a very efficient programming-learning environment. Compared to the distance learning, the CD-type model can save the high cost of the infrastructure as well as decrease the learners' learning cost and time. How to design a good course is an important issue; however, it is beyond the scope of this paper. "Good" course content take time to design, test, and modify. As soon as the content converges, novices will be able to learn programming languages very fast. Until then, programming will no longer be a privileged learning activity.

ACKNOWLEDGMENT

I would like to thank anonymous referees for valuable suggestions.

REFERENCES

- [1] Griffin, R., "Management", Houghton Mifflin, Internet-Based Education, 4th edition, 1992.
- [2] Tschritzis, D., "Reengineering the University", CACM, June 1999, p93-100.
- [3] Macek, T., Manova, B., Kolar, J., "Global Cooperation Project in Computer Programming Course", SIGCSE'99 Proceedings, p208-211....
- [4] Carrasquel, J., "Teaching CS1 on-line: the good, the bad, and the ugly", Proceedings of SIGCSE'99, p212-216.
- [5] Shyu, Y., Chen, P., "MPTE: Multimedia Programming Training Environment", Proceedings of ICMCS'99, Volume I, p822-826....
- [6] Doube, W., "Distance Teaching Workloads", SIGCSE'00 Proceedings, p347-351.
- [7] Anderson, R., Dickey, M., Perkins, H., "Experiences with Tutored Video Instruction for Introductory Programming Courses", SIGCSE'01 Proceedings, p347-351.
- [8] Rapanotti, L., Blake, C., and Griffiths, R., "eTutorials with Voice Groupware: real-time conferencing to support computing students at a distance", ITICSE 2002 Proceedings, p116-120.
- [9] Ari Korhonen, A., Malmi, L., Myllyselkä, P., Scheinin, P., "Does it Make a Difference if

- Students Exercise on the Web or in the Classroom", ITICSE 2002 Proceedings, p121-124.
- [10] Zupancic, B., Horz, H., "Lecture Recording and Its Use in a Traditional University Course", ITICSE 2002 Proceedings, p24-28.
 - [11] McDonald, M., Dorn, B., McDonald, G., "A Statistical Analysis of Student Performance in Online Computer Science Courses", SIGCSE'04, p71-74

Topics	Data size (Bytes)
Turbo C++	57.6 M
Borland C++	36.9 M
Call by Value	120 M
Recursive Function	89.9 M
Linear Sort	71.6 M
Bubble Sort	142 M
Recording Format: 22.05 KHz, 8 bits stereo	

Table 1. Media sizes of some course topics

Title	Disc no.	Size	Lecture length
BASIC	3 discs	1550 MB	393 min
C-Stage 1	3 discs	1514 MB	408 min
C-Stage 2	3 discs	1650 MB	461 min
C-Stage 3	3 discs	1600 MB	450 min
Content of C language is divided into 3 stages.			

Table 2. Sizes of designated course contents

Yuh-Huei Shyu is an associate professor in the Department of Information Engineering, Tamkang University, Taiwan, R.O.C. He has a Ph.D. in Mathematics from Illinois Institute of Technology and a M.S. in Computer Science from University of Arizona. His current interest is in knowledge transfer, especially in the techniques of transferring knowledge in multiple times of learning speed.

Agile Systems for Clinical Research

Ted D. Wade and J.R. Murphy

Abstract— *Reliance on paper research records slows the progress of clinical research. Clinical culture and expensive and/or complex software impede the adoption of electronic methods. Institutions are addressing strategic issues like networks and knowledge bases for broad-based data sharing while ignoring the unmet need for flexible, affordable systems for basic data collection. We are evolving a system to replace paper records with an agile, metadata-based system that we hope will be quickly configured and maintained, readily understood, and self-documenting. We are currently testing the use of metadata-generated web forms to replace a paper/fax front end.*

Index Terms— *Clinical research, metadata, clinical trial, electronic data capture, informatics.*

1. INTRODUCTION

Both medical care and medical research still rely heavily on paper documents [1,2,3]. Compared to other professions there seems to be no concerted professional effort to change to electronic formats. Medicine, like any other profession, has its unique aspects, but the resistance to the use of computers may be rooted in its culture and the shortcomings of computer technology to accommodate this culture [4].

A. Clinical Culture versus Data Quality

Clinicians are trained to be data synthesizers. The differential diagnosis is a Gestalt resulting in extracting the correct patient problem from a myriad of details.

Manuscript received September 28, 2004.
Ted D. Wade is with the National Jewish Medical and Research Center, USA
(e-mail: WadeT@NJC.ORG).

Details that don't agree with the developing pattern are frequently ignored. The important result is the physician's narrative note and not the data elements that support that note. Data quality control is done on these notes and not on individual data items. Current hospital information systems were created to catalogue these narratives for census and billing purposes and not to capture the individual data elements that inform clinical research.

Clinicians' notes and their associated data items are collected together into the medical record, which is a legal document. Despite its fragility paper is seen as the proper way to have these legal documents readily available. Until recently clinicians have had little to no training in how to collect data beyond those items they gather about the subject's medical history and physical examination. Other data is presented to the physicians, and they have very little understanding of its variability or error sources.

Individual clinicians tend to believe that their way of collecting data is the best because it serves their unique way of synthesizing information. It is therefore difficult to get a group of clinicians to agree on a standard way of collecting and dealing with data. Hospital administrators can get what they need for billing and census from current flawed but functional systems, so activities for clinical research data collection have been seriously under-funded. Until recently government funding agencies have felt that data collection should be covered by institutional overhead costs. Since no one in these institutions was a vigorous advocate for using these funds in this manner, medical informatics of this type has languished [5].

B. Institutional Responses

With recent advances in genetics and medical imaging brought about by the computer, clinicians, administrators and granting agencies are starting to think that computation is a necessary tool for medical science and that it could be used in non-administrative areas to great effect. Large, multi-center networks are being created to provide more subjects for clinical trials and disease consortia [6,7]. Federated databases and data mining [8,9] are being considered as a way to find out more about diseases. Data sharing has become a standard feature of research grant announcements [10]. The sheer volume of correct, detailed information being required by government agencies [11] to meet the requirements of doing research on human subjects is moving medicine in the direction of electronic data capture and management [5]. A variety of efforts by government and standards bodies are underway to standardize research vocabularies, metadata, and ontologies [11,12,13].

Much work on the above problems is organized around the computational grid concept [14]. We think that the grid work does not address the clinical research productivity issues described below. These are mostly obstacles to basic data collection. They are small in data volume and not computationally intensive – the opposite of grid problems. Grid database issues concern the integration of heterogeneous, independent databases. Multi-center clinical research tends to need central, standardized data collection.

2. MEETING BASIC NEEDS WITH AGILE SOFTWARE

The medical culture will have to make compromises with the increasing use of computers but computer technology should also harmonize with this culture whenever possible. The paper infrastructure needs to become electronic, in a way that is affordable, is secure, addresses regulatory requirements, and, perhaps above all, is flexible enough to fit into the clinical work environment. We think that the ideal clinical data collection system is *agile*.

A. Agile Characteristics

Clinical research, although it emphasizes plans, protocols, and methodologic consistency, is also a very dynamic endeavor, especially in the early stages of a research project. Our experience in the data analysis stages of many studies underscores the need for a system to be self-documenting and historical in nature. We see many researchers who have limited resources and a strong need for efficient setup and execution of their work, but not at the expense of flexibility.

Once initial planning is complete studies should be quick to set up. Funding for research is subject to institutional and regulatory roadblocks, and so it often arrives late. Researchers need to *get underway fast*.

Data collection should be able to *change as new research ideas occur*, especially during the early or pilot study stage. Medical research exists in a very dynamic environment, where medical knowledge, instrumentation capabilities, patient populations, staff resources, and research objectives all change fast. Additionally, biological phenomena are inherently hard to predict, so that the range of measures of biological responses can only be approximately known before a study starts. And amid all this, researchers, who are inherently creative people, generate many more ideas than they have time or resources to test. Software should not add to this bottleneck.

The subset of clinical research known as clinical trials is heavily regulated. The planning stages of trials are longer, and once underway a trial can be changed very little. The dynamic tools that we advocate would facilitate the planning and setup stages of clinical trials as well as the more fluid networked data collection arrangements of other kinds of clinical research

An agile clinical data system should facilitate cooperation both within and across

institutions. Many studies are now conducted by scientists who are dispersed across multiple locations. This is partly to share clinical or scientific skills, but mostly to draw from a larger pool of patients, so the study can be completed sooner.

An agile data collection system needs short feedback loops for quality control. Clinical research needs to achieve near-perfect accuracy of data collection in a chaotic environment. Data must be verified using poorly structured, easily misplaced, "source documents". One of the main expenses of clinical studies is dealing with the data exceptions, discrepancies, and corrections known, technically, as "queries" [15]. Deciding what the true value of a datum should be increases greatly with the time elapsed from its original generation.

Once you have a data system that is easily modified, then that system should be mostly self-documenting -- including a complete record of historical changes to data definitions. The record of such changes tend to fall by the wayside as research teams rush to adapt the data collection system to the above-mentioned dynamic circumstances. Data that are collected earlier seem not to be comparable to those collected later, and no one remembers why.

An agile system should quickly condition data for the analysis stage of a study. Studies often take longer to complete than planned. Further progress, and especially further funding, is contingent upon completing analysis of data already in hand. Part of the process, traditionally, is "conditioning": transforming data from one format, or formats, to something suitable for standard statistical analysis tools. Such transformations both generate errors and uncover previously un-noticed problems, such as referential integrity, format heterogeneity, and the like. The ideal, agile, system would enforce formats, value ranges, and database integrity up front, and store data in a form close to that usable by analytic software.

An agile system should, like any good software, minimize the training needed to master it. Most people who staff medical research studies also have extensive clinical duties and responsibilities. They resist learning new computer systems because they rightly suspect that another system will not help them get their work done. For these people, the best system should work in ways that are both obvious and self-teaching. There are many principles that can help in designing for ease of use [16]. It is also important to give software the minimum number of features needed to do a job, and to replace "feature creep" with inherent flexibility, which would be defined as the ratio of functional variety relative to the effort in configuring and managing those functions.

B. Existing Solutions

Our Data Coordinating Center's job is to support dozens of research studies -- from large, multi-center trials to internally funded pilot studies -- in collecting data, assuring its quality, and analyzing the data. Studies must cover our costs. Small studies rarely have adequate funding, yet their data needs can be fairly complex. To help them we must be efficient, especially at the low end.

A few years ago we began by trying to improve on the ad hoc methods that are still widely used in academic health centers (AHC's) and Clinical Research Organizations (CRO's) the world over:

- Write data down on more or less refined paper forms.
- Have clerical staff type the data into whatever word processor, spreadsheet or database program is easiest to learn.
- Use batch programming with statistical software to condition data
- Perform statistical analysis

We saw no suitable informatics solution available for our adoption. It is well known among data centers that capital costs of fairly complete commercial systems are too expensive for self-supporting organizations such as ours. There is commercial

competition for supplying the software infrastructure for clinical research, yet adoption is slow and satisfaction is low [3], with most complaints about flexibility, startup time, and service. This situation may reflect the effects of feature creep instead of inherent flexibility. Non-commercial efforts, with rare exceptions [17,18] have been aimed at sharing data [19] across institutional boundaries.

Lacking an outside solution, we started to evolve our own clinical research infrastructure, embodied in software. The work proceeded in several stages, and is still underway.

3. EVOLUTION OF OUR STUDY DESIGNER / MANAGER

A. Stage 1 - Paper/Fax Front End

We first tried to improve the use of paper forms by adapting a commercial forms-processing system with software interpretation of hand-written data. Forms are designed for each study, printed, and then filled out by researchers. The filled forms are scanned or faxed to a central facility, the Data Coordinating Center, where staff verifiers oversee the software interpretation of data and correct errors in that interpretation. The software then transfers the data to a database. Investigators get to retain paper, a familiar medium that fits into their workflow, while the study database is managed centrally. Other stages of data conditioning and analysis follow in the traditional way.

This is a usable hybrid of paper and electronic approaches, with a number of problems. One is that the data acquisition path is one-way: the forms-processing system does not allow direct feedback to the researchers, even though data is acquired fairly quickly. The second problem is that the commercial forms system is essentially closed to metadata. It can neither import nor export information about how data are defined.

It is thus a laborious, manual process to keep congruence between the conditioning and analysis programs, on the one hand, and the forms design and data collection on the other hand. Thirdly, the forms system's data type and range checking are poorly suited for scientific data.

Another set of problems arises in connection with the faxing process - poor images, missing or misplaced pages, and repeated submissions. Even though the software tries to manage these issues, they still cause a lot of work. The problems seem to be inherent in this kind of paper-to-electronic transition

B. Stage 2: Custom-coded Web Front End

For a small clinical trial study we tried replacing some paper forms with hand-crafted web forms that feed the same database. This worked well, but the programming was expensive and hard to change. The web forms were complex because the researchers wanted them to replace "worksheet"-type features of paper documents, such as data accumulation over time and computational aids.

C. Stage 3: Hybrid Paper/Fax Front End with Electronic Management

Our next opportunity to improve our system was when we contracted to coordinate data management in a large five-year study with ten different locations around the United States and England. Instead of using custom-coded web forms we kept the commercial forms package as the data input subsystem. We married this to a database of metadata, with programming to automatically generate system screens and procedures. We wanted quicker feedback to researchers with less programming costs, easier data conditioning and analysis, better documentation of procedures, and improved setup costs. Our approach got us all but the last objective, as we shall explain.

Our system was driven by a metadata database, which we called the *data dictionary*.

- Electronic source (pdf) for the paper forms can be distributed over the web, helping clinical centers to keep up with changes.
- The data feedback loop is shortened because data are available to researchers on the web just as soon as they have passed through our human-supervised automatic recognition process. If our staff verifiers can not determine from the handwritten data what a value is intended to be, they post a question that is automatically passed back to the researchers via the web.
- Reports on the data are relatively easy to design because of the data dictionary. Many types of reports and data snapshots are generated automatically.
- Automatic email alerts about data mean that study oversight staff can go check those data immediately and address any anomalies quickly. This is one of our system's most popular features.
- The automatic audit trail and double-checking of data edits is valuable.
- We use the Secure Socket Layer protocols built into web servers and browsers to meet standards for internet data privacy.
- We also address privacy by minimizing personal identifying information in the research database, while maintaining accuracy by using identifier codes with built-in integrity checks.
- Studies can copy each others' data attribute and form definitions. This looks as if it will become more effective as more studies use the system.
- Our screen designs minimize the number of different types of objects on a display, and explanations of all items are available either next to the object or via a pop-up window.

We viewed the Stage 3 system as part of transitioning from dependence on paper, but we found its hybrid nature was the source of nearly all its *problems*:

- Researchers found it hard to focus on

abstract data definitions, so they gave primacy to the visual appearance of a form. They usually iterated through form designs. With each iteration we had to laboriously make the data dictionary agree with the forms-processing package data.

- There was no way to programmatically make the web data-editing forms resemble the paper input forms, so investigators could not easily visually compare their paper copy to the contents of the central database. This made more work for the clinic data coordinators, who must review data quality by comparing data in the database with the data on their paper forms.
- Study coordinators, who were still focused on the paper forms, wanted to re-fax forms to correct errors instead of going to the web. They also found it hard to track of what had been faxed, and were sometimes not sure if the faxing process had worked. If they did fax the same form again, it was tedious to determine which data values to retain.
- Given the last two problems, some researchers began to ask for primary data collection by web forms, which they realized would have the advantages of a 2-way pipe.

D. Stage 4: Metadata-centered, Agile System

In subsequent stages we shall try to achieve our goals for an agile system with emphasis on the following new features.

Forms generated from data dictionary. We now have 2-way data forms that are generated directly from metadata. To keep software implementation and form maintenance simple, form design is not graphic or layout-centric. Instead it is parameter-based. To change the form you change properties in the data dictionary -- layout parameters, grouping, labeling, explanatory text, field order, and so forth. Our current implementation uses standard html/javascript. A future version might use

This is becoming more or less a standard when a study is archived for its sponsor [11]. But this approach makes it hard to search for something across studies. In our data dictionary we keep, for all studies ever defined in it, all the metadata plus pointers to regulatory documents, analysis programs, and results files. This approach is subject to pointer error when things get moved, but it allows searching and sharing data across studies.

The next step will be to have the data system be fine-grained enough to reflect metadata change during the course of a study. Most change will either be in the storage formats of data items that appear on a case report form, or in the changing definitions of individual data items. Relational databases do not deal with either of these situations. If you change the definition of data in a relational table, the previous state of affairs - different data and different formats - is replaced and thus lost.

A better way would be to include pointers to metadata with the data. With data stored in XML format you could include the name and version number of each data item along with the value of the item. The name and version would point back to the data dictionary where metadata details are kept. Any analytic program would then be able to know exactly how each data value was defined when it was captured. This should allow more powerful procedures for data conditioning and analysis, with better-informed choices about what measurements are comparable in what situations.

4. CONCLUSION

We have argued that very many clinical studies are still going without basic data processing support, mostly for lack of understanding their complex and changing needs. We have developed a core toolset that has promise in meeting those needs. We hope to speed the evolution of an affordable, flexible system by harnessing a cooperative development mode, using open source licensing [23] of our Study Designer / Manager software.

REFERENCES

- [1] 1. Henderson L. (1999) 'Will EDC finally catch on?', *CenterWatch*, Vol. 6(9):1, pp. 5-11.
- [2] Kachnowski, S. (2004) 'Finding a Pathway for EMR and EDC Integration', *Bio-IT World*, <http://www.bioitworld.com/printlinks/article.cfm?ArticleID=3779>
- [3] Uehling, M.D. (2003) 'Data Rapture?' *Bio-IT World*, May 9, <http://www.bioitworld.com/archive/050903/data.html>
- [4] Anderson, J.G., Jay, S.J., Schweer, H.M. and Anderson, M.M. (1986) 'Why doctors don't use computers: some empirical findings.' *Journal of the Royal Society of Medicine*, Vol. 79, pp. 142-144.
- [5] Sung, N.S., Crowley, W.F. Jr., Genel, M., Salber, P., Sandy, L., Sherwood, L.M., Johnson, S.B., Catanese, V., Tilson, H., Getz, K., Larson, E.L., Scheinberg, D., Reece, E.A., Slavkin, H., Dobs, A., Grebb, J., Martinez, R.A., Korn, A., Rimoin, D. 'Central Challenges Facing the National Clinical Research Enterprise' (2003) *Journal of the American Medical Association*, Vol. 289(10), pp. 1278-1287
- [6] 'caBIG: Cancer Biomedical Informatics Grid' (2004), National Cancer Institute, <http://cabig.nci.nih.gov/caBIG/>
- [7] 'NIH Establishes Rare Diseases Clinical Research Network' (2003) *NIH News*, November 3, <http://www.nih.gov/news/pr/nov2003/ncrr-03.htm>
- [8] Biomedical Informatics Research Network (2004) <http://www.nbim.net/index.htm>
- [9] Hastings, S., Gray, M., Kurc, T., Catalyurek, U., Gingerich, T., Young, D. and Salz, J. (2002) 'Grid Support of Collaborative Clinical and Biomedical Research Studies', *Advancing Pathology Informatics, Imaging, and the Internet Conference*, <http://medicine.osu.edu/informatics/DIGC/pubs/Abstract-Kurc.pdf>
- [10] 'Axiopie review of NIH data sharing policy' (2004) Axiopie Scientific Data Management, <http://axiopie.anc.ed.ac.uk/nihstats.html>
- [11] 'Clinical Data Interchange Standards Consortium', (2004) CDISC, <http://www.cdisc.org/>
- [12] 'Cancer Data Standards Repository (caDSR)', (2004), National Cancer Institute, <http://ncicb.nci.nih.gov/core/caDSR>
- [13] 'Unified Medical Language System' (2004) National Library of Medicine, <http://www.nlm.nih.gov/research/umls/umlsmain.html>
- [14] Banks T., Djaoui, A., Parastatids, S., Mani, A., Tuecke, S., Czajkowski, K., Foster, I., Frey, J., Graham, S., Kesselman, C., Maguire, T., Sandholm, T., Snelling, D., Vanderbilt, P., 'Open Grid Service Infrastructure Primer', (2004), Global Grid Federation, <http://www.ggf.org/documents/GWD-I-E/GFD-I.031.pdf>
- [15] 'Clinical Pipelines Clogged with Paper' (2003) *Bio-IT World*, May 9, http://www.bioitworld.com/archive/050903/data_sidebar_2451.html
- [16] Constantine, L.L. (2000) 'What Do Users Want? Engineering Usability into Software', Constantine & Lockwood, Ltd, <http://www.foruse.com/articles/whatusers.htm>.

- [17] Nadkarni, P. and Brandt, C.A. (2004) 'TrialDB: A Generic System for Management of Clinical Study Data', Yale Center for Medical Informatics, <http://ycmi.med.yale.edu/trialdb/>
- [18] Salgado, N.C. and Gouveia-Oliveira, A. (2000) 'Towards a Common Framework for Clinical Trials Information Systems', Proceedings of the American Medical Informatics Association Symposia 2000, pp.754-758
- [19] San Diego Supercomputing Center (2004) 'SDSC Storage Resource Broker' <http://www.sdsc.edu/DICE/SRB/index.html>
- [20] 'CDER Data Standards Manual' (2002) Center for Drug Evaluation and Research, <http://www.fda.gov/cder/dsm/>
- [21] Dubinko, M. (2003) 'Xforms Essentials', <http://dubinko.info/writing/xforms/book.html>
- [22] 'Medical Subject Headings' (2004) National Library of Medicine, <http://www.nlm.nih.gov/mesh/meshhome.html>
- [23] 'Open Source Initiative OSI' (2004) <http://www.opensource.org/>

Reasoning Procedure and Implementation for Logic Programs as Managing Schemes to Extract Demand

Susumu Yamasaki, Kenichi Iwata, and Mariko Sasakura

Abstract— *We apply a nonmonotonic reasoning based on logic programming to the managing scheme to get information on some composition of assumed materials and demanded ones. This scheme can be applied to the infrastructure on the internet by extracting the demand for supply. To extract the demand for a composition, the well-renowned negation as failure rule is available in logic programming. Explicit negation may be included for expressive capabilities, for which some contradiction-removal procedure must be fairly implemented, to avoid the contradiction of a predicate competing with its explicit negation.*

Index Terms—Logic Programming, Managing, Nonmonotonic Reasoning

1. INTRODUCTION

THERE has been a growth of order-made PC compositions to make use of modules to be combined for the PC customer needs. The PC composition should get modules (materials) from the more original producer and make variations of PCs for the customer (consumer).

The composition process is regarded as a methodology, or a management in material compositions for making PCs. The composition induces a managing scheme, being an interface between the producer and the consumer.

For the interface, reasoning is of much use as [3] presents in wide variations of areas. In this paper, a managing scheme with reasoning facilities is examined as regards the PC composition domain. As an automated interface between the producer and consumer, a managing scheme is of use, where the interface is required to reason the process of getting materials from the producer, and composing the materials to supply goods to the consumer. Through the composition, some extraction of demanded materials is to be made by reasoning so that the demand is passed to the producer, and some possible supply is made. The managing of compositions can be organized with the supply goal and inferred demand of materials, which is illustrated in Fig. 1.

Manuscript received November 5, 2004.

Authors are with Department of Intelligence Computing and System, Graduate School of Natural Science and Technology, Okayama University, 31-1 Tsushima-Naka, Okayama, Japan (e-mail: yamasaki@momo.it.okayama-u.ac.jp).

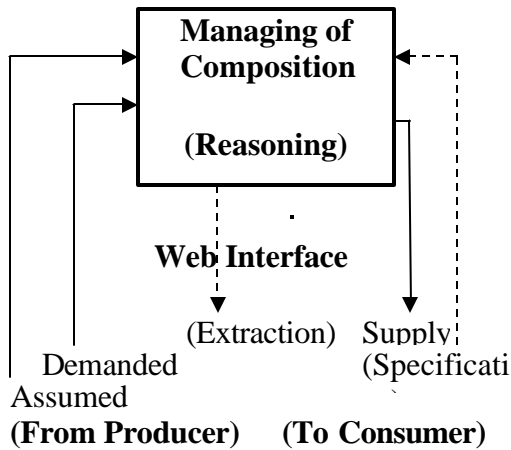


Figure 4 A managing scheme and Web-interface.

The scheme is constructed as a function or a procedure with some reasoning capabilities. For the extraction of demand to be made reasonably, logic programming is effective, because the extraction of demand can be fairly implemented by explanation reasonings, which originated from [5],[8]. As [6] presents the basis, a logical system (a class of extended logic programs) containing two kinds of negation, explicit negation and negation as failure, is regarded as one of standard frameworks with both simplicity and expressive capability. Based on the class, the abduction framework in [4] might offer a good model of managing scheme. With the role of negation as failure to extract information, this paper rather deals with a contradiction-free procedure for an extended logic program with some model theory. For simplicity, we adopt only negation as failure to extraction, apart from the treatment (as in [4]) of both positive abducibles and negative defaults.

To cope with a possible contradiction involved in a managing scheme, we make use of even contradiction-removal techniques as regards explicit negation for expressive facilities. There have been studies on proof procedures for extended logic programs (containing both explicit negation and negation as failure), as in [1],[12],[13], with an introduction of the coherence principle that the explicit negation implies the negation as failure. In accordance with the exception notion of [10]

for removing any contradiction between a pair of a predicate and its explicit negation, a proof procedure is constructed in [18]. It is adopted for implementation. In this paper, a more abstract representation of the proof procedure is given for the reasoning capability. The realization of a managing scheme by means of logic programming is shown in Table 1.

To realize a managing scheme, we deal with logic programming as follows:

- (i) Both explicit negation and negation as failure are treated.
- (ii) A proof procedure is presented, based on negation as failure to extract the demand of subjects. The proof procedure is concerned with dynamically removing the contradiction (through a procedure execution) between a pair of a predicate and its explicit negation. The contradiction removal is required, because the scheme may involve some contradiction.
- (iii) An abstract representation of the proof procedure is given for its demonstration.

Table 1 Realization of Managing Scheme by Logic Programming

A managing Scheme	An Extended Logic Program
Reasoning Facilities for Composition and Extraction	Proof Procedure for Nonmonotonic reasoning with Negation as Failure

2. LOGICAL EXPRESSIONS

D. Notations

We now present the terminology for propositional logic programming.

- (1) A set of symbols to stand for propositions is assumed.
- (2) Two kinds of negation signs are taken: the explicit negation " \neg " and the negation as failure "*not*".
- (3) A literal l is either an atom a or an explicit negation $\neg a$ of an atom a . An atom is an expression consisting of a symbol to denote a proposition.

We deal with the extended logic program. An extended logic program (ELP, for short) is a set of clauses of the form

$$l \leftarrow l_1, \dots, l_m, \text{not } l_{m+1}, \dots, \text{not } l_n \quad (0 \leq m \leq n)$$

where l and l_i are literals, and "not" stands for the negation as failure (NAF). The expression $\text{not } l$ is also called an NAF literal. The literal l of the clause is said to be its head, and the literal sequence $l_1, \dots, l_m, \text{not } l_{m+1}, \dots, \text{not } l_n$ is its body. The explicit negation $\neg l$ means

- (i) $\neg a$ if $a = l$, and
- (ii) a if $l = \neg a$,

for an atom a . The pair of literals a and $\neg a$ is said to be complementary.

The expressions $L, L_1, \dots, L_m, \dots, M, M_1, \dots, M_n, \dots$ are reserved to denote literals or NAF literals. The expressions $\mathbf{a}, \mathbf{a}_1, \dots, \mathbf{a}_m, \dots, \mathbf{b}, \mathbf{b}_1, \dots, \mathbf{b}_n, \dots$ are reserved to denote sequences of literals or NAF literals.

The ELP is said to be a general logic program if it does not contain any explicit negation. For details of ELPs, see [1],[2],[4],[10],[12],[13]. As a means of treating consistency for a pair of a literal and its explicit negation, an exception notion is introduced in [10]. In [1], a proof procedure with coherence principle is given, while some condition for the program to be consistent is shown. In [4], the abduction with the integrity constraint is formulated.

A goal is an expression of the form

$$\leftarrow l_1, \dots, l_m, \text{not } l_{m+1}, \dots, \text{not } l_n \quad (0 \leq m \leq n),$$

where l_i are literals. The goal of the form $\leftarrow \text{not } m_1, \dots, \text{not } m_q$ ($q \geq 0$) is said to be a negative goal. The negative goal is the empty clause, denoted by *empty*, if it contains no literal. Note that the goal $\leftarrow l$ may be interpreted as a proposition $\neg l$. We briefly mention SLD resolution and negation as failure (see [11],[15] for details), in which the reasoning capability consists of SLD resolution and negation as failure in logic programming.

The SLD resolution (deduction) is a

derivation formed by finitely many times applying a rule to infer a goal $\leftarrow L_1, \dots, L_{i-1}, M_1, \dots, M_n, L_{i+1}, \dots, L_m$

from a goal $\leftarrow L_1, \dots, L_{i-1}, l, L_{i+1}, \dots, L_m$ and a clause $l \leftarrow M_1, \dots, M_n$ in the given ELP.

That is, a literal l of a goal may be replaced by the body M_1, \dots, M_n of a clause $l \leftarrow M_1, \dots, M_n$, whose head is just the literal l . Negation as failure is a rule to infer an NAF literal $\text{not } l$, when the literal l is not derived by some proof procedure. We next refine negation as failure in relation to SLD resolution such that the succeeding and failing (derivations) of a goal (with reference to a given ELP) are defined recursively as follows:

- (i) The goal $\leftarrow l$ succeeds if $\leftarrow l$ is reduced to *empty* by applying SLD resolution.
- (ii) The goal $\leftarrow l$ fails if one of the following conditions holds:
 - (a) There is no clause whose head is the literal l .
 - (b) The goals $\leftarrow \mathbf{a}_1, L_1, \mathbf{b}_1, \dots, \leftarrow \mathbf{a}_n, L_n, \mathbf{b}_n$ ($n \geq 1$) are derived by SLD resolution for the goal $\leftarrow l$ such that all the goals $\leftarrow L_1, \dots, \leftarrow L_n$ fail.
 - (c) The goal $\leftarrow \neg l$ succeeds.
- (iii) The goal $\leftarrow \text{not } l$ succeeds if the goal $\leftarrow l$ fails.
- (iv) The goal $\leftarrow \text{not } l$ fails if the goal $\leftarrow l$ succeeds.

Note the sense that if the goal $\leftarrow l$ may succeed, then $\neg l$ may be concluded.

In relation to the definition (ii)(c) above, we will present a method of removing contradictory succeeding derivations of both the goal $\leftarrow a$ and the goal $\leftarrow \neg a$. If both goals may succeed, contradictory conclusions $\neg a$ and a (that is, $\neg \neg a$) are regarded as inferred.

E. Illustration of a Program ELP

We present a simple example as regards managing a PC composition for the conditions.

- (1) The PC consists of modules Processor, HDD, and CD-ROM.
- (2) The module Processor is a Standard-type of 512 MB, or no Standard-type of 256MB.
- (3) A processor ProcessorA is a Standard-type.
- (4) No processor ProcessorB is not a Standard-type.
- (5) The module HDD contains a disk of 80 GB with no case Overcost, or a disk of 40 GB.
- (6) If there is no disk of 40 GB, then it is in the case Overcost.
- (7) The module CD-ROM has a driver DriveA but not a driver DriveB, or is in no case of not containing a driver DriveB.

The conditions are expressed by an ELP :

- (a) $PC \leftarrow Proc, HDD, CD-ROM$
- (b) $Proc \leftarrow M512, Standard$
- (c) $Proc \leftarrow M256, not Standard$
- (d) $Standard \leftarrow ProcA$
- (e) $\neg Standard \leftarrow not ProcB$
- (f) $HDD \leftarrow HD80, Overcost$
- (g) $HDD \leftarrow HD40$
- (h) $Overcost \leftarrow not HD40$
- (i) $CD-ROM \leftarrow DriveA, \neg DriveB$
- (j) $CD-ROM \leftarrow not \neg DriveB$

If there is a processors *ProcA* and no processor *ProcB*, that is, there is a clause $ProcA \leftarrow$ and a clause $\neg ProcB \leftarrow$, there is a contradiction by means of the clauses (d) and (e).

The cases with hard disks of 80 GB and 40 GB are regarded as in "exclusive-or" for the module HDD, that is, one of two clauses $HD80 \leftarrow$ and $HD40 \leftarrow$ is exclusively selected for the clause $HDD \leftarrow$ to be reasoned.

The cases with drivers *DriveA* and *DriveB* are in "exclusive or". Although the clauses (f), (g) and (h) take a different combination with a pair of clauses (i) and (j), we see from the pair clauses (i) and (j) that one of two clauses $DriveA \leftarrow$, and $DriveB \leftarrow$ is exclusively selected for the clause $CD-ROM \leftarrow$ to be reasoned.

3. FORMAL REPRESENTATION OF REASONING PROCEDURE

The managing scheme can be expressed by an ELP with reasoning facilities. The reasonings are presented in [18]. For a brief demonstration, a relational form of the ground version procedure is herein given. We have two sets to be kept, which are to be transformed through succeeding and failing derivations:

- (i) the set of literals to remove contradictory succeeding derivations
- (ii) the set of NAF literals

Note that the first set is in accordance with a semantic view of [18], and the second is primarily a means of information extraction.

The former set is expressed by Σ, Σ_1, \dots , and the second is by Δ, Δ_1, \dots . Given an ELP P , the predicate $suc_p(G \ \Sigma_1 \ \Delta_1 \ \Sigma_2 \ \Delta_2)$ is true, when a goal G succeeds with the assumed sets Σ_1 and Δ_1 to acquire the sets Σ_2 and Δ_2 . The predicate $fail_p(G \ \Sigma_1 \ \Delta_1 \ \Sigma_2 \ \Delta_2)$ is true when a goal G fails with the assumed sets Σ_1 and Δ_1 to acquire the sets Σ_2 and Δ_2 . The set $\Delta_2 - \Delta_1$ is interpreted as extracted. Following the derivations in [18], we have these rules:

- (0) The goal *empty* trivially succeeds.
- (1) SLD resolution is applicable to the succeeding derivation.
- (2) The goal $\leftarrow not a$ succeeds if the NAF literal *not a* is in the assumed set Δ_1 .
- (3) The goal $\leftarrow not a$ succeeds if $not a \notin \Delta_1$ and the goal $\leftarrow a$ fails with the assumed sets Σ_1 and $\Delta_1 \cup \{not a\}$.
- (4) The goal $\leftarrow a$ fails when there is no clause for SLD resolution to apply to.
- (5) SLD resolution is applied for the goal to exhaust the failing (sub-)derivations.
- (6) The goal $\leftarrow \neg a$ fails if the goal $\leftarrow a$ succeeds.
- (7) The NAF literal *not a* is removed from a goal in the failing derivation when the

NAF literal *not a* is included in the assumed set Δ_1 .

- (8) The goal $\leftarrow \text{not } a$ fails when $\text{not } a \notin \Delta_1$ and the goal $\leftarrow a$ succeeds with the assumed set Δ_1 .

We assume an ELP P and we now have the relational representations of succeeding and failing derivations, in which the relations suc_p and fail_p are defined simultaneously by recursion to be the least set satisfying the following closure. The relational representation of derivations for general logic programs is given in 0. The relations demonstrate the implementation as procedural methods for the given ELP. The subscript P for the ELP P may be omitted if it is clear in the context.

- (0) $\text{suc}_p(\text{empty}; \Sigma; \Delta; \Sigma; \Delta)$ for any Σ and Δ .
- (1) $\text{suc}_p(\leftarrow L_1, \dots, L_{i-1}, M_1, \dots, M_m, L_{i+1}, \dots, L_n; \Sigma_1 \cup \{l\}; \Delta_1; \Sigma_2; \Delta_2)$ for $\neg l \notin \Sigma_1$ and $(l \leftarrow M_1, \dots, M_m) \in P \Rightarrow \text{suc}_p(\leftarrow L_1, \dots, L_{i-1}, l, L_{i+1}, \dots, L_n; \Sigma_1; \Delta_1; \Sigma_2; \Delta_2)$.
- (2) $\text{suc}_p(\leftarrow L_1, \dots, L_{i-1}, L_{i+1}, \dots, L_n; \Sigma_1; \Delta_1; \Sigma_2; \Delta_2)$ and $\text{not } l \in \Delta_1 \Rightarrow \text{suc}_p(\leftarrow L_1, \dots, L_{i-1}, \text{not } l, L_{i+1}, \dots, L_n; \Sigma_1; \Delta_1; \Sigma_2; \Delta_2)$.
- (3) $\text{fail}_p(\leftarrow l; \Sigma_1; \Delta_1 \cup \{\text{not } l\}; \Sigma'_2; \Delta'_2)$ for $\text{not } l \notin \Delta_1$, and $\text{suc}_p(\leftarrow L_1, \dots, L_{i-1}, L_{i+1}, \dots, L_n; \Sigma'_2; \Delta'_2; \Sigma_2; \Delta_2) \Rightarrow \text{suc}_p(\leftarrow L_1, \dots, L_{i-1}, \text{not } l, L_{i+1}, \dots, L_n; \Sigma_1; \Delta_1; \Sigma_2; \Delta_2)$.
- (4) There is no clause in P , which contains the literal l in the head $\Rightarrow \text{fail}_p(\leftarrow L_1, \dots, L_{i-1}, l, L_{i+1}, \dots, L_n; \Sigma; \Delta; \Sigma; \Delta)$ for any Σ and Δ .
- (5) For any clause $l \leftarrow M_1^j, \dots, M_{n_j}^j$ in P ($1 \leq j \leq k$) of all the clauses that

contain the literal l in the head,

$$\text{fail}_p(\leftarrow L_1, \dots, L_{i-1}, M_1^j, \dots, M_{n_j}^j, L_{i+1}, \dots, L_n; \Sigma_j; \Delta_j; \Sigma_{j+1}; \Delta_{j+1}) \Rightarrow \text{fail}_p(\leftarrow L_1, \dots, L_{i-1}, l, L_{i+1}, \dots, L_n; \Sigma_1; \Delta_1; \Sigma_{k+1}; \Delta_{k+1}).$$

- (6) $\text{suc}_p(\leftarrow \neg l; \Sigma_1; \Delta_1; \Sigma_2; \Delta_2) \Rightarrow \text{fail}_p(\leftarrow L_1, \dots, L_{i-1}, l, L_{i+1}, \dots, L_n; \Sigma_1; \Delta_1; \Sigma_2; \Delta_2)$.
- (7) $\text{fail}_p(\leftarrow L_1, \dots, L_{i-1}, L_{i+1}, \dots, L_n; \Sigma_1; \Delta_1; \Sigma_2; \Delta_2)$ and $\text{not } l \in \Delta_1 \Rightarrow \text{fail}_p(\leftarrow L_1, \dots, L_{i-1}, \text{not } l, L_{i+1}, \dots, L_n; \Sigma_1; \Delta_1; \Sigma_2; \Delta_2)$.
- (8) $\text{suc}_p(\leftarrow l; \Sigma_1; \Delta_1; \Sigma_2; \Delta_2)$ for $\text{not } l \notin \Delta_1 \Rightarrow \text{fail}_p(\leftarrow L_1, \dots, L_{i-1}, \text{not } l, L_{i+1}, \dots, L_n; \Sigma_1; \Delta_1; \Sigma_2; \Delta_2)$.

We present the soundness of the reasoning procedure for the managing scheme. The extraction of demand may be correctly implemented because of the procedural soundness. The paper [18] suggests a semantics which supports a contradiction-free procedure.

In the following sense, the relation suc_p , which may involve the effect of the relation fail_p , is sound. This is paraphrased to the sense of consistency that if a goal $\leftarrow l$ succeeds, then $\text{not } l$ cannot be included in the set of NAF literals.

Definition 1. For a set Σ , we define the set $\Gamma[\Sigma]$ to be $\{\text{not } \neg l \mid l \in \Sigma\}$.

Note in the following theorem that we regard the set $\Delta_2 \cup \Gamma[\Sigma_2]$ as extracted.

Theorem 1. Assume the predicate $\text{suc}_p(\leftarrow l; \Sigma_1; \Delta_1; \Sigma_2; \Delta_2)$ such that $\Sigma_1 = \Delta_1 = \emptyset$. Then $\text{not } l \notin \Delta_2 \cup \Gamma[\Sigma_2]$.

Proof (Outline) (1) By the definition of the relation suc_p , $\neg l \notin \Sigma_2$. It follows that $\text{not } l \notin \Gamma[\Sigma_2]$.

(2) Contrary to the assumption that

not $l \in \Delta_2$, suppose that $not\ l \in \Delta_2$. Then, by the construction of Δ_2 , $fail_p(\leftarrow l; \Sigma'_1; \Delta'_1; \Sigma'_2; \Delta'_2)$ for $\Sigma'_1, \Delta'_1, \Sigma'_2, \Delta'_2$ such that $l \in \Delta'_1 \subseteq \Delta'_2 \subseteq \Delta_2$. We next examine the cases supporting this relation $fail_p$.

(i) When $suc_p(\leftarrow \neg l; \Sigma'_1; \Delta'_1; \Sigma'_2; \Delta'_2)$, this is a contradiction to the first assumption that $suc_p(\leftarrow \neg l; \Sigma_1; \Delta_1; \Sigma_2; \Delta_2)$.

(ii) When there is no clause whose head is l for the relation $fail_p$, this is a contradiction to the first assumption that $suc_p(\leftarrow \neg l; \Sigma_1; \Delta_1; \Sigma_2; \Delta_2)$ such that there is some clause whose head is l .

(iii) When there is the predicate $fail_p(\leftarrow not\ m_1, \dots, not\ m_n; \Sigma''_1; \Delta''_1; \Sigma''_2; \Delta''_2)$

for some sets $\Sigma''_1, \Delta''_1, \Sigma''_2, \Delta''_2$, which is possibly caused by $\leftarrow not\ m_1, \dots, not\ m_n$ derivable from the goal $\leftarrow l$, it is concluded that $n \neq 0$. Otherwise, $\leftarrow not\ m_1, \dots, not\ m_n$ is empty and it contradicts $fail_p$. For

$\leftarrow not\ m_1, \dots, not\ m_n$, there is some literal $m_i \notin \Delta_1$ ($1 \leq i \leq n$) such that we have $suc_p(\leftarrow m_i; \Sigma''_1; \Delta''_1; \Sigma''_2; \Delta''_2)$, because of $fail_p$. On the assumption that $not\ m_i \in \Delta_2$, we repeat the same discussion. We finally reach the case in which

$fail_p(\leftarrow not\ n_1, \dots, not\ n_k; \Sigma_{f_1}; \Delta_{f_1}; \Sigma_{f_2}; \Delta_{f_2})$

for some sets $\Sigma_{f_1}, \Delta_{f_1}, \Sigma_{f_2}, \Delta_{f_2}$, but there is no literal $not\ n_i$ such that $not\ n_i \notin \Delta_{f_1}$. This causes the case in which $\leftarrow not\ n_1, \dots, not\ n_k = empty$, which is a contradiction to the possible predicate that

$fail_p(\leftarrow not\ n_1, \dots, not\ n_k; \Sigma_{f_1}; \Delta_{f_1}; \Sigma_{f_2}; \Delta_{f_2})$

q.e.d.

4. MANAGING SCHEME WITH WEB-INTERFACE

A. Architecture

The architecture of the presented managing scheme with Web-interface is shown as in Fig. 2. That is, a managing scheme is

$M = (Reasoning\ system, Database, Web)$ where:

(1) The subsystem Reasoning system operates on the program stored in Database and a goal given by a manager so that the reasoning result and the extracted information are provided to the manager. The demand may be communicated to the producer. It is constructed as a Java applet, following the procedure of [14],[18] in accordance with the relational representation of this paper.

(2) The subsystem Database contains an order list of the consumer (customer), a stock list of the producer and the parts information of the manager. Database takes a form of ELP. It is described by XML forms.

(3) Web-interfaces Web are installed for the manager, the consumer and the producer. They are constructed as Java servlets.

B. Implementation illustration

We have an experiment for the ELP which is presented in Section 2. B. Assume that:

- (i) The consumer requests the statement (1) which is translated to the clause (a) in Database.
- (ii) The manager keeps the statements (2)-(7), which are directly translated to the clauses (b)-(j), respectively, in Database.
- (iii) The producer presents the parts *M512*, *HD80*, *DriveA*, but not *DriveB*. They are translated to the clauses :

$M\ 256 \leftarrow; HD\ 40 \leftarrow; DriveA \leftarrow;$

$\neg DriveB \leftarrow.$

The procedure is implemented as follows.

- (i) The goal $\leftarrow PC$ is given by the manager to evoke the reasoning procedure.
- (ii) The next goal is easily obtained: $\leftarrow Proc, HDD, CD-ROM$. The subgoal $\leftarrow Proc$ succeeds, because the goal sequence is formed as below:

5. CONCLUDING REMARKS

$\leftarrow M256, not\ Standard; \leftarrow not\ Standard.$

The goal $\leftarrow Standard$ fails, because of succeeding goal $\leftarrow \neg Standard$. Note that the goal $\leftarrow not\ ProcB$ succeeds for the failure of the goal $\leftarrow ProcB$. Among the extracted, we have *notStandard* and *not ProcB*.

- (iii) As is straightforward, we have got the goal $\leftarrow HDD, CD-ROM$. Then we have the goal $\leftarrow HDD$, because of the clause $HDD \leftarrow$.
- (iv) We now have the goal $\leftarrow CD-ROM$. Because of the clauses $DriveA \leftarrow$ and $\neg DriveB$, $\leftarrow CD-ROM$ succeeds.

We finally have a demand of the processor *ProcB*. The result of reasoning for (i) and (ii) is shown as in Fig.3, where the set Total Sigma \cup Total Delta without negation as failure sign (“wave” in the figure) is regarded as the extracted.

We present a managing scheme as an automated interface between the producer and the consumer. It can be implemented as a Web system on the internet. The scheme involves reasoning facilities, which are caused by an extended logic program.

- (1) A managing scheme of the material compositions can be represented by an extended logic program, in which the reasoning procedure for the extraction may be implemented. This paper presents an abstract representation for the reasoning of a scheme, which differs from an abduction of [4].
- (2) The background of the soundness of succeeding and failing derivations may be given with reference to model theory, following [1], [12], [13], [16], [19]. The first-order version of derivations are presented in [18]. The non-ground extraction is expected in relation to the model theory of [7].

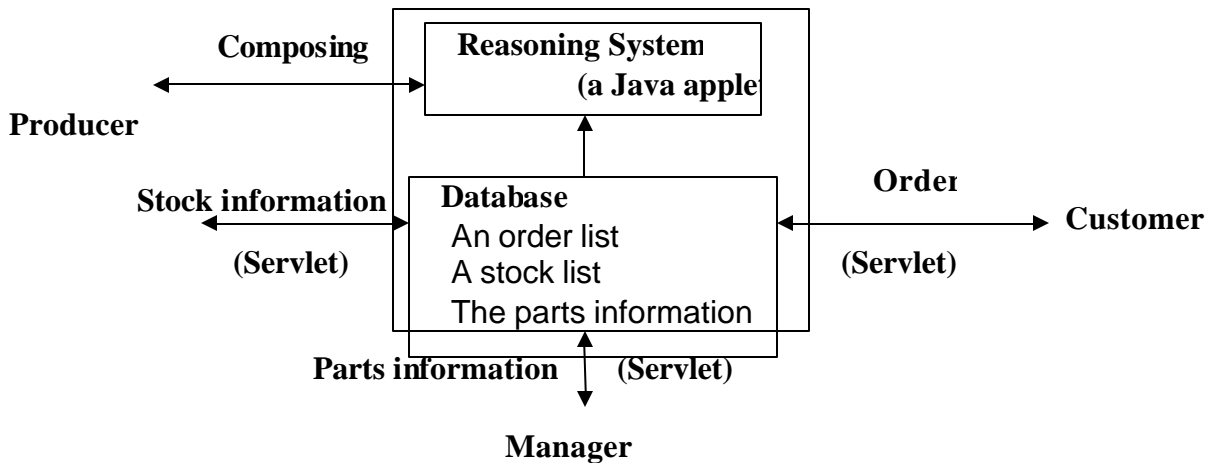


Figure 5 An architecture of the managing scheme.

- (3) The notion of exceptions, as in [10], can be implemented in the derivations we present. We now apply a contradiction-free procedure to the implementation of a managing scheme on the internet, even if the managing scheme faces some contradictory state.

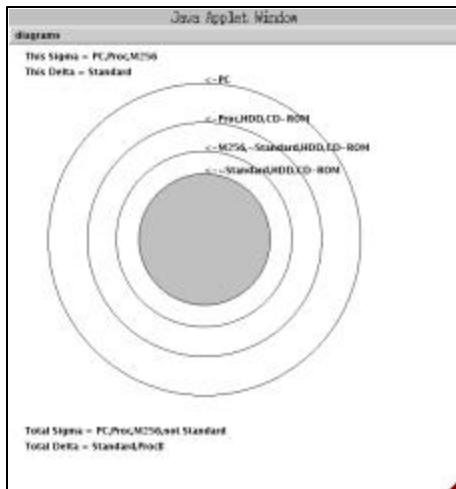


Figure 6 An output figure of the managing scheme.

REFERENCES

- [1] Alferes, J.J., Damasio, C.V. and Pereira, L.M., A logic programming system for nonmonotonic reasoning, *J. of Automated Reasoning*, 14, pp.93-147, 1995.
- [2] Analyti, A. and Pramanik, S., Reliable semantics for extended logic programs with rule prioritization, *J. of Logic and Computation*, 5, pp.303-324, 1995.
- [3] Baral, C., *Knowledge Representation, Reasoning and Declarative Problem Solving*, Cambridge University Press, 2003.
- [4] Brogi, A., Lamma, P., Mancarella, P. and Mello, P., A unifying view for logic programming with non-monotonic reasoning, *Theoretical Computer Science*, 184, 1-2, pp.1-59, 1997.
- [5] Dung, P.M., An argumentation-theoretic foundation for logic programming, *J. of Logic Programming*, 22, pp.151-177, 1995.
- [6] Gelfond, M. and Lifschitz, V., Logic programs with classical negation, *Proc. of th ICLP*, pp.579-597, 1990.
- [7] Gottlob, G., Marcus, S., Nerode, A., Salzer, G. and Subrahmanian, V. S., A non-ground realization of the stable and well-founded semantics, *Theoretical Computer Science*, 166, pp.221-262, 1996.

- [8] Eshghi, K. and Kowalski, R.A., Abduction compared with negation by failure, *Proc. of 6th ICLP*, pp.234-255, 1989.
- [9] Kunen, K., Signed data dependencies in logic programming, *J. of Logic Programming*, 7, pp.231-245, 1989.
- [10] Kowalski, R.A. and Sadri, F., Logic programs with exceptions, *Proc. of 7th International Conference on Logic Programming*, pp.598-613, 1990.
- [11] Lloyd, J.W., *Foundations of Logic Programming*, 2nd, Extended Edition, Springer-Verlag, 1993.
- [12] Pereira, L.M., Alferes, J. J. and Aparicio, J.N., Contradiction removal within well-founded semantics, *Proc. of 1st International Workshop on Logic Programming and Nonmonotonic Reasoning*, pp.105-119, 1991.
- [13] Pereira, P.M., Joaquim, N., Aparicio, J.N. and Alferes, J.J., Non-monotonic reasoning with logic programming, *J. of Logic Programming*, 17, pp.227-263, 1993.
- [14] Sasakura, M. and Yamasaki, S., An explanation reasoning procedure applicable to loop transformation in compiler, *Proc. of ACM ESEC/FSE International Workshop on Intelligent Technologies for Software Engineering, WITSE 03*, pp.34-39, Helsinki, 2003.
- [15] Shepherdson, J.C., Negation in Logic Programming, in: Minker, J. (ed.), *Foundations of Deductive Databases and Logic Programming*, pp.19-88, 1987.
- [16] Van Gelder, A., The alternating fixpoint of logic programs with negation, *J. of Computer and System Sciences*, 47, pp.185-221, 1993.
- [17] Yamasaki, S. and Kurose, Y., A sound and complete procedure for a general logic program in non-floundering derivations with respect to the 3 valued stable model semantics, *Theoretical Computer Science*, 266, pp.489-512, 2001.
- [18] Yamasaki, S. and Sasakura, M., A contradiction-free proof procedure with visualization for extended logic programs, *Proc. of SSGRR02*, 4, 2002.
- [19] You, J.-H. and Yuan, L.Y., On the equivalence of semantics for normal logic programs, *J. of Logic Programming*, 22, pp.211-222, 1995.

Susumu Yamasaki received Eng.D. in information science from Kyoto University in 1980. In 1985/86, he was a visiting fellow at Dept. of Computer Science, University of Warwick, U.K. In 1987, he joined Okayama University as a professor.

Kenichi Iwata is a Ph.D. candidate at Graduate School of Natural Science and Technology, Okayama University.

Mariko Sasakura received Eng.D. in information science from Kyushu University in 2000. She is an assistant professor of Okayama University.

REVIEWERS:

Hojjat Adeli,
Ohio State University, USA

David A Banks,
University of South Australia, Australia

Peter Bellows,
Rockwell Collins Inc., USA

Robert Baumgartner,
Vienna University of Technology, Austria

Peter J Bentley,
University College London, UK

Estela Blaisten-Bajoras,
George Mason University, USA

Marilyn A Campbell,
Queensland University of Technology, Australia

Clara Ceneteno,
Institute for Prospective Technological Studies, Spain

Zachary Coffin,
Coffin Capital, USA

Jeffrey T. Chang,
Duke University, USA

Shu-Ching Chen,
Florida International University, USA

Daniel Leon Chester,
University of Delaware, USA

Alina Chircu,
University of Texas at Austin, USA

Bob Crisp,
University of Arkansas, USA

Michael J. Earl,
University of Oxford, UK

Mohamed Essam,
Khedr University of Ottawa, Canada

Terrence P. Fries,
Coastal Carolina University, USA

John Gill,
Royal National Institute of the Blind, UK

Larry C. Giunipero,
Florida State University, USA

Esther Grassian,
UCLA College Library, USA

Ismet Gucuyener,
Uludag University, Turkey

Werner Haas,
Inst. of Inf. Systems & Inf. Management, Austria

Fadi A. Haddadin,
University of Chicago, USA

Carmen de Pablos Heredero,
Rey Juan Carlos University, Spain

Chuan-Gan Hu,
Nankai University, China

Patrick C. K. Hung,
Institute of Technology, Canada

Colin Johnson,
University of Kent at Canterbury, England

Uzay Kaymak,
Erasmus University Rotterdam, Netherlands

W. B. Langdon,
University College, England

Stavroula Leka,
University of Nottingham, UK

Barbara A. Lewis,
Lamar University, USA

Tao Li,
Azusa Pacific University, USA

Yingjiu Li,
Singapore Management University, Singapore

Marc Light,
University of Iowa, USA

Ioannis Maghiros,
Institution European Commission, Spain

William H. Melody,
Technical University of Denmark, Denmark

Gardoni Mickael,
INPG / GILCO laboratory, France

Artur Mikitiuk,
University of Texas at Tyler, USA

Miguel Mira da Silva,
Instituto Superior Tecnico, Portugal

Paul Murschetz,
University of Cologne, Germany

Robert J Palmer,
Bishop's University, Canada

Sharath Pankanti,
IBM T. J. Watson Research Center, USA

Manish Parashar,
University of New Jersey, USA

Wayne Pease,
University of Southern Queensland, Australia

James I. Penrod,
University of Memphis, USA

Davide Pierattoni,
University of Udine, Italy

Benjamin C. Pierce,
University of Pennsylvania, USA

Axel Polleres,
University of Innsbruck, Austria

Anand Raghunathan,
NEC Labs America, USA

John Reisner,
Air Force Institute of Technology, USA

Tommo Reti,
HIIT, Finland

Elizabeth B. Richmond,
University of Wisconsin-Eau Claire, USA

Luis G. R. Rivas,
Universidad Pontificia Bolivariana, Colombia

Paul C Robins,
Aston University Business School, UK

Horacio Saggion,
University of Sheffield, United Kingdom

Julian Satran,
IBM Research Laboratory, Israel

Gad Shani,
Ben Gurion University, Israel

Fabio Sirocchi,
CINECA, Italy

Oleg Skljarov,
Saint-Petersburg Research Institute, Russia

Ingjerd Skogseid,
Western Norway Research Institute, Norway

Junichi Suzuki,
University of Massachusetts, USA

Richard Szanto,
Corvinus University of Budapest, Hungary

Douglas E. Turner,
State University of West Georgia, USA

Fábio Luciano Verdi,
School of Electrical and Computer Engineering, Brazil

Jelena Vucetic,
Alpha Mission Inc., USA

Yujun Wang,
University of California, USA

James L. Wayman,
San Jose State University, USA

Josh Wiener,
Oklahoma State University, USA

Jeannette Wilhelmy,
University of Cologne, Germany

Welcome to IPSI BgD Conferences and Journals!

<http://www.internetconferences.net>

<http://www.internetjournals.net>

IPSI-2005 HAWAII

Hilton Waikoloa
Big Island, Hawaii, USA
January 6 to 9, 2005

IPSI-2005 AMALFI

Hotel Santa Caterina
Amalfi, Italy
February 17 to 20, 2005

IPSI-2005 FRANKFURT

Schlosshotel Kronberg
Frankfurt, Germany
March 26, 2005

IPSI-2005 FRANCE

Hotel de la Cite,
Carcassonne, France
April 23 to 26, 2005

IPSI-2005 SPAIN

Hostal de la Gavina,
Costa Brava,
near Barcelona, Spain
April 28 to May 1, 2005

IPSI-2005 BELGRADE

Belgrade, Serbia
June 2 to 5, 2005
With Visits to
Famous Serbian Monasteries

IPSI-2005 USA

Hotel@MIT, Cambridge
Massachusetts, USA
July 7 to 10, 2005

IPSI-2005 PESCARA

Hotel Castello Chiola
Loreto Aprutino, Italy
July 27 to Aug 1, 2005

IPSI-2005 STOCKHOLM

Grand Hotel
Stockholm, Sweden
September 24, 2005

IPSI-2005 MONTENEGRO

Hotel Sveti Stefan
October 1 to 8, 2005

Keynote:

Professor de Gennes, Nobel Laureate, France

IPSI-2005 VENICE

Hotel Luna Baglioni
Venice, Italy
November 9 to 14, 2005

IPSI-2005 SLOVENIA

Hotel Toplice
Bled, Slovenia
December 8 to 11, 2005

IPSI-2006 HAWAII

Hilton Waikoloa
Big Island, Hawaii, USA
January 5 to 8, 2006

IPSI-2006 AMALFI

Hotel Santa Caterina
Amalfi, Italy
February 16 to 19, 2006

IPSI-2006 AMSTERDAM

Amsterdam, Holand
March 10 to 12, 2006

IPSI-2006 PARIS

Paris, France
April 22 to 25, 2006

IPSI-2006 MARBELLA

Marbella, Spain
May, 2006

IPSI-2006 BELGRADE

Belgrade, Serbia
June 1 to 4, 2006
With Visits to
Famous Serbian Monasteries

IPSI-2006 USA

Hotel@MIT, Cambridge
Massachusetts, USA
July 6 to 9, 2006

IPSI-2006 PESCARA

Hotel Castello Chiola
Loreto Aprutino, Italy
July 29 to Aug 5, 2006

IPSI-2006 STOCKHOLM

Grand Hotel
Stockholm, Sweden
September 23, 2006

***CIP – Katalogizacija u publikaciji
Narodna biblioteka Srbije, Beograd***

***ISSN 1820 – 4511 = THE IPSI BGD
Transactions on Advanced Research
COBISS.SR – ID 119128844***