

INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

**Anais do 3º Workshop de Teses de Doutorado
em Andamento do IC-UNICAMP**

*Claudia M. Bauzer Medeiros, Alan Massaru Nakai,
Carla Geovana do Nascimento Macario, Diego
Zuquim Guimaraes Garcia, Gilberto Zonta
Pastorello Jr, Jorge Lima de Oliveira Filho, Neumar
Costa Malheiros, Nielsen Cassiano Simões e
Vânia Paula de Almeida Neris (Eds.)*

Technical Report - IC-07-36 - Relatório Técnico

December - 2007 - Dezembro

The contents of this report are the sole responsibility of the authors.
O conteúdo do presente relatório é de única responsabilidade dos autores.

Apresentação

Este relatório técnico contém resumos de 27 trabalhos apresentados no III Workshop de Teses de Doutorado do Instituto de Computação da UNICAMP. O workshop, realizado de 21 a 23 de novembro de 2007, permitiu que doutorandos do Instituto apresentassem os principais aspectos de suas pesquisas. Cada capítulo corresponde a uma apresentação, sendo o texto limitado a 4 páginas. A participação foi voluntária e o perfil acadêmico dos participantes foi variado, cobrindo desde alunos recém-admitidos no programa até aqueles que já tinham defesa marcada em dezembro de 2007. A publicação dos resumos sob forma de um relatório técnico tem por objetivo uma maior divulgação de trabalhos de doutorado em andamento no IC. Além disso, é um registro sucinto do estado de várias dessas pesquisas. Como coordenadora do Workshop, destaco o trabalho dos cinco alunos de doutorado que foram os efetivos organizadores do evento e que são co-editores deste relatório - Alan, Carla, Diego, Gilberto, Jorge, Neumar, Nielsen e Vânia. Ressalto, igualmente, o trabalho de todos os autores de capítulos, que se dispuseram a apresentar trabalhos em andamento e seus orientadores que incentivaram tal participação. Finalmente, agradeço à coordenação de pós-graduação do IC pelo apoio.

Claudia Bauzer Medeiros
Coordenadora do 3º WTD
Professora - Instituto de Computação - UNICAMP

Sumário

1 Workflow-based Sensor Data Management	
Gilberto Zonta Pastorello Jr and Claudia Bauzer Medeiros	5
2 Usando Quadrados Latinos para Colorir Grafos Split	
Sheila Morais de Almeida, Célia Picinin de Mello e Aurora Morgana	9
3 Programação para multi-core e memória transacional	
Alexandro Baldassin e Paulo Centoducatte	12
4 Consistency Priorities Towards a User Oriented Multi-Device Design	
Rodrigo de Oliveira, Heloísa Vieira da Rocha	16
5 O método da geração de colunas aplicado ao problema do m-anéis-estrelados capacitado	
Edna A. Hoshino e Cid C. de Souza	19
6 Classificação de Padrões Utilizando Floresta de Caminhos Ótimos	
João Paulo Papa e Alexandre Xavier Falcão	22
7 Comunicação Multimídia com QoS em Redes de Sensores Sem Fio de Larga Escala	
Cláudio S. de Carvalho e Edmundo R. M. Madeira	24
8 Detecção e Localização de Displasias em Imagens de Ressonância Magnética do Cérebro Humano	
Felipe P.G. Bergo, Alexandre X. Falcão	27
9 Um Método Híbrido de Agrupamento para Recuperação de Informação	
Jurandy Almeida, Siome Goldenstein e Ricardo Torres	30
10 Grafos Pfaffianos e Problemas Relacionados	
Alberto Alexandre Assis Miranda, e Cláudio Leonardo Lucchesi	34
11 Especificação de Modelos de Dados para Bibliotecas Digitais de Objetos Complexos	
Nádia P. Kozievitch e Ricardo Torres	38
12 Design of Flexible User Interfaces	
Vânia Paula de Almeida Neris and Maria Cecília Calani Baranauskas	40

13	Especificação de Mecanismo para Anotação Semântica de Dados Geoespaciais Disponíveis na Web	
	Carla Geovana do Nascimento Macário, Claudia Maria Bauzer Medeiros	44
14	Metodologias de teste para descrições de hardware de alto nível	
	Bruno Albertini, Sandro Rigo e Guido Araujo	47
15	Tempo de Convergência para Equilíbrio de Nash em Jogos Não-Cooperativos	
	André Vignatti e Flávio Miyazawa	50
16	Uplink Scheduling with Quality of Service in IEEE 802.16 Networks	
	Juliana Freitag Borin and Nelson L. S. da Fonseca	53
17	Testes baseados em modelos e injeção de falhas na validação da robustez de protocolos de comunicação	
	Thaise Yano, Eliane Martins e Fabiano Luís de Sousa	57
18	Um modelo para suporte à avaliação da qualidade da educação à distância no ensino superior via Internet	
	Ricardo Luís Lachi e Heloísa Vieira da Rocha	61
19	Redes Mesh Autonômicas	
	Neumar Malheiros, Edmundo Madeira	65
20	Políticas Semânticas para Serviços Web	
	Diego Zuquim Guimarães Garcia e Maria Beatriz Felgar de Toledo	69
21	CoGPlat: Uma Plataforma para e-Government Orientada a Serviços na Web Semântica	
	Ivo José Garcia dos Santos e Edmundo Roberto Mauro Madeira	72
22	Uma Arquitetura para gerência de mobilidade em redes de acesso sem fio heterogêneas	
	Jorge Lima de Oliveira Filho, Edmundo R. M. Madeira	76
23	Métodos Geométricos e Fotométricos para Visão Computacional 3d	
	Rafael Felipe Veiga Saracchini, e Jorge Stolfi	78
24	Geração Automática de Storyboards para Vídeos de Futebol Televisivos	
	Nielsen Cassiano Simões, Neucimar J. Leite	80

25 An Approach for Development of Reactive Systems with Distributed Behavior	
Carlos Neves Júnior and Hans Kurt Edmund Liesenberg	83
26 Dynamic Task Scheduling on Computational Grids	
Luiz Fernando Bittencourt and Edmundo R. M. Madeira	86
27 Balanceamento de Carga para Serviços Web Altamente Disponíveis	
Alan Massaru Nakai, Edmundo Madeira, Luiz Eduardo Buzato	90

1. Workflow-based Sensor Data Management

Authors: Gilberto Zonta Pastorello Jr and Claudia Bauzer Medeiros

The proliferation, variety and ubiquity of sensing devices, from satellites to tiny sensors are making huge amounts of data available. This has opened up new possibilities for us to understand, manage and monitor a given environment, from the small – e.g., a room – to the large – e.g., the planet. However, this has added a new dimension to the classic problem of heterogeneous data management – how to handle increasing volumes of sensing data from a wide range of sensors. Small sensors with communication capabilities present several new research challenges [1–4, 6, 13, 14], giving birth to the so-called sensor networks [5, 15] area. Several issues have still to be tackled in this area, including efficient energy consumption, communication protocols, data processing and transmission, self-regulated organization, and others.

The main goal of this thesis is to specify an infrastructure to support sensor data management, aiming at data publication and usage at large. More specifically, we use scientific data and scientific experiments as a testbed for our solution. This work aims at creating a framework with a threefold solution: (i) using a specific kind of component – DCC – for homogeneous sensor data acquisition; (ii) using scientific workflows for flexible composition of sensor data and manipulation software; and (iii) publication of sensor data tailored to distributed scientific applications.

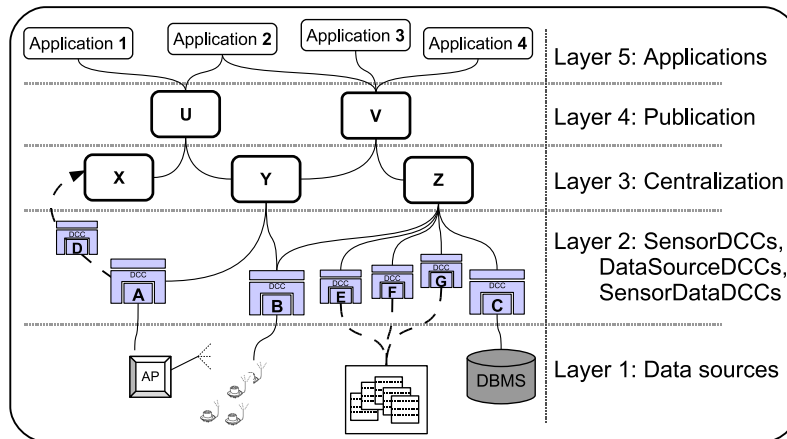


Figura 1: Management layers

Figure 1 outlines our proposal to deal with these issues, where layers denote different data access and manipulation levels, from data sources to applications. For more details the reader is referred to [10]. The continuous lines denote a data flow path and the dotted lines denote a reference to data sources. The first layer contains the data sources such as services, DBMS, files (e.g., text, imagery), and, in particular, sensor and their auxiliary devices. The second layer contains a specific kind of component: *Digital Content Components* (DCC) [12], which have Semantic

Web conformant annotations. Both data and data sources¹ are encapsulated within DCCs, so uniform interfaces are available to the applications, for dynamic (e.g., sensors) and stable (e.g., time series files) sensor data sources; and for dynamic (e.g., stream) and stable (e.g., on image) sensor generated data.

The third layer contains elements that centralize and manipulate the encapsulated sensor data from layer 2, offering functions such as data fusion, summarization, classification and sampling. In this layer, the processing elements can be single DCCs and/or scientific workflows. Workflows are applied to control and compose the basic functions so that the data are tailored to fit applications' needs [9]. The fourth layer has the publication and data access mechanisms, offering high abstraction level interfaces. Applications in Layer 5 are regarded as clients of Layers 3 and 4.

Sensor data have particular requirements to be dealt with, specially concerning data streams manipulation and data fusion schemes. Sensor networks present an even bigger challenge as some of these solutions can be implemented within the network. Our DCC implementations provide solutions to these aspects, including the possibility of using intra and/or extra-network algorithm implementations. The composition of these solutions is done using workflow activities, which are transparently executed by invoking DCC operations.

Figure 2 illustrates the main modules of the architecture. Scientists (the main users) interact with it via the GUI (Graphical User Interface) to design workflows and DCCs, and monitor workflow and/or DCC execution. The architecture relies on the following subsystems: (i) WOODSS [7] – a scientific workflow specification and documentation environment developed at UNICAMP; (ii) ANIMA [12], an infrastructure developed to support DCC execution and management; (iii) a set of modules to design DCCs, eventually reusing and modifying DCCs from a repository [11]; and (iv) a basic infrastructure needed to monitor and execute workflows, including the WFMS. These subsystems (except for the Workflow Execution and Monitoring Modules) were implemented separately.

This proposal has the following advantages (presented in detail in [10]):

- (1) provides homogeneous access to heterogeneous sensing devices;
- (2) enables applications to have multiple views of sensing data, by taking advantage of scientific workflows to mediate sensor data access. This fosters reuse of solutions for managing these data;
- (3) eliminates the need for an application to concern itself with whether a data source is static or dynamic.

Ongoing work involves the implementation of the last two modules of the architecture (Workflow Execution and Monitoring), and the integration of the modules.

¹We distinguish between data – the bits and bytes – and data sources – device or resources that provide data. This distinction can become blurred at times – e.g., a file can be treated either as data or as a data source.

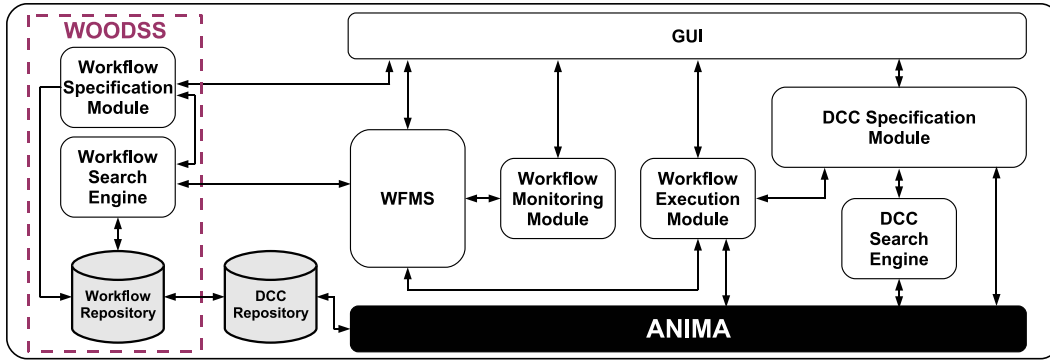


Figure 2: The main modules of the architecture.

We also intend to explore data publication standards – e.g., OGC’s GML [8] – and mapping standards’ terms into ontology terms, enabling the use of semantic annotations with correspondence to the standards.

Referências

- [1] ACM Conference On Embedded Networked Sensor Systems. 2003 and 2004 editions.
- [2] ACM Int. Conference on Mobile Computing and Networking. 1999 to 2004 editions.
- [3] ACM Int. Conference on Wireless Sensor Networks and Applications. 2002 and 2003 editions.
- [4] Com. of the ACM. June 2004 edition – Special Edition on “Wireless sensor networks”.
- [5] J. M. Hellerstein, W. Hong, and S. Madden. The Sensor Spectrum: Technology, Trends and Requirements. *SIGMOD Record*, 32(4):22–27, 2003.
- [6] IEEE Computer. August 2004 edition – Special Edition on “Sensor Networks”.
- [7] C. B. Medeiros, J. J. Pérez-Alcazar, L. Digiampietri, G. Z. Pastorello Jr, A. Santanchè, R. S. Torres, E. R. M. Madeira, and E. Bacarin. WOODSS and the Web: annotating and reusing scientific workflows. *SIGMOD Record*, 34(3):12–23, 2005.
- [8] Open Geospatial Consortium. Geography Markup Language (GML) 3.0. https://portal.opengeospatial.org/files/?artifact_id=7174 (as of Nov 2007).

- [9] G. Z. Pastorello Jr, C. B. Medeiros, and A. Santanchè. Applying Scientific Workflows to Manage Sensor Data. In *Proc. of the 1st e-Science Workshop*, pages 09–18, 2007.
- [10] G. Z. Pastorello Jr, C. B. Medeiros, and A. Santanchè. Providing Homogeneous Access for Sensor Data Management. Technical Report IC-07-012, Institute of Computing, State University of Campinas, May 2007.
- [11] A. Santanchè and C. B. Medeiros. Self Describing Components: Searching for Digital Artifacts on the Web. In *Proceedings of XX SBBD*, 2005.
- [12] A. Santanchè, C. B. Medeiros, and G. Z. Pastorello Jr. User-centered Multimedia Building Blocks. *Multimedia Systems Journal*, 12(4):403–421, 2007.
- [13] SIGMOD Record. December 2003 edition – Special Edition on “Sensor Network Technology and Sensor Data Management”.
- [14] SIGMOD Record. March 2004 edition – Special Edition on “Sensor Network Technology and Sensor Data Management (Part II)”.
- [15] R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, and D. Estrin. Habitat Monitoring with Sensor Networks. *Com. of the ACM*, 47(6):34–40, 2004.

2. Usando Quadrados Latinos para Colorir Grafos Split

Autores: Sheila Morais de Almeida, Célia Picinin de Mello e Aurora Morgana

Seja $G = (V(G), E(G))$ um grafo simples, onde $V(G)$ e $E(G)$ são os conjuntos dos vértices e das arestas de G , respectivamente. Uma *coloração de arestas* de G é uma atribuição de cores às arestas de G de forma que todas as arestas incidentes em um mesmo vértice possuam cores distintas. O *índice cromático* de G , $\chi'(G)$, é o número mínimo de cores necessárias para uma coloração de arestas de G .

É fácil ver que $\chi'(G) \geq \Delta$, onde Δ denota o grau máximo de G . Em 1964, Vizing [9] mostrou que, para um grafo simples, o índice cromático é, no máximo, $\Delta + 1$. Assim surgiu o *Problema da Classificação*, que consiste em decidir, dado um grafo simples G , se $\chi'(G) = \Delta$ ou se $\chi'(G) = \Delta + 1$. No primeiro caso, dizemos que G pertence à *Classe 1*; no segundo, dizemos que G pertence à *Classe 2*.

Mesmo com a poderosa restrição de Vizing, decidir qual o índice cromático de um grafo simples é NP-difícil [5]. Por outro lado, existem classes que tiveram o seu índice cromático determinado: os grafos bipartidos pertencem à Classe 1; os grafos completos e os ciclos sem cordas pertencem à Classe 1 se, e somente se, possuem um número par de vértices. Adicionalmente, Planthold [7,8] provou que grafos com vértice universal ou com vértice quase-universal podem ser classificados em tempo polinomial. Hoffman e Roger [4] produziram resultados semelhantes para grafos multipartidos completos. Além disso, existem resultados parciais para algumas classes, por exemplo, grafos indiferença com Δ ímpar [2], split-indiferença [6], duplamente cordais com Δ ímpar [3] e grafos split com Δ ímpar [1].

Seja G um grafo. Uma *clique* é um conjunto de vértices dois a dois adjacentes em G . Uma *clique maximal* é uma clique que não está propriamente contida em nenhuma outra clique de G . Um *conjunto independente* é um conjunto de vértices dois a dois não adjacentes em G .

Um *grafo split* é um grafo cujo conjunto de vértices pode ser particionado em uma clique e um conjunto independente. O Problema da Classificação ainda não está completamente resolvido para esta classe. Em [1], Chen, Fu e Ko mostraram que grafos split com Δ ímpar pertencem à Classe 1. Nosso interesse é resolver este problema para os grafos split com Δ par. Neste trabalho, usamos quadrados latinos para produzir uma coloração de arestas com Δ cores para alguns grafos split que possuem Δ par.

Um *diagrama de cores* $\mathcal{C} = (C_1, \dots, C_k)$ é uma seqüência de vetores de cores, onde cada vetor de cores $C_i = [c_{i,1}, \dots, c_{i,d_i}]$, $1 \leq i \leq k$, é formado por um conjunto de cores distintas. Um diagrama de cores \mathcal{C} é *monotônico* se a cor $c_{i,j}$ aparece, no máximo, $d_i - j$ vezes antes de C_i , para $1 \leq i \leq k$ e $1 \leq j \leq d_i$.

Um diagrama de cores monotônico pode ser usado para se obter uma coloração de arestas para um grafo bipartido.

Este resultado é apresentado no lema a seguir.

Lema 1 [1] *Seja B um grafo bipartido com bipartição $\{U, V\}$ e considere um diagrama de cores monotônico $\mathcal{C} = (C_{u_1}, \dots, C_{u_k})$. Então B tem uma coloração de arestas que usa as cores de C_{u_i} , para cada $u_i \in U$.*

Uma matriz $k \times k$ cujos elementos são valores do conjunto $\{0, \dots, k-1\}$ é chamada *quadrado latino* se cada elemento de $\{0, \dots, k-1\}$ aparece exatamente uma vez em cada linha e em cada coluna. Um quadrado latino $M = [m_{i,j}]$ é *comutativo* se $m_{i,j} = m_{j,i}$ para $0 \leq i \leq j \leq k-1$.

Considere a matriz $M = [m_{i,j}]$, $0 \leq i \leq k-1$ e $0 \leq j \leq k-1$, tal que $m_{i,j} = (i+j) \pmod{k}$. Note que M é um quadrado latino comutativo e é possível construir um diagrama de cores monotônico, usando os elementos das linhas de M para obter uma seqüência de vetores de cores. Por exemplo, é fácil ver que a seqüência $\mathcal{C} = (C_0, \dots, C_j)$, onde $C_i = [m_{i,j}, m_{i,j+1}, \dots, m_{i,k-1}]$, $0 \leq i \leq j \leq k$, é um diagrama de cores monotônico.

Seja G um grafo split com uma partição $\{Q, S\}$, onde Q é uma clique maximal e S é um conjunto independente. A cada grafo split G podemos associar um grafo bipartido B_G , obtido de G pela remoção de todas as arestas de Q . Seja $d(Q) = \max_{v \in Q} \{d_B(v)\}$. Se $(d(Q))^2 \geq 2|Q| + 1$ e S contém um vértice v cujo grau em G , denotado por $d(v)$, é pelo menos $\frac{|Q|}{2}$, então é possível colorir as arestas de G com Δ cores, usando a matriz A e o diagrama de cores \mathcal{C} devolvidos pelo Algoritmo ColorDiagrams apresentado abaixo.

Algorithm ColorDiagrams($\Delta, |Q|, d(v)$)

Construa uma matriz $M_{\Delta-1, \Delta-1}$ onde $m_{i,j} = (i+j) \pmod{\Delta-1}$.

Construa uma seqüência $\mathcal{C} = (C_0, \dots, C_{|Q|-1})$, onde

$$C_i = [m_{i,|Q|}, \dots, m_{i,\Delta-2}], \quad 0 \leq i < |Q|.$$

Inclua $m_{i,i}$ como primeiro elemento de C_i , $d(v) \leq i < |Q|$.

Inclua $\Delta-1$ como primeiro elemento de C_i , $0 \leq i < |Q|$.

Construa uma matriz $A_{|Q|, |Q|}$, onde $a_{i,j} \leftarrow m_{i,j}$, $0 \leq i, j < |Q|$;

$l \leftarrow 0$; $l' \leftarrow |Q|-1$; $x \leftarrow -1$; $c \leftarrow |Q| + x$;

Se c é ímpar, $count \leftarrow \lfloor \frac{\Delta-|Q|-x-1}{2} \rfloor$;

Se c é par, $count \leftarrow \lfloor \frac{\Delta-|Q|-x-2}{2} \rfloor$;

Enquanto $(l < l')$ e $(c < \Delta-2)$ faça

Substitua a cor c de $a_{l,l'}$ e $a_{l',l}$ na matriz A por $\Delta-1$;

Substitua a cor $\Delta-1$ de C_l e $C_{l'}$ por c ;

$l \leftarrow l+1$; $l' \leftarrow l'-1$; $count \leftarrow count-1$;

Se $count = 0$, então

$x \leftarrow x+1$; $c \leftarrow |Q| + x$;

Se c é ímpar, então $count \leftarrow \lfloor \frac{\Delta-|Q|-x-1}{2} \rfloor$;

Se c é par, então $count \leftarrow \lfloor \frac{\Delta-|Q|-x-2}{2} \rfloor$;

$l \leftarrow l+1$;

Devolva A e \mathcal{C} .

Quando o Algoritmo ColorDiagrams é executado para um grafo split que satisfaz as condições anteriormente citadas, a matriz A e o diagrama de cores \mathcal{C} devolvidos pelo algoritmo possuem as propriedades descritas nos lemas 2 e 3.

Lema 2 *A matriz A devolvida pelo Algoritmo ColorDiagrams é comutativa, seus elementos pertencem ao conjunto $\{0, \dots, \Delta - 1\}$ e em cada linha e cada coluna os elementos são dois a dois distintos. Além disso, se Δ é par, os elementos da diagonal principal de A são dois a dois distintos.*

Lema 3 *Seja G um grafo split com Δ par. Se G tem um vértice v pertencente a S com $d(v) \geq \frac{|Q|}{2}$ e se $(d(Q))^2 \geq 2|Q| + 1$, então a seqüência \mathcal{C} devolvida pelo Algoritmo ColorDiagrams é um diagrama de cores monotônico.*

Estes lemas auxiliam na demonstração do Teorema 4.

Teorema 4 *Seja G um grafo split com Δ par. Se G tem um vértice v pertencente a S com $d(v) \geq \frac{|Q|}{2}$ e $(d(Q))^2 \geq 2|Q| + 1$, então G pertence à Classe 1.*

Referências

- [1] B-L. Chen, H-L. Fu and M. T. Ko. Total chromatic number and chromatic index of split graphs. *J. Combin. Maths. Combin. Comp.*, 17:137–146, 1995.
- [2] C. M. H. de Figueiredo, J. Meidanis and C. P. de Mello. On edge-colouring indifference graphs. *Theoretical Computer Science*, 181:91–106, 1997.
- [3] C. M. H. de Figueiredo, J. Meidanis and C. P. de Mello. Total-chromatic number and chromatic index of dually chordal graphs. *Information Processing Letters*, 70:147–152, 1999.
- [4] D. G. Hoffman e C. A. Rodger. The chromatic index of complete multipartite graphs. *J. of Graph Theory*, 16:159–163, 1992.
- [5] I. Holyer. The NP-completeness of edge-coloring. *SIAM J. Comput.*, 10:718–720, 1981.
- [6] C. Ortiz, N. Maculan and J. L. Szwarcfiter. Characterizing and edge-colouring split-indifference graphs. *Discrete Applied Math.*, 82:209–217, 1998.
- [7] M. Plantholt. The chromatic index of graphs with a spanning star. *Journal of Graph Theory*, 5:45–53, 1981.
- [8] M. Plantholt. The chromatic index of graphs with large maximum degree. *Discrete Applied Math.*, 47:91–96, 1983.
- [9] V. G. Vizing. On an estimate of the chromatic class of a p -graph. *Diket. Analiz.*, 3:25–30, 1964.

3. Programação para multi-core e memória transacional

Autores: Alexandro Baldassin e Paulo Centoducatte

Durante as últimas décadas, o desempenho alcançado pelos microprocessadores cresceu exponencialmente. Esse crescimento deveu-se principalmente ao avanço da tecnologia de fabricação que possibilitou a construção de transistores menores e mais rápidos, além de permitir a adição de mais funcionalidades à microarquitetura. A partir do começo do século 21 esse panorama vem se alterando. Embora o processo de fabricação ainda possibilite transistores mais rápidos e em maior quantidade, seguindo a lei de Moore, fatores novos estão limitando o crescimento do desempenho dos microprocessadores: aumento da potência dissipada, limite na extração de paralelismo no nível de instruções e complexidade de projeto. A tendência atual é desenvolver projetos mais simples, com frequência de operação mais baixa, e integrar em um mesmo *chip* dois ou mais núcleos de processamento, ou seja, hoje predominam os *processadores multi-core* (também conhecidos como *single-chip multiprocessors*) [11].

É importante notar que, até então, o aumento da velocidade de operação dos microprocessadores implicava um aumento proporcional no desempenho do *software*. Com o advento dos processadores *multi-core* isso não é mais verdade. Para fazer uso do potencial do *hardware* atual, o *software* deve ser paralelizado. A implicação decorrente deste fato é que o modelo seqüencial de programação agora deve ser trocado por um modelo paralelo. O maior desafio atual está no desenvolvimento de novos mecanismos para programação concorrente que sejam tangíveis para a maioria dos programadores, e que possam tirar o máximo desempenho possível dos processadores *multi-core*.

O modelo de programação concorrente mais utilizado hoje é o *multi-threaded*. Neste modelo, *threads* concorrentes se comunicam através da escrita e/ou leitura em variáveis compartilhadas. Acessos às mesmas posições de memória precisam ser explicitamente sincronizados para evitar interferências entre as *threads*. Os mecanismos de sincronização atuais empregam bloqueios (*locks*). Sob o ponto de vista dos processadores *multi-core*, bloqueios possuem uma grande desvantagem: limitação de paralelismo. Uma região de código protegida por um bloqueio (região crítica) permite acesso a somente uma *thread* em um determinado instante de tempo, introduzindo um gargalo serial. Qualquer tentativa de acesso à região crítica enquanto ela estiver ocupada implica bloqueio da *thread* solicitante. Para contornar o problema e aumentar o nível de paralelismo, a região crítica pode ser fatorada em regiões menores. Infelizmente esta técnica requer que mais bloqueios sejam utilizados, aumentando o risco de erros de implementação e podendo causar instabilidade no sistema, como *deadlock*.

Um mecanismo de sincronização alternativo que vem despertando bastante interesse por parte de pesquisadores nos últimos anos é a *Memória Transacional*

(*Transactional Memory*) [8]. O modelo transacional usa como base o conceito de *transação* para garantir sincronismo entre *threads* concorrentes. Uma transação é uma seqüência de instruções com garantia de atomicidade e isolamento. Durante sua execução, uma transação armazena localmente os acessos de leitura e escrita feitos aos dados compartilhados e, caso nenhum conflito tenha ocorrido, torna visível instantaneamente suas alterações locais para o restante do sistema. Caso contrário, a transação é cancelada, suas alterações locais descartadas e sua execução reiniciada. Esta característica otimista proporciona que sistemas transacionais explorem mais paralelismo, aumentando seu desempenho e escalabilidade. O uso de Memória Transacional também facilita a programação *multi-threaded* porque o programador não precisa se preocupar em garantir a sincronização como nas abordagens baseadas em bloqueios. Todo o controle de acesso à memória compartilhada é feito automaticamente pelo sistema que implementa Memória Transacional.

A implementação do mecanismo de Memória Transacional pode ser feita exclusivamente em *software* (*Software Transactional Memory*, ou STM) [4, 5, 14–16], ou contar com suporte em *hardware* (*Hardware Transactional Memory*, ou HTM) [1, 3, 6, 10, 12, 13]. Ambas abordagens devem prover: (1) armazenamento especulativo dos dados; (2) detecção e resolução de conflitos entre as transações; e (3) efetivação e cancelamento atômicos. STM e HTM diferem no modo como esses requisitos são implementados: STMs empregam bibliotecas, compiladores e/ou sistemas de execução (*runtime systems*), enquanto HTMs modificam componentes arquiteturais (CPU, cache, barramento, protocolo de coerência). Abordagens híbridas, que empregam ao mesmo tempo STM e HTM, também existem [2, 7, 9].

Referências

- [1] C. S. Ananian, K. Asanovic, B. C. Kuszmaul, C. E. Leiserson e Sean Lie. Unbounded transactional memory. *IEEE Micro*, 26(1):59–69, janeiro/fevereiro de 2006.
- [2] Peter Damron, Alexandra Fedorova e Yossi Lev. Hybrid transactional memory. Em *Proceedings of the 12th International Conference on Architectural Support for Programming Languages and Operating Systems*, pp. 336–346, outubro de 2006.
- [3] Lance Hammond, Vicky Wong, Mike Chen, Brian D. Carlstrom, John D. Davis, Ben Hertzberg, Manohar K. Prabhu, Honggo Wijaya, Christos Kozyrakis e Kunle Olukotun. Transactional memory coherence and consistency. Em *Proceedings of the 31st Annual International Symposium on Computer Architecture*, pp. 102–113, junho de 2004.
- [4] Tim Harris e Keir Fraser. Language support for lightweight transactions. Em *Proceedings of the 18th Annual ACM SIGPLAN Conference on Object-Oriented*

- Programming, Systems, Languages, and Applications*, pp. 388–402, outubro de 2003.
- [5] Maurice Herlihy, Victor Luchangco, Mark Moir e William N. Scherer. Software transactional memory for dynamic-sized data structures. Em *Proceedings of the 22nd Annual Symposium on Principles of Distributed Computing*, pp. 92–101, julho de 2003.
 - [6] Maurice Herlihy e J. Eliot B. Moss. Transactional memory: Architectural support for lock-free data structures. Em *Proceedings of the 20th Annual International Symposium on Computer Architecture*, pp. 289–300, junho de 1993.
 - [7] Sanjeev Kumar, Michael Chu, Christopher J. Hughes, Partha Kundu e Anthony Nguyen. Hybrid transactional memory. Em *Proceedings of the 11th Symposium on Principles and Practice of Parallel Programming*, pp. 209–220, março de 2006.
 - [8] James R. Larus e Ravi Rajwar. *Transactional Memory*. Morgan & Claypool Publishers, 2007.
 - [9] Chi Cao Minh, Martin Trautmann, JaeWoong Chung, Austen McDonald, Nathan Bronson, Jared Casper, Christos Kozyrakis e Kunle Olukotun. An effective hybrid transactional memory system with strong isolation guarantees. Em *Proceedings of the 34th International Symposium on Computer Architecture*, pp. 69–80, junho de 2007.
 - [10] K. E. Moore, J. Bobba, M. J. Moravan, M. D. Hill e D. A. Wood. LogTM: Log-based transactional memory. Em *Proceedings of the 12th International Symposium on High-Performance Computer Architecture*, pp. 254–265, fevereiro de 2006.
 - [11] Kunle Olukotun, Basem A. Nayfeh, Lance Hammond, Ken Wilson e Kunyung Chang. The case for a single-chip multiprocessor. Em *Proceedings of the 7th International Conference on Architectural Support for Programming Languages and Operating Systems*, pp. 2–11, outubro de 1996.
 - [12] Ravi Rajwar, Maurice Herlihy e Konrad Lai. Virtualizing transactional memory. Em *Proceedings of the 32nd Annual International Symposium on Computer Architecture*, pp. 494–505, junho de 2005.
 - [13] Hany E. Ramadan, Christopher J. Rossbach, Donald E. Porter, Owen S. Hofmann, Aditya Bhandari e Emmett Witchel. MetaTM/TxLinux: Transactional memory for an operating system. Em *Proceedings of the 34th International Symposium on Computer Architecture*, pp. 92–103, junho de 2007.

- [14] Bratin Saha, Ali-Reza Adl-Tabatabai, Richard L. Hudson, Chi Cao Minh e Benjamin Hertzberg. McRT-STM: A high performance software transactional memory system for a multi-core runtime. Em *Proceedings of the 11th Symposium on Principles and Practice of Parallel Programming*, pp. 187–197, março de 2006.
- [15] Nir Shavit e Dan Touitou. Software transactional memory. Em *Proceedings of the 14th Annual ACM Symposium on Principles of Distributed Computing*, pp. 204–213, agosto de 1995.
- [16] Fuad Tabba, Cong Wang, James R. Goodman e Mark Moir. NZTM: Non-blocking zero-indirection transactional memory. Em *Second ACM SIGPLAN Workshop on Transactional Computing*, agosto de 2007.

4. Consistency Priorities Towards a User Oriented Multi-Device Design

Authors: Rodrigo de Oliveira, Heloísa Vieira da Rocha

The recent miniaturization of digital components and the vast technological and commercial development have contributed to the emerging of mobile devices (*e.g.* laptops, PDAs, smartphones) considered of imperative need by a great deal of the population. Therefore, realizing one task through different devices has become part of everyday life and one of the biggest challenges in Human-Computer Interaction. Since then, many approaches have been proposed for multi-device design, such as linear transformations to build mobile interfaces from scratch, used by commercial organizations like Avantgo (www.avantgo.com) and UsableNet (www.usablenet.com), or the dynamic and automatic adaptations focusing on the device's constraints [1,2,7]. These and other related approaches certainly have their merit, but the generated interfaces are different from the original in certain aspects that complicate interaction with more than one device to perform the same task, especially when refinding and/or comparing information [6,8]. Many works addressed consistency and continuity problems focusing on user interface generation [3,4] and task migration [10], but their guidelines are still restricted to the device, despite not being sufficiently concrete for an automatic interface framework [9]. A recent proposal [9] addresses the multi-device design problem by passing the control of every appliance to a handheld interface generated automatically. Despite the valued ideas, many device specific interaction types important to each context of use can be lost on the process, besides the need to carry a mobile device to control any appliance.

After a careful review of the main works on multi-device design and also considering what is already known about the importance of the user's cognitive processes on their mental model development, we stated the following hypothesis: *The lack of consistency on task perception and execution of multi-device interfaces has generated ambiguous mental models upon which is difficult to make predictions about the effects of activating a control mechanism towards a task execution.*

This assumption is based on an understanding of consistency that is user oriented. In this approach, we suggest consistency as a practice towards the users' expectations, and not to fill requirements of different devices and contexts' constraints. In a more pragmatic way, consistency should be provided on the user's mental model update cycle, which in our conception, has at least four stages. On a first instance, the user scans the interface and infers the most appropriate control mechanism to perform the given task. On the next step, after perceiving the task, he/she actually execute it by activating the control mechanism. The third phase is dependent on the system's feedback, when the user perceives results and acts according to it. If it was expected, the fourth stage may store this information as a direct confirmation using short-term dynamics or adapt the whole structure to embody the new situation using long-term dynamics [12]. However, if the feedback wasn't

expected, the structure of this user's model will have to change and the information stored may generate new ambiguities in the future. According to [11], inaccurate mental models of more complex systems, such as an airplane or nuclear reactor, can lead to disastrous accidents.

In order to help users form an accurate and useful mental model of a system while interacting with any of its interfaces, we suggest applying consistency on multi-device contexts using the following priorities:

1. *Task Perception* - the same control mechanisms to execute a task and their disposal on the interface. If these requirements cannot be followed with good usability on devices with different control attributes (*e.g.* size, weight, material, etc.) and properties (*e.g.* fluidity, flexibility, opacity, etc.), perception should be mapped to each device's interaction type maintaining usability.
2. *Task Execution* - the same actions flow to execute a task. If the control mechanisms available on a given interface had to be adapted for the others by the task perception priority, the actions flow should be maintained on a logical perspective. Although this may repress bad design decisions and lose opportunity to improve usability on each interface independently, user's decision making is supported under a consistent multi-device context, providing ease of learning/remembering and safety of use. Next priority improves efficiency.
3. *Task Personalization* - the ability to change task perception and execution according to the users' preferences. The goal is to achieve the best design for any user which is the configuration that user expects. This can be related to the *personally consistent design* concept [9], but with an active position for the user. As a result, efficiency and ease of expert use is provided to avoid the downsides of consistency [5].

It's important to understand the correct application of these consistency priorities as they can be easily misunderstood. For example, if an individual wants to check an account balance through an ATM machine, a tablet PC and a telephone, there is no possibility to perceive and perform the task in the same way. If the mentioned devices are important to the end user, adaptation to the contrasting interaction types is a price they are willing to pay. In fact, that's part of the *task personalization* priority. So the focus is to provide the same task perception and execution under a logical perspective, be it through words typed, written or said.

The evaluation process is being conducted through the application of this proposal to the Summary Thumbnail [6], a prototype designed to automatically adapt desktop web pages for handhelds. Some issues on this project could be identified and fixed using the consistency priorities. First prototypes were informally tested on a few institutions with much better impressions than the awarded commercial solution by Opera (www.opera.com/products/mobile/reviews). Next prototypes are still under development and user evaluations will be taken with them to verify the advantages of this proposal.

Referências

- [1] S. Berti, F. Correani, G. Mori, F. Paternò, and C. Santoro. Teresa: a transformation-based environment for designing and developing multi-device interfaces. In *CHI Extended Abstracts*, pages 793–794, 2004.
- [2] K. Coninx, K. Luyten, C. Vandervelpen, J. V. den Bergh, and B. Creemers. Dygimes: Dynamically generating interfaces for mobile computing devices and embedded systems. In *Mobile HCI*, pages 256–270, 2003.
- [3] C. Denis and L. Karsenty. Inter-usability of multi-device systems - a conceptual framework. In A. Seffah and H. Javahery, editors, *Multiple User Interfaces: Cross-Platform Applications and Context-Aware Interfaces*, pages 373–384. John Wiley and Sons, 2004.
- [4] M. Florins, D. G. Trevisan, and J. Vanderdonckt. The continuity property in mixed reality and multiplatform systems: a comparative study. In *CADUI*, pages 321–332, 2004.
- [5] J. Grudin. The case against user interface consistency. *Commun. ACM*, 32(10):1164–1173, 1989.
- [6] H. Lam and P. Baudisch. Summary thumbnails: readable overviews for small screen web browsers. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 681–690, New York, NY, USA, 2005. ACM Press.
- [7] J. Lin. *Using Design Patterns and Layers to Support the Early-Stage Design and Prototyping of Cross-Device User Interfaces*. PhD thesis, University of California, Berkeley, 2005.
- [8] B. MacKay, C. R. Watters, and J. Duffy. Web page transformation when switching devices. In *Mobile HCI*, pages 228–239, 2004.
- [9] J. Nichols. *Automatically Generating High-Quality User Interfaces for Appliances*. PhD thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania, December 2006.
- [10] P. S. Pyla, M. Tungare, and M. Pérez-Quñones. Multiple user interfaces: Why consistency is not everything, and seamless task migration is key. In *Proceedings of the CHI 2006 Workshop on The Many Faces of Consistency in Cross-Platform Design.*, 2006.
- [11] J. Reason. *Human Error*. Cambridge University Press, Cambridge UK, 1990.
- [12] Y. Waern. On the dynamics of mental models. In *Informatics and Psychology Workshop*, pages 73–93, 1987.

5. O método da geração de colunas aplicado ao problema do m -anéis-estrelados capacitado

Autores: Edna A. Hoshino e Cid C. de Souza

O *problema de roteamento de veículos*, *vehicle routing problem* (VRP), introduzido por Dantzig e Ramser [2] está relacionado a aplicações em que determinadas mercadorias, localizadas em um *depósito*, devem ser distribuídas em pontos dispersos geograficamente, chamados *clientes*. Em tais aplicações, cada cliente precisa de uma demanda fixa da mercadoria e uma frota de m veículos idênticos localizados no depósito está disponível para efetuar a distribuição das mercadorias e, assim, atender as demandas dos clientes. Em geral, os veículos possuem uma capacidade limitada de carga, o que restringe o número de clientes que um veículo pode visitar. Neste contexto, o termo *rota* é usado para designar a seqüência de clientes visitados por um veículo. O objetivo do VRP consiste em determinar m rotas disjuntas nos clientes, uma para cada veículo, de modo a atender todas as demandas dos clientes e minimizar os custos das rotas. VRP é NP-difícil, uma vez que generaliza o problema do caixeiro viajante, *Traveling Salesman Problem* (TSP), um clássico problema NP-difícil. A grande aplicabilidade do problema deu origem a diferentes variantes. Quando os veículos têm capacidade limitada de carga, o problema é conhecido por *capacitated vehicle routing problem* (CVRP). Em determinadas aplicações, os clientes podem estar muito dispersos geograficamente de tal modo que torna-se inaceitável, do ponto de vista dos custos das rotas, atender todos os clientes. Nestas situações, pode ser mais viável traçar rotas que não passam por todos os clientes e dispor de um serviço de entrega que distribui a posteriori mercadorias de um ponto da rota até os clientes que ficaram fora das rotas. Exemplos de tais aplicações incluem logística de distribuição de jornais, fretamento de veículos, projeto de redes de fibras óticas, entre outros.

O *problema do m -anéis-estrelados capacitado* (mCRSP) pode ser visto como essa variante do VRP, na qual se admitem pontos intermediários com demanda nula, chamados *nós Steiner*, a ocorrerem nas rotas e permitir que clientes não sejam atendidos diretamente por nenhuma das rotas. Neste caso, cada um de tais clientes tem sua demanda atendida indiretamente associando-o a um dos pontos da rota. Esta associação, chamada *conexão*, implica num custo, que do ponto de vista da aplicação pode ser visto como o custo de deslocar as mercadorias de um ponto da rota até o cliente não visitado. No contexto de mCRSP, uma rota é chamada *anel*, o conjunto das conexões até uma rota é denominado *estrela* e as duas juntas formam uma configuração que lembra a topologia estrela e é designada por *anel-estrelado*. Como no CVRP, os veículos possuem uma capacidade limitada de carga, mas desta vez restringindo o número de clientes que um veículo pode atender direta e indiretamente. O objetivo do mCRSP consiste em encontrar m anéis-estrelados disjuntos nos vértices, um para cada veículo e não violando sua capacidade, para atender as demandas dos clientes e minimizando os custos das rotas e das conexões.

O mCRSP é NP-difícil uma vez que ele é uma generalização do CVRP. Este problema foi introduzido por Baldacci, Dell’Amico e Salazar [1] como um modelo para aplicações no contexto de redes e telecomunicações, como o projeto de redes de fibras ópticas. Duas formulações de programação linear inteira baseadas na formulação *de dois índices e fluxos de duas mercadorias* do VRP [9] foram propostas para modelar o mCRSP. Pelo nosso conhecimento, o único algoritmo exato para resolver este problema é um algoritmo *branch-and-cut* proposto no mesmo artigo em que o problema foi introduzido. Embora este algoritmo exato tenha apresentado bons resultados computacionais, classes de instâncias com 76 e 101 vértices não puderam ser resolvidas na otimalidade em até 7200 segundos.

Neste trabalho, uma formulação por cobertura de conjuntos para modelar o mCRSP é discutida e diferentes implementações de um algoritmo *branch-and-price* deste modelo são avaliadas. Modelos de cobertura de conjuntos, em geral, apresentam limitantes duais mais apertados que os de outros modelos, como já reportado na literatura para o VRP [9].

Entretanto, estes modelos tem o inconveniente de possuírem um número exponencial de variáveis, o que exige que estas sejam geradas ao longo do algoritmo e apenas quando se mostram interessantes para obtenção da solução ótima. Esta técnica é conhecida como geração de colunas, ou *branch-and-price* quando embutida em um algoritmo enumerativo do tipo *branch-and-bound*, e requer a resolução de um subproblema que, no caso do mCRSP é NP-difícil. Problemas correlatos a este incluem o problema do caixeiro viajante com coleta de lucros [3], o problema anel-estrelado de Steiner [8] e o problema do ciclo recobridor [5] (*covering tour problem*). No entanto, nenhum deles trata exatamente o subproblema por não considerar pesos nas conexões ou capacidades dos veículos. Para contornar a dificuldade do subproblema, uma relaxação do subproblema é considerada, usando-se uma idéia similar ao que foi usado com sucesso para o problema de roteamento de veículos com janelas de tempo [7] e CVRP [4]. A relaxação consiste, basicamente, em permitir a repetição de vértices ao longo dos anéis-estrelados. Uma rotina de programação dinâmica foi projetada para resolver, em tempo pseudopolinomial, o subproblema relaxado. Uma vez que tal relaxação pode levar a limitantes duais fracos, determinadas estruturas, tais como os k -ciclos [6], são proibidas para apertar a relaxação. Diferentes estruturas foram avaliadas para esse propósito.

Comparativos entre as diferentes implementações *branch-and-price* com respeito aos limitantes duais, gap de otimalidade e tempo de processamento foram realizados. Os resultados foram confrontados com aqueles obtidos por uma implementação *branch-and-cut* proposta por Baldacci et al.

Referências

- [1] Baldacci, R., M. Dell’Amico, and J.J. Salazar, *The Capacitated m -Ring Star Problem*, University of Modena and Reggio Emilia, Technical Report 42 (2003).

- [2] Dantzig, G.B., and J.H. Ramser. The truck dispatching problem. *Management Science*, 6:80–91, 1959.
- [3] Feillet, D., P. Dejax, and M. Gendreau, *Traveling salesman problems with profits*, *Transportation Science*, **39(2)** (2005), 188–205.
- [4] Fukasawa, R., H. Longo, J. Lysgaard, M.P. de Aragão, M. Reis, E. Uchoa, and R.F. Werneck, *Robust Branch-and-Cut-and-Price for the Capacitated Vehicle Routing Problem*, *Mathematical Programming*, **106(3)** (2006), 491–511.
- [5] Gendreau, M., G. Laporte, and F. Semet, *The Covering Tour Problem*, *Operations Research*, **45 (4)** (1997), 568–576.
- [6] Irnich, S., G. Desaulniers, *Shortest path problems with resource constraints*, 2004. Unpublished manuscript available at http://www.dpor.rwth-aachen.de/de/publikationen/pdf/or_2004-01.pdf.
- [7] Irnich, S., D. Villeneuve, *The shortest path problem with resource constraints and k -cycle elimination for $k \geq 3$* , *Inform Journal on Computing* **18**, issue 3, 2006.
- [8] Labbé, M., G. Laporte, I. Rodríguez Martín, and J.J. Salazar González, *The ring-star problem: Polyhedral analysis and exact algorithm*, *Networks*, **43** (2004), 177–189.
- [9] Toth, P., and D. Vigo, “The vehicle routing problem”, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2001.

6. Classificação de Padrões Utilizando Floresta de Caminhos Ótimos

Autores: João Paulo Papa e Alexandre Xavier Falcão

Algoritmos de classificação de padrões são divididos em categorias conforme a sua informação *a priori* sobre o conjunto de dados rotulados. Metodologias não supervisionadas não possuem conhecimento algum sobre os rótulos (classes) dos dados, sendo seu principal objetivo encontrar uma função que melhor agrupe os mesmos, de tal forma que amostras "semelhantes" pertençam a um mesmo agrupamento. Tais metodologias não garantem que amostras pertencentes a um mesmo agrupamento possuam a mesma classe. Algumas aplicações disponibilizam informação parcial a respeito do conjunto de dados, ou seja, conhecemos os rótulos de apenas algumas amostras de um conjunto de treinamento. Neste caso, temos as técnicas de classificação semi-supervisionadas, as quais baseiam-se em amostras rotuladas e não rotuladas. Porém, estas técnicas tratam de resolver o problema localmente, pois trabalham na tarefa de rotular um conjunto de teste específico, ao contrário das metodologias supervisionadas, as quais possuem total conhecimento sobre seu conjunto de treinamento, possuindo como objetivo encontrar uma função que melhor dicotomize o espaço de características para qualquer amostra do conjunto de teste e para futuras amostras. Contudo, as abordagens supervisionadas atuais possuem certas limitações. Métodos supervisionados (árvores de decisão, por exemplo) quando treinados para obter erro zero no conjunto de treinamento, normalmente apresentam baixo desempenho no conjunto de teste. Este problema é conhecido como *overfitting*. A maioria dos métodos, porém, nem garante erro zero no conjunto de treinamento. Supondo que este deveria representar as diversidades das amostras de teste, não seria um indicativo subótimo para o desempenho no conjunto de teste? Além da robustez, o estado da arte em métodos supervisionados (Redes neurais e SVM) assume que as classes são separáveis no espaço de atributos ou em algum espaço estendido, o que nem sempre é verdade.

Assim sendo, propomos um método para projeto de classificadores supervisionados que permite tratar classes não separáveis, eliminar *outliers* no projeto do classificador e obter erro zero no conjunto de treinamento. Ademais, problemas de análise de imagens podem requerer a descrição dos objetos por múltiplos descritores de forma, cor e textura. O *framework* proposto permite combinar estes descritores com Programação Genética. Nossa abordagem é baseada em um grafo completo, onde as amostras de treinamento são os nós e uma métrica de distância entre os vetores de atributos das amostras define o peso dos arcos. O classificador constrói uma floresta de caminhos ótimos neste grafo com raízes em amostras protótipos, onde uma amostra de teste é rotulada pelo protótipo que lhe oferece o caminho de custo mínimo. As amostras podem ser *pixels*, regiões e contornos de imagens, sendo que toda classe deve ter ao menos um protótipo e uma mesma classe pode ser representada por vários protótipos. O classificador de floresta de caminhos ótimos (OPF

- *Optimum Path Forest*) [1] utiliza o algoritmo IFT [2] (*Image Foresting Transform*) no espaço de atributos, e cada protótipo se torna uma raiz de uma árvore de caminhos ótimos neste espaço onde as amostras de uma árvore são mais fortemente conexas a sua raiz do que a qualquer outro protótipo. Os protótipos são escolhidos de forma a gerar erro zero de classificação no conjunto de treinamento. Estes importantes elementos são as amostras mais próximas de diferentes classes, sendo encontradas a partir de uma MST (Árvore de Espalhamento Mínima - *Minimum Spanning Tree*) no grafo completo. Se encontrássemos o conjunto de protótipos ideal, não teríamos também erros de classificação no conjunto de teste. Isto vai depender da diversidade das amostras de treinamento em representar o problema. Frequentemente, as aplicações limitam o conjunto de treinamento dificultando a obtenção de protótipos que realmente representam o problema. Estas situações são contornadas com um algoritmo de aprendizado sobre o conjunto maior de amostras de avaliação rotuladas. Comparamos os resultados preliminares em três bases (duas sintéticas e uma real) com relação a uma Rede Neural Multicamadas (ANN-MPL - *Artificial Neural Network-Multilayer Perceptron*) [4] e com Máquinas de Vetores de Suporte (SVM - *Support Vector Machines*) [3]. O OPF demonstrou superioridade em todas as bases testadas, tanto em relação a eficiência quanto em efetividade.

Referências

- [1] J. P. Papa and A. X. Falcão and P. A. V. Miranda and C. T. N. Suzuki and N. D. A. Mascarenhas. Design of robust pattern classifiers based on optimum-path forests. *Proceedings of the 8th Int. Symposium on Mathematical Morphology*, 337–348, 2007.
- [2] A.X. Falcão and J. Stolfi and R.A. Lotufo. The image foresting transform: theory, algorithms, and applications, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, 19–20, 2004.
- [3] B.E. Boser and I.M. Guyon and V.N. Vapnik. A training algorithm for optimal margin classifiers. *Proceedings of the 5th Workshop on Computational Learning Theory*, 144–152, 1992.
- [4] S. Haykin *Neural networks: a comprehensive foundation*. Prentice Hall, 1994.

7. Comunicação Multimídia com QoS em Redes de Sensores Sem Fio de Larga Escala

Autores: Cláudio S. de Carvalho e Edmundo R. M. Madeira

Prover comunicações multimídia eficientes em redes de sensores sem fio (RSSF) é uma tarefa complicada devido a diversos fatores. Por exemplo, enquanto as comunicações multimídia demandam uma alta largura de banda e requerem baixos atrasos fim-a-fim na entrega de pacotes, os nós sensores possuem restrições de largura de banda, memória, processamento e energia que dificultam a oferta de tais requisitos de qualidade de serviço (QoS). Prover suporte a esta comunicação multimídia eficientes na camada de rede é uma área de pesquisa recente que apresenta vários desafios ainda pouco explorados [2, 4, 6].

As aplicações para redes de sensores sem fio multimídia (RSSFM) apresentam contribuições significativas em diversas áreas. Em algumas delas, tais como o monitoramento de habitat de animais e o controle de congestionamento de veículos, é necessário centenas de nós sensores para cobrir a região a ser monitorada. No entanto, não foi encontrado na literatura nenhum protocolo de roteamento para o provisionamento de comunicações multimídia em RSSFM de larga escala. Neste trabalho propomos um protocolo de roteamento para tal finalidade.

Tipicamente, uma rede de sensores possui somente um ponto de acesso (*sink*) para onde os dados coletados são enviados. No caso de uma rede de sensores de larga escala é importante que a rede tenha múltiplos pontos de acesso (*multisink*) para permitir que comunicações entre nós distantes sejam otimizadas e que a carga da rede seja balanceada entre os pontos de acesso [9]. Além de considerar a existência de múltiplos pontos de acesso na rede, o protocolo de roteamento proposto deve incluir mecanismos capazes de evacuar grande volume de dados e prover os requisitos de QoS das comunicações multimídia.

Os nós sensores utilizam, tipicamente, a tecnologia ZigBee para transmissão de dados na rede. Um único canal de transmissão ZigBee pode alcançar no máximo 250kb/s [5], no entanto, esta largura de banda é insuficiente para transmissão de um fluxo multimídia na rede. Para cobrir este problema, o algoritmo de roteamento a ser desenvolvido deve apresentar capacidades para encaminhar os fluxos da rede utilizando várias rotas que conectam a origem ao destino. A tolerância a falhas a ser considerada no protocolo proposto está na replicação parcial dos dados transmitidos nas rotas. Ambas as soluções consideradas são utilizadas em [6], no entanto, tal trabalho não considera o uso de múltiplos pontos de acesso.

Os atrasos inseridos em uma comunicação multimídia estão relacionados com diversos fatores como a distância a ser percorrida pelos dados, enfileiramento e processamento de dados nos nós sensores, e mecanismos para construção de rotas e transmissão de dados. A seguir são discutidos a abordagem do protocolo proposto sobre alguns destes fatores.

A transmissão de dados em múltiplas rotas contribui para a redução do atraso na

transmissão. Eren Gürses et al [7] sugerem o uso de rotas proativas (pré-calculadas) como um elemento importante para cumprir os requisitos de atraso das comunicações multimídia. Embora o consumo de energia seja relativamente alto para manter tabelas de rotas proativas, pretendemos investigar este mecanismo para esta pesquisa. O uso de rotas reativas (construídas sob demanda) demandam um tempo excessivo para tal finalidade.

Eren Gürses et al. também sugerem o uso de algoritmos de roteamento geográficos (baseados na localização dos nós). Embora tais algoritmos de roteamento permitam obter um requisito importante das aplicações multimídia, isto é, identificar geograficamente um nó sensor ou um grupo de nós sensores na rede, pretendemos investigar mecanismos que sejam mais escaláveis, assim como os mecanismos utilizados pelos algoritmos de roteamento hierárquicos. Uma solução interessante seria o uso de um algoritmo roteamento híbrido. É importante que o algoritmo de roteamento proposto tenha um baixo consumo de energia.

O protocolo de roteamento proposto será implementado no *Network Simulator 2* (NS2) [1]. Há outros simuladores na literatura, no entanto, o NS2 têm sido bastante utilizado para validar protocolos de roteamento no contexto das RSSF. Dentre as métricas para avaliação do protocolo proposto incluem-se: o tempo de transmissão de fluxo multimídia entre um nó sensor fonte e um ponto de acesso, e o tempo de vida e a escalabilidade da rede. A avaliação destas métricas é altamente influenciada pelo espalhamento dos nós sensores na rede [3, 8]. Por exemplo, diferentes posicionamentos para os pontos de acesso da rede resultam em diferentes resultados de desempenho para um protocolo de roteamento. Mohamed Younis et al [10] sugerem que os experimentos em redes de sensores sejam realizados em um cenário mais próximo da realidade, ao invés de simplesmente utilizar uma topologia onde os nós sensores estejam espalhados uniformemente na rede. Neste sentido, propomos também um algoritmo de espalhamento de nós para melhor avaliar o protocolo de roteamento.

As principais contribuições desta pesquisa são a modelagem e implementação de um protocolo de roteamento com QoS e de um algoritmo de espalhamento de nós, ambos para RSSFM de larga escala.

Referências

- [1] Network Simulator 2. <http://www.isi.edu/nsnam/ns/>, Visitado em Novembro de 2007.
- [2] Kemal Akkaya and Mohamed Younis. An energy-aware qos routing protocol for wireless sensor networks. In *Proceedings of the 23rd International Conference on Distributed Computing Systems Workshops*, pages 710 – 715, Providence, Rhode Island USA, Maio 2003.

- [3] Kemal Akkaya and Mohamed Younis. Positioning of base stations in wireless sensor networks. *IEEE Communications Magazine*, 45(4):96 – 102, Abril 2007.
- [4] I. F. Akyildiz, T. Melodia, and K. R. Chowdury. A Survey on Wireless Multimedia Sensor Networks. *Computer Networks (Elsevier)*, 51(4):921–960, Março 2007.
- [5] Paolo Baronti, Prashant Pillai, Vince W. C. Chook, Stefano Chessa, Alberto Gotta, and Y. Fun Hu. Wireless sensor networks: A survey on the state of the art and the 802.15.4 and zigbee standards. *Computer Communications*, 30(7):1655–1695, 2007.
- [6] Min Chen, Victor C. M. Leung, Shiwen Mao, and Yong Yuan. Directional geographical routing for real-time video communications in wireless sensor networks. *Computer Communications*, 30(17):3368 – 3383, Novembro 2007.
- [7] Eren Gürses and Özgür B. Akan. Multimedia Communication in Wireless Sensor Networks. *Annales des Télécommunications*, 60(7-8):872–900, Julho/Agosto 2005.
- [8] Haeyong Kim, Yongho Seok, Nakjung Choi, Yanghee Choi, and Taekyoung Kwon. *Information Networking*, volume 3391/2005 of *Lecture Notes in Computer Science*, chapter Optimal Multi-sink Positioning and Energy Efficient Routing in Wireless Sensor Networks, pages 264–274. Springer Berlin / Heidelberg, Janeiro 2005.
- [9] E. Ilker Oyman and Cem Ersoy. Multiple sink network design problem in large scale wireless sensor networks. In *Proceedings of IEEE International Conference on Communications*, volume 6, pages 3663 – 3667, Junho 2004.
- [10] Mohamed Younis and Kemal Akkaya. Strategies and techniques for node placement in wireless sensor networks: a survey. *Ad Hoc Networks*, 2007. Article in press.

8. Detecção e Localização de Displasias em Imagens de Ressonância Magnética do Cérebro Humano

Autores: Felipe P.G. Bergo, Alexandre X. Falcão

Displasias corticais focais (FCDs) [13] são uma das principais causas de epilepsias resistentes ao tratamento medicamentoso, e seu diagnóstico não invasivo é um procedimento complexo, demorado e subjetivo [2, 11, 14]. Propomos um método automático de detecção de displasias a partir de imagens de ressonância magnética (RM). O método baseia-se na análise automática de uma reformatação curvilínea do cérebro, computada de forma também automática. O resultado final do método é uma visualização interativa do cérebro com possíveis localizações de displasias classificadas por probabilidade, permitindo uma inspeção rápida onde há indícios de lesões.

FCDs são lesões congênitas do córtex cerebral, caracterizadas pela distribuição anormal de neurônios nas camadas corticais [13]. Epilepsia é uma condição neurológica crônica caracterizada por crises espontâneas recorrentes. A presença de FCDs é a segunda causa mais comum de epilepsias resistentes ao tratamento medicamentoso [14].

Trabalhos Anteriores. As FCDs foram descritas por Taylor [13] na década de 1970. A partir de 1990 a técnica de RM passou a oferecer exames volumétricos com resolução adequada, e os indícios de FCDs puderam ser detectados visualmente de forma não invasiva. Este ainda é o método mais comum na prática clínica. Lesões sutis requerem tratamento especial das imagens [2], realizado de forma interativa e não repetível. Diversos trabalhos recentes propõem métodos para detecção automática de FCDs, porém com falhas metodológicas ou premissas inadequadas [1, 7, 9, 12], com taxas de acerto entre 71% e 85%.

Proposta. Nossa proposta é um método automático para classificar possíveis localizações de displasias a partir de imagens tridimensionais de RM. Como resultado final, o método desenvolvido permitirá a visualização tridimensional interativa de reformatações curvilíneas [2] com indicações gráficas das localizações de possíveis displasias, permitindo que o médico inspecione apenas as regiões suspeitas em vez de centenas de imagens.

Nosso método é composto por 4 operações: (1) segmentação do cérebro, (2) correção de alinhamento, (3) análise de textura e classificação de lesões e (4) visualização interativa do resultado.

A segmentação do cérebro é realizada através da técnica de Tree Pruning [3,4,10], que é tolerante a ruídos e independente de modelos (templates), sendo aplicável a pacientes de qualquer idade. A correção de alinhamento é realizada com uma oti-

mização local de um plano de divisão entre os hemisférios [5]. O mapeamento cada superfície utiliza uma transformada de distância Euclideana baseada na Transformada Imagem-Floresta [8].

Displasias aparecem como borramentos assimétricos nas imagens planarizadas. A análise de textura usa medidas de textura e medidas no domínio da frequência para alimentar um classificador baseado em grafos. A visualização 3D das lesões detectadas pode ser facilmente realizada com o hardware atualmente disponível [6].

O método proposto não requer a segmentação de WM e GM, independe de templates e automatiza o procedimento visual de detecção de displasias por especialistas [2, 11], que se baseia na identificação de assimetrias de textura entre os hemisférios do cérebro em planos tangentes à sua curvatura.

O método está sendo validado em um banco de imagens disponibilizado pela FCM-Unicamp, que inclui tanto pacientes com FCDs confirmadas como controles saudáveis. Nosso objetivo é detectar 100% das FCDs, ainda que sejam detectados alguns falsos positivos, já que o diagnóstico final será dado por especialistas.

Referências

- [1] SB Antel et al. Automated detection of focal cortical dysplasia lesions using computational models of their MRI characteristics and texture analysis. *NeuroImage*, 19(4):1748–1759, Aug 2003.
- [2] AC Bastos et al. Diagnosis of subtle focal dysplastic lesions: Curvilinear reformatting from three-dimensional magnetic resonance imaging. *Annals of Neurology*, 46(1):88–94, 1999.
- [3] FPG Bergo and AX Falcão. Fast and automatic curvilinear reformatting of MR images of the brain for diagnosis of dysplastic lesions. In *Proc. of 3rd Intl. Symp. on Biomedical Imaging*, pages 486–489. IEEE, Apr 2006.
- [4] FPG Bergo, AX Falcão, PAV Miranda and LM Rocha. Automatic Image Segmentation by Tree Pruning. *J Math Imag Vis*, to appear.
- [5] FPG Bergo, GCS Ruppert, LF Pinto and AX Falcão. Fast and Robust Midsagittal Plane Location in 3D MR Images of the Brain In *Biosignals 2008*, Funchal, Portugal, to appear (Jan 2008).
- [6] F de Goes, FPG Bergo, AX Falcão, S Goldenstein, and L Velho. Adapted dynamic meshes for deformable surfaces. In *XIX Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, pages 213–220. IEEE, Oct 2006.
- [7] O Colliot et al. Segmentation of focal cortical dysplasia lesions on MRI using level set evolution *NeuroImage*, 32(4):1621–1630, Oct 2006.

- [8] AX Falcão, J Stolfi, and RA Lotufo. The image foresting transform: Theory, algorithms, and applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(1):19–29, 2004.
- [9] J Kassubek, HJ Huppertz, J Spreer, and A Schulze-Bonhage. Detection and localication of focal cortical dysplasia by voxel-based 3-D MRI analysis. *Epilepsia*, 43(6):596–602, Jun 2002.
- [10] PAV Miranda, FPG Bergo, LM Rocha, and AX Falcão. Tree-pruning: A new algorithm and its comparative analysis with the watershed transform for automatic image segmentation. In *XIX Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, pages 37–44. IEEE, Oct 2006.
- [11] MA Montenegro, LM Li, MM Guerreiro, CAM Guerreiro, and F Cendes. Focal cortical dysplasia: Improving diagnosis and localization with magnetic resonance imaging multiplanar and curvilinear reconstruction. *J Neuroimaging*, 12(3):224–230, Jul 2002.
- [12] S Srivastava et al. Feature-based statistical analysis of structural MR data for automatic detection of focal cortical dysplastic lesions. *NeuroImage*, 27(2):253–266, Aug 2005.
- [13] DC Taylor, MA Falconer, CJ Bruton, and JA Corsellis. Focal dysplasia of the cerebral cortex in epilepsy. *J Neurol Neurosurg Psychiatry*, 34(4):369–387, Aug 1971.
- [14] H Urbach, J Hattingen, J von Oertzen, C Luyken, H Clusmann, T Kral, M Kurthen, J Schramm, I Blümcke, and HH Schild. MR imaging in the pre-surgical workup of patients with drug-resistant epilepsy. *Am J Neuroradiol*, 25(6):919–926, Jun 2004.

9. Um Método Híbrido de Agrupamento para Recuperação de Informação

Autores: Jurandy Almeida, Siome Goldenstein e Ricardo Torres

A evolução das tecnologias de informação e comunicação permitiram a criação de grandes repositórios de dados, aumentando a necessidade de desenvolver sistemas de informação eficientes para gerenciar essas coleções. A busca de uma informação é uma das tarefas mais importantes que deve ser suportada por esses sistemas.

O algoritmo mais simples para realizar uma consulta é uma busca sequencial. Nesse método, todos os dados são analisados sequencialmente. Embora simples, esse método é inviável para grandes coleções, uma vez que o tempo gasto para processar uma consulta é proporcional ao tamanho do banco [8].

Existem extensivos estudos em técnicas indexação e estruturas de dados para acelerar o processo de consulta nesses repositórios de forma que informações relevantes possam ser encontradas rapidamente [3, 4]. Entretanto, na maioria dessas técnicas, a eficiência é comprometida pela sobreposição dos dados, característica comum na recuperação de informação [5, 6].

Em contraste, agrupamento (*clustering*) é uma das técnicas de descoberta de conhecimento mais úteis para identificar correlações em grandes conjuntos de dados. Nesse método, os dados são organizados em grupos, aos quais é associado um elemento representativo. Assim, uma consulta é realizada utilizando-se apenas um pequeno número de elementos representativos, reduzindo o espaço de busca [7].

Em geral, os algoritmos de agrupamento podem ser hierárquicos ou particionais. Os algoritmos hierárquicos encontram grupos sucessivos utilizando grupos pré-estabelecidos, enquanto os algoritmos particionais encontram todos os grupos de uma única vez [2].

As estratégias de agrupamento hierárquico podem ser divididas em dois paradigmas básicos: aglomerativos e divisivos. Uma estratégia aglomerativa começa com cada elemento como um grupo independente e os une, sucessivamente, em grupos cada vez maiores. Por outro lado, uma estratégia divisiva começa com um grande grupo e o divide, sucessivamente, em grupos menores [2].

Este trabalho baseia-se no DAH-Cluster (*Divisive-Agglomerative Hierarchical Clustering*), um método híbrido (hierárquico divisivo e aglomerativo) de agrupamento [1]. Esse método combina características dos modelos hierárquicos divisivo e aglomerativo de agrupamento, reduzindo os erros obtidos por cada um desses paradigmas e melhorando a qualidade das tarefas de agrupamento. Além disso, o DAH-Cluster introduz um novo conceito, chamado fator de reagrupamento, que permite agrupar elementos similares que seriam separados pelos paradigmas tradicionais.

Sejam:

- c – um grupo;
- $c.rep$ – o elemento representativo do grupo c ;

- $c.elements$ – o conjunto de elementos do grupo c ;
- $c.son$ – um apontador para o próximo nível da hierarquia de grupos;
- C – um conjunto de grupos;
- k – o número de grupos em todas as tarefas de agrupamento;
- $f \in [0, 1)$ – um fator de reagrupamento;
- E – o conjunto de elementos analisados;
- D – uma medida para avaliar as dissimilaridades entre os elementos em E .

A Figura 3 ilustra uma representação esquemática do DAH-Cluster. Inicialmente, têm-se E elementos para serem agrupados. Após a primeira iteração, é construído o nível C_1 da hierarquia, com k grupos $c_1 \dots c_k$. Para cada grupo c_i em C_1 , são selecionados os $\lfloor f \times k \rfloor$ grupos mais próximos, criando um novo conjunto de elementos E_{c_i} . Para cada conjunto de elementos E_{c_i} é executado um novo processo de agrupamento, gerando um novo nível na hierarquia, composto pelos grupos $c_{11} \dots c_{1k} \dots c_{k1} \dots c_{kk}$. Esse processo é repetido enquanto $|E_{c_i}| > k$.

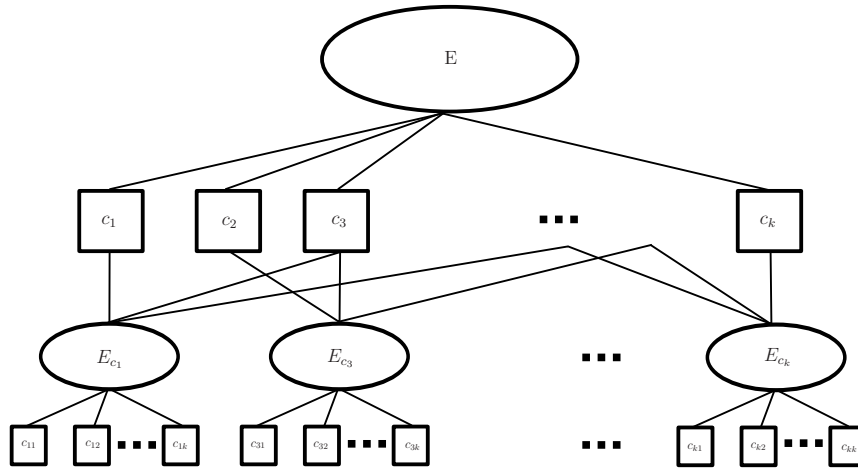


Figura 3: Representação esquemática do DAH-Cluster.

Esse processo consiste em três etapas: (1) inicialmente, executa-se um estágio de agrupamento no conjunto de elementos E , gerando um conjunto de grupos C ; (2) para cada grupo $c \in C$, é construído um novo conjunto de elementos E^* agrupando-se os elementos $c.elements$ dos $\lfloor f \times k \rfloor$ grupos mais próximos ao representante $c.rep$; (3) se o número de elementos em E^* for maior que o número de grupos k , então

um novo nível *c.son* é criado na hierarquia de *c*, reiniciando as etapas (1–3) para o conjunto de elementos E^* .

Esse método representa um paradigma hierárquico aglomerativo no sentido que começa com cada elemento como um grupo representativo e encontra k grupos. Por outro lado, representa um paradigma hierárquico divisivo no sentido que começa com um conjunto de elementos E e, iterativamente, particiona E em subconjuntos E^* . Esse particionamento é feito por meio de um fator de reagrupamento f que permite agrupar elementos similares que seriam separados pelos paradigmas tradicionais devido a sobreposição dos dados.

Este trabalho tem como objetivo o estudo e o desenvolvimento de estruturas de indexação que sejam capazes de resolver problemas ligados à sobreposição de dados em grandes repositórios de informação. Dessa forma, várias extensões ao DAH-Cluster, tanto do ponto de vista teórico quanto prático, serão alvo de pesquisas deste trabalho. Algumas dessas extensões são apresentadas a seguir:

- **Atualização da estrutura hierárquica.** A utilização do DAH-Cluster na recuperação de informação consiste em duas etapas: (1) construção de uma estrutura hierárquica que melhor respresenta os relacionamentos semânticos entre os dados; e (2) utilização dessa estrutura para reduzir o tempo de processamento de uma consulta. Assim, a definição e implementação de regras que permitam atualizar essa estrutura precisam ser investigadas.
- **Comparação com estruturas de indexação.** As estruturas de indexação [5, 6] são as técnicas mais utilizadas na recuperação de informação para reduzir o tempo de processamento de uma consulta. Entretanto, na maioria dessas técnicas, a eficiência é comprometida pela sobreposição dos dados, característica comum na recuperação de informação, uma vez que um mesmo dado pode ser visto sob diferentes interpretações. Nesse contexto, uma das etapas deste trabalho consiste em comparar essas técnicas tanto sob o ponto de vista quantitativo quanto qualitativo.

Referências

- [1] J. Almeida. Recuperação de Imagens por Cor utilizando Análise de Distribuição Discreta de Características. Master's thesis, Instituto de Computação, Unicamp, Campinas, SP, August 2007.
- [2] P. Berkhin. A survey of clustering data mining techniques. In J. Kogan, C. Nicholas, and M. Teboulle, editors, *Grouping Multidimensional Data: Recent Advances in Clustering*, volume 12, chapter 2, pages 25–71. Springer, Berlin, Heidelberg, Germany, 2006.
- [3] E. Chávez, G. Navarro, R. A. Baeza-Yates, and J. L. Marroquín. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, September 2001.

- [4] V. Gaede and O. Günther. Multidimensional access methods. *ACM Computing Surveys*, 30(2):170–231, June 1998.
- [5] C. Traina Jr., A. J. M. Traina, C. Faloutsos, and B. Seeger. Fast indexing and visualization of metric data sets using slim-trees. *IEEE Transactions on Knowledge and Data Engineering*, 14(2):244–260, March 2002.
- [6] C. Traina Jr., A. J. M. Traina, B. Seeger, and C. Faloutsos. Slim-trees: High performance metric trees minimizing overlap between nodes. In C. Zaniolo, P. C. Lockemann, M. H. Scholl, and T. Grust, editors, *Proceeding of the International Conference on Extending Database Technology*, volume 1777 of *Lecture Notes in Computer Science*, pages 51–65, Konstanz, Germany, March 27–31 2000. Springer.
- [7] M. Shyu, S. Chen, M. Chen, and C. Zhang. A unified framework for image database clustering and content-based retrieval. In S. Chen and M. Shyu, editors, *Proceedings of the ACM International Workshop on Multimedia Databases*, pages 19–27, Washington, DC, USA, November 13 2004. ACM Press.
- [8] R. O. Stehling, M. A. Nascimento, and A. X. Falcão. Techniques for color-based image retrieval. In C. Djeraba, editor, *Multimedia Mining - A Highway to Intelligent Multimedia Document*, volume 22 of *Multimedia Systems and Applications*, chapter 4, pages 61–80. Kluwer Academic Publishers, Norwell, MA, USA, 2002.

10. Grafos Pfaffianos e Problemas Relacionados

Autores: Alberto Alexandre Assis Miranda, e Cláudio Leonardo Lucchesi

Seja D um grafo orientado. Seja Q um circuito com número par de vértices de D . Diz-se que a orientação de Q em D é *ímpar* se, ao fixarmos um sentido de percurso de Q , o número de arestas orientadas no mesmo sentido do percurso é ímpar. Um subgrafo H de um grafo G é *conforme* se $G - V(H)$ tem emparelhamento perfeito. As seguintes afirmações sobre D são equivalentes:

- D é uma orientação *Pfaffiana*;
- todo circuito conforme de D com um número par de vértices tem orientação ímpar;
- seja M um emparelhamento perfeito qualquer de D , então todo circuito M -alternado tem orientação ímpar.

Um grafo não orientado é Pfaffiano se e somente se existe uma orientação de suas arestas que é uma orientação Pfaffiana.

A definição de grafo Pfaffiano deriva da idéia de Tutte de usar a estrutura matemática Pfaffiano (não definida aqui) na teoria de emparelhamentos. Em seu livro “Graph Theory As I Have Known It” [15], ele descreve como chegou à idéia de usar os Pfaffianos para determinar uma fórmula para o número de emparelhamentos perfeitos de um grafo. Apesar de não ter sido bem sucedido em encontrar essa fórmula, Tutte conseguiu utilizar identidades envolvendo Pfaffianos para demonstrar o seu teorema famoso que caracteriza grafos que têm emparelhamentos perfeitos [14].

Surpreendentemente, o problema de se decidir se um dado grafo é Pfaffiano está relacionado a outros problemas fundamentais de teoria dos grafos, e aparentemente não relacionados. Por exemplo, o problema de decidir se um dado grafo orientado tem ou não um circuito orientado par é equivalente ao problema de decidir se um grafo bipartido dado é ou não Pfaffiano [16].

O problema está também relacionado a outras questões em física, química e economia (veja o livro de Lovász e Plummer [10, Capítulo 8] e o artigo de McCuaig [11].) Motivado por problemas externos à teoria dos grafos, Kasteleyn [6] demonstrou que todo grafo planar tem uma orientação Pfaffiana. O grafo $K_{3,3}$ é o menor grafo não Pfaffiano.

Little [7] demonstrou que um grafo bipartido é Pfaffiano se e somente se não contém subgrafo conforme que é uma bissubdivisão de $K_{3,3}$. Uma *bissubdivisão* de um grafo é obtida substituindo-se arestas do grafo por caminhos com um número par de vértices internos. Um subgrafo H de um grafo G é *conforme* se $G - V(H)$ tem um emparelhamento perfeito. Assim, o problema de decidir se um dado grafo bipartido é Pfaffiano está em co-NP. Este resultado imediatamente sugere uma pergunta natural: decidir se o problema geral está ou não em NP. Vazirani e Yanakakis [16] mostraram que decidir se uma dada orientação de um grafo é ou não Pfaffiana é tão difícil quanto decidir se o grafo tem ou não uma orientação Pfaffiana.

McCuaig [11], e, independentemente, Robertson, Seymour e Thomas [13], descobriram algoritmo polinomial para decidir se um grafo bipartido tem ou não uma orientação Pfaffiana.

Um *grafo coberto por emparelhamentos* é um grafo conexo e não trivial no qual toda aresta pertence a algum emparelhamento perfeito. Maiores detalhes sobre grafos cobertos por emparelhamentos podem ser encontrados em [2–4, 9, 10]. O estudo de grafos Pfaffianos pode ser restrito aos grafos cobertos por emparelhamentos, de forma natural. As propriedades de grafos cobertos por emparelhamentos podem então auxiliar na compreensão das propriedades de grafos Pfaffianos.

Um *corte justo* em um grafo coberto por emparelhamentos é um corte que contém precisamente uma aresta em cada emparelhamento perfeito do grafo. Seja $C := \partial(X)$ um corte não trivial em um grafo G . O grafo H obtido a partir de G pela contração de todos os vértices em X a um único vértice x é chamado de C -*contração*, e representado por $H := G\{X \rightarrow x\}$. As duas C -contrações de G são $G\{X \rightarrow x\}$ e $G\{\bar{X} \rightarrow \bar{x}\}$. Caso G seja coberto por emparelhamentos e C um corte justo, ambas as C -contrações de G são cobertas por emparelhamentos. Naturalmente, se algumas das C -contrações ainda possui algum corte justo não trivial, a operação pode ser repetida até que os grafos obtidos sejam livres de cortes justos não triviais. Este processo denomina-se *decomposição em cortes justos*. Lovász [9] provou que a família de grafos resultantes de uma decomposição em cortes justos de um grafo coberto por emparelhamentos é única a menos de arestas múltiplas. Além disso, pode-se reduzir o problema de se decidir se um grafo G coberto por emparelhamentos é Pfaffiano ao problema de se decidir se os grafos cobertos por emparelhamentos livres de cortes justos são Pfaffianos.

Os grafos cobertos por emparelhamentos livres de cortes justos não triviais são separados em duas classes: *tijolos* e *presilhas*. As presilhas são bipartidas, e os tijolos não são bipartidos. A decomposição em cortes justos de um grafo bipartido coberto por emparelhamentos produz somente presilhas. A decomposição de um grafo não bipartido coberto por emparelhamentos resulta em tijolos e presilhas. Sendo assim, o problema Pfaffiano se reduz no caso bipartido a presilhas e no caso geral a tijolos e presilhas. Maiores informações podem ser encontradas em [5]. Mais especificamente, o algoritmo de McCuaig e Robertson, Seymour e Thomas decidem se uma presilha é Pfaffiana. No entanto, até hoje não se sabe nada sobre a complexidade de se decidir se um tijolo é Pfaffiano.

Um corte separador C em um grafo coberto por emparelhamentos G é um corte tal que ambas as C -contrações de G são cobertas por emparelhamentos. Similarmente ao que é feito com cortes justos, pode-se definir uma decomposição de um grafo coberto por emparelhamentos em cortes separadores. Grafos cobertos por emparelhamentos livres de cortes separadores são chamados de *grafos sólidos*. Recentemente, Carvalho, Lucchesi e Murty caracterizaram grafos sólidos Pfaffianos [1]. Novamente, não se sabe até hoje se existe algoritmo polinomial para decidir se um grafo sólido é Pfaffiano. Também não se sabe se existe algoritmo para encontrar um corte separador em um grafo não sólido.

Outro problema relacionado é se existe algoritmo polinomial para decidir se um grafo quase-bipartido é Pfaffiano. Um grafo *quase-bipartido* é um grafo não bipartido G tal que existem duas arestas e e f e $G - e - f$ é bipartido. Recentemente, Little e Fischer caracterizaram grafos quase-bipartidos Pfaffianos [8], de uma forma similar àquela dos bipartidos, através de subgrafos conformes proibidos. Dessa forma, o problema Pfaffiano para esta classe está em co-NP. Até recentemente, não se sabia se existe algoritmo polinomial para o problema Pfaffiano para esta classe de grafos.

O objetivo deste projeto é determinar a classe de complexidade (P, NP, NP-completo, etc) de problemas sobre grafos Pfaffianos. Em particular, estamos interessados em determinar a classe de complexidade a que pertence cada um dos problemas a seguir:

- determinar se um tijolo é Pfaffiano;
- determinar se um grafo quase-bipartido é Pfaffiano;
- determinar se um grafo sólido é Pfaffiano;
- encontrar um corte separador em um grafo não-sólido.

Recentemente, apresentamos um algoritmo de tempo polinomial para determinar se um grafo quase-bipartido é Pfaffiano [12].

Referências

- [1] M. H. de Carvalho, C. L. Lucchesi, and U. S. R. Murty. A characterization of solid pfaffian matching covered graphs. *Comunicação Particular*.
- [2] M. H. de Carvalho, C. L. Lucchesi, and U. S. R. Murty. Ear decompositions of matching covered graphs. *Combinatorica*, 19:151–174, 1999.
- [3] M. H. de Carvalho, C. L. Lucchesi, and U. S. R. Murty. On a conjecture of Lovász concerning bricks. I. The characteristic of a matching covered graph. *J. Combin. Theory Ser. B*, 85:94–136, 2002.
- [4] M. H. de Carvalho, C. L. Lucchesi, and U. S. R. Murty. On a conjecture of Lovász concerning bricks. II. Bricks of finite characteristic. *J. Combin. Theory Ser. B*, 85:137–180, 2002.
- [5] M. H. de Carvalho, C. L. Lucchesi, and U. S. R. Murty. On the number of dissimilar Pfaffian orientations of graphs. *Theor. Inform. Appl.*, 39:93–113, 2005.
- [6] P. W. Kasteleyn. Dimer statistics and phase transitions. *J. Math. Phys.*, 4:287–293, 1963.

- [7] C. H. C. Little. A characterization of convertible $(0, 1)$ -matrices. *J. Combin. Theory Ser. B*, 18:187–208, 1975.
- [8] C. H. C. Little and I. Fischer. A characterisation of Pfaffian near bipartite graphs. *J. Combin. Theory Ser. B*, 82:175–222, 2001.
- [9] L. Lovász. Matching structure and the matching lattice. *J. Combin. Theory Ser. B*, 43:187–222, 1987.
- [10] L. Lovász and M. D. Plummer. *Matching Theory*. Number 29 in Annals of Discrete Mathematics. Elsevier Science, 1986.
- [11] W. McCuaig. Pólya’s permanent problem. *The Electronic J. of Combin.*, 11, 2004.
- [12] A. A. A. Miranda and C. L. Lucchesi. A polynomial time algorithm for recognizing near-bipartite Pfaffian graphs. Technical Report IC-07-15, May 2007.
- [13] N. Robertson, P. D. Seymour, and R. Thomas. Permanents, Pfaffian orientations and even directed circuits. *Ann. of Math. (2)*, 150:929–975, 1999.
- [14] W. T. Tutte. The factorization of linear graphs. *J. London Math. Soc.*, 22:107–111, 1947.
- [15] W. T. Tutte. *Graph Theory as I Have Known It*. Number 11 in Oxford Lecture Series in Mathematics and its Applications. Clarendon Press, Oxford, 1998.
- [16] V. V. Vazirani and M. Yannakakis. Pfaffian orientation of graphs, 0,1 permanents, and even cycles in digraphs. *Discrete Applied, Math.*, 25:179–180, 1989.

11. Especificação de Modelos de Dados para Bibliotecas Digitais de Objetos Complexos

Autores: Nádia P. Kozievitch e Ricardo Torres

A necessidade de mecanismos para troca de informações, preservação, arquivamento e indexação de dados digitais (documentos, imagens, vídeos, modelos, catálogos, informações geo-referenciadas, séries temporais, etc.) favoreceram o surgimento de Bibliotecas Digitais (BD). Esta grande quantidade de dados heterogêneos disponível na internet facilitou a busca pela informação e enriqueceu o conceito que já tínhamos das bibliotecas tradicionais, sem as restrições de espaço e tempo.

Como exemplos de bibliotecas digitais, podemos citar a digitalização de várias bibliotecas tradicionais, como a Biblioteca Digital da Unicamp (<http://libdigi.unicamp.br/>) e Biblioteca Digital de Oxford (<http://www.odl.ox.ac.uk/>) e Biblioteca Nacional da França (<http://www.bnf.fr/>); museus, como o Australian Museums On Line, (<http://www.nma.gov.au/AMIS>) e outras iniciativas, como a Biblioteca do Congresso Americano (<http://memory.loc.gov/ammem/index.html>) e a biblioteca do Vaticano (http://www.vatican.va/biblioteca_vatic/index.html) que antes tinham seu acervo disponível somente de forma local. Na comunidade científica podemos citar o projeto ETANA-DL [1], uma biblioteca digital criada para gerenciar informações arqueológicas e seus serviços, e o protótipo StoryBank [2], que mistura apresentações visuais e desenhos com áudio.

Do ponto de vista computacional, uma BD é construída a partir de componentes simples chamados de objetos digitais. Estes objetos podem ser agregados a outros tipos de dados, metadados e algumas vezes serviços, criando o conceito de Objetos Complexos. Por exemplo, ao invés de ter somente um arquivo executável, um objeto complexo poderia agregar além disso os metadados, as regras de permissão de acesso e os links para os serviços de conversão de formato e/ou outros objetos complexos relacionados.

Da mesma forma que diferentes formatos de metadados se proliferam, como XML e RDF (Resource Description Format), surgem inovações em objetos complexos [3], como os formatos NetCDF, HDF e ELFS na computação científica. Já na indústria, os objetos complexos podem ser encontrados no armazenamento persistente de bancos de dados e codecs multimídia. Desta evolução surgiram ainda outras tecnologias, como o Framework Multimídia MPEG-21 e formatos de objetos digitais, como o MPEG-21 DIDL e METS. E apesar de haverem vários padrões especializados para objetos complexos, existe ainda muita incompatibilidade entre os mesmos, motivando pesquisas na busca de soluções de integração. Em [4], por exemplo, é apresentado um estudo comparativo de 2 padrões para representação de objetos digitais complexos: na área de Educação (Im Content Package - IMS CP) e em desenvolvimento de software (Reusable Asset Specification - RAS). E apesar de cada padrão possuir características específicas, o trabalho concluiu que existem muitas características compatíveis que permitem um mapeamento direto entre os

padrões.

Este projeto tem como objetivo propor um modelo de dados de objetos complexos em bibliotecas digitais, visando sua utilização na especificação, implementação, de sistemas para gerenciamento objetos complexos e serviços. Para isso utilizaremos o Modelo 5S (Streams, Structures, Spaces, Scenarios e Societies) [5] como uma base formal para definir, relacionar e unificar conceitos de bibliotecas digitais, metadados, coleções e serviços. A utilização do 5S permite modelar uma biblioteca digital em diferentes dimensões, incluindo tipos de dados suportados (streams model), a sua organização e estrutura (structures model), propriedades espaciais e de interface da biblioteca (spaces model), aspectos comportamentais da biblioteca digital (scenarios model) e as diferentes comunidades envolvidas (societies model).

As contribuições esperadas são: (i) proposta de um modelo de dados para objetos complexos; (ii) especificação e implementação de bibliotecas digitais para gerenciamento de dados complexos; (iii) formalização teórica de bibliotecas digitais de objetos complexos, usando o formalismo 5S.

Referências

- [1] Ravindranathan U., Shen R., Gonçalves M.A, Fan W., Fox E.A., Flanagan J.W. *ETANA-DL: a digital library for integrated handling of heterogeneous archeological data*. JCDL, 2004.
- [2] Jones M., Harwood W., Buchanan G., Lalmas M.R. StoryBank: an Indian village community digital library. *JCDL, 2007*.
- [3] Nelson M.L., Sompel H. IJDL special issue on complex digital objects. In *IJDL, 2006*.
- [4] Dourado P., Ferreira P., Santanchè A. Representação unificada de objetos digitais complexos: confrontando o RAS com o IMS CP. in III Workshop de Bibliotecas Digitais, 2007.
- [5] Gonçalves M.A, Fox E.A., Watson L.T., Kipp N.A. Streams, Structures, Spaces, Scenarios, Societies (5S): A formal model for digital libraries. *TOIS, 2004*.

12. Design of Flexible User Interfaces

Authors: Vânia Paula de Almeida Neris and Maria Cecília Calani Baranauskas

Introduction

The socialization of computer systems brought a new challenge to Human-Computer Interaction (HCI) researchers: how to design interfaces that are easy to use by so many people with the most different profiles, in different situations of use? The natural answer to this question seems to be to offer the end-user the possibility of adapting the interface to his/her preferences, needs and situations of use. Nevertheless, the solutions to the adaptation problem have not been that simple.

Tailoring is the expression used in literature to define the activity of changing a computer application according to its context of use [3]. Applications that allow tailoring offer to end-users the possibility to adapt the software to their personal preferences or changes in the task, after the software implementation [7]. Tailoring involves the concept of “design for change”, offering the flexibility of being adapted to different organizational contexts or not anticipated situations of use, or those that have changed. The main benefits that can be obtained with this type of flexibility, emphasized by literature, are: more efficiency [5], more satisfaction of use [6] and a smaller learning curve when an application is replaced [4, 6].

While the term tailoring has been well accepted and used, some related concepts can also be found such as: customization, end-user modification, extension, personalization etc. In this work we will use the terms tailoring or adaptation as synonymous and the focus is on the tailoring that is done explicitly by the end user and not automatically by the software. In this context, interfaces can be understood as flexible.

However, solutions of adaptation, already implemented in some applications, have not been effective. Two main problems still persist: (1) interfaces do not communicate clearly the possibility of been tailored; (2) and when they do, they ask for skills that unsophisticated users don't have. Considering this problem, the proposal for the PhD. Thesis entitled “A Framework for Flexible User Interfaces Design” aims at finding new solutions for tailoring, to help the interface engineer in the design of flexible interfaces that ask for less effort from final users, during the adaptation of these interfaces.

Next sections present the main results obtained until now, conclusions and further work.

Semiotic-informed Categorization for Tailoring

We argue that designing applications that make possible some level of tailoring by the end-user should consider aspects such as: (a) the system architecture and also questions regarding implementation; (b) documentation and the expression of the

possibility of tailoring; and also (c) aspects regarding the social impact of possible changes [1].

According to Organizational Semiotics [8], an organization is understood as any information system in which people use signs for communication toward purposeful and coordinated actions. In this sense, any technological artifact, e.g. a software application, is embedded in a formal system that, by its turn, exists in the context of an informal system.

The Semiotic-informed categorization for tailoring that we proposed [2] considers the impact of the potential changes not only in the software application, but in the whole organization. The possibilities of tailoring are offered in the user interface as a reflection of changes that can happen in any of the three layers (technical, formal and informal) of the information system. This wider view of end-user tailoring demands the designers' knowledge about the sub-culture and rules of the organization to reflect them in the software application in order to provide end-user tailorable applications that demand less effort to be adapted. Figure 4 illustrates the semiotic perspective for end-user tailoring.

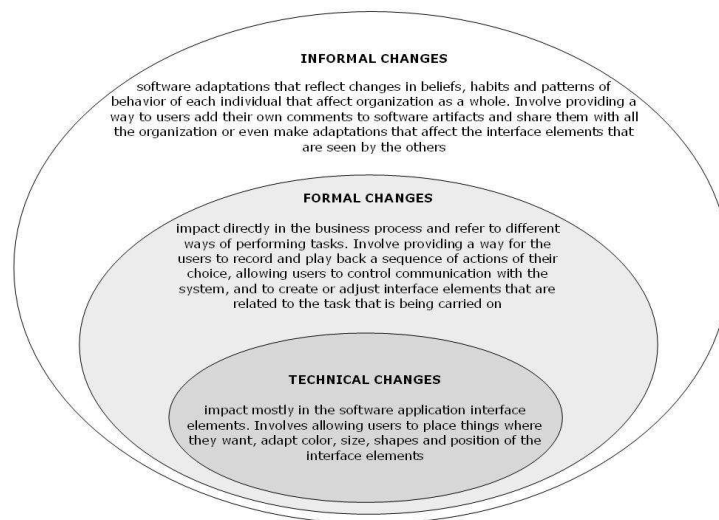


Figura 4: A Semiotic-informed categorization for end-user tailoring.

The Semiotic-informed categorization for end-user tailoring may contribute to the design and development of tailorable applications as (a) it considers tailoring from the point of view of the impact of prospective changes in the social context of the system usage; (b) designers concerned with the sub-culture and rules of the organization could develop a deeper understanding of the users' needs; (c) by taking a view of the interactive system embedded in the organization's formal and informal layers, the designer is likely to conceive interface elements in a way the user could effortlessly perceive the tailorable features.

Elicitation Patterns for Tailorable User Interfaces

As interaction patterns have been considered a promising approach to bridge the gaps between analysis, design and implementation of usability related features, we investigated and synthesized from literature a set of interaction patterns related to tailoring activities. From this analysis, we proposed a group of Tailorable User Interface Features (TUIFs), that can be understood as characteristics that interfaces that allow tailoring should have. To cope with the design process of tailorable applications that implement the TUIFs, we have elaborated a set of issues to be discussed with stakeholders during the elicitation of software requirements [1]. Table 1 shows knowledge about TUIFs structured in a pattern format.

Table 1. TUIF Elicitation Patterns.

Name	Problem	Context	Solution - Tailoring Features Elicitation Guide Issues to be discussed with stakeholders
I can customize	Which information needs to be elicited in order to provide users with the “show the tailoring options” feature?	One of the main reasons users do not tailor is they don’t know they can do it.	<ul style="list-style-type: none"> - Will the application offer different actions to be customized? - If so, what are the actions that will be customized? - Is that possible to offer the tailoring function really close to where the action will be performed? - Is that possible to show the changes in a WYSIWYG format? - How users are going to save changes? - How application will manage to show saved changes next time user comes back?
Rearrange layout	Which information needs to be elicited in order to provide users with the “change the position of interface elements” feature?	Users may want to change interface elements positions, rearranging layout.	<ul style="list-style-type: none"> - Which interface elements could be moved? - Which movement options will be offered (stacking, grouping, aligning, nesting)? - Is that possible to move the artifact directly? - How users are going to save changes? - How application will manage to show saved changes next time user comes back?
Create and Save My Interface Elements	Which information needs to be elicited in order to provide users with the “define own interface elements” feature?	Users may want to create or change layout and graphical presentation of interface elements, according to their own understandings.	<ul style="list-style-type: none"> - Will the application allow users to create new elements with new functions? - If so, could new elements be created in any part of the interface? - How will users create the new elements (a completely new or a modified copy)? - If copies will be allowed, how to offer the possibility to make changes in only one copy or in all copies? - How will these new elements be recorded? - How are users going to name the new elements? - Will the application allow users to organize and reuse their new elements? - If so, how the new elements can be organized?
Record and Repeat	Which information needs to be elicited in order to provide users with the “recording actions and points of interest” feature?	In a navigable software system, when the system is possibly large and complex, allows the user to move freely through it in ways not directly supported by the initial design	<ul style="list-style-type: none"> - Will the application allow users to record different actions (depending on the kind of application action means a specific functionality performed by the user, a place visited, etc.) they perform? - If so, how many actions can be recorded? - How will these actions be recorded? - How are users going to name the record? - Will the application allow users to organize their records? - If so, how the records can be organized? - How is playback going to be presented? Is it possible to offer a single button?
My had-own understanding	Which information needs to be elicited in order to provide users with the “insert comments” feature?	Interface elements sometimes are not as meaningful as they should be. Users may want to add comments about interface elements that can be read by the same user or by others	<ul style="list-style-type: none"> - Will the application allow users to make notes to all interface elements? - Will the application allow users to make different types of notes (in different colors, different formats)? - How will these notes be recorded? - How are users going to name the notes? - Will the application allow users to organize their notes? - If so, how can the notes be organized?

Conclusions and Ongoing Work

This paper presented a Semiotic-informed categorization for end-user tailoring that considers tailoring from the point of view of the impact of prospective changes in the social context of the system usage. Also, we formalized Tailorable User Interface Features and some Elicitation Patterns. Future works involve promoting design sections with users to better understand their needs and behavior when tailoring interfaces. Afterward a framework with guidelines, methods and software applications to help the design of tailorable applications will be formalized. This research follows an empiric methodology, more specifically qualitative. Case studies will be done to validate the framework.

Referências

- [1] M.C.C. Baranauskas and V.P.A. Neris. Using Patterns to Support the Design of Flexible User Interaction. *12th International Conference on Human-Computer Interaction*. Interaction Design and Usability LNCS 4550. Heidelberg: Springer. v. 4550. p. 1033-1042. 2007.
- [2] V.P.A. Neris, and M.C.C. Baranauskas. End-user Tailoring: a Semiotic-informed Perspective. *Proc. International Conference on Organizational Semiotics*. Sheffield. p. 47-53. 2007.
- [3] H. Kahler, A. Morch, O. Stiemerling, and V. Wulf. Computer Supported Cooperative Work : The Journal of Collaborative Computing. *Computer Supported Cooperative Work*. Vol. 9(I), p:1-4. 2000.
- [4] J. Ma, M. Kienle, and P. Kaminski. Customizing Lotus Notes to Build Software Engineering Tools. *Proc. Center for Advanced Studies on Collaborative*. p:211 - 222, Toronto. 2003.
- [5] W. Mackay Triggers and Barriers to Customization Software. *Proc. Human factors in computing systems: Reaching through technology*. New Orleans, p:153 - 160. 1991.
- [6] D. Rivera The Effect of Content Customization on Learnability and Perceived Workload. In *CHI05 Human factors in computing systems*. p:1749 - 1752 Portland, USA, 2005.
- [7] R. Slagter, M. Biemans, and H. Hofte. Evolution in Use of Groupware: Facilitating Tailoring to the Extreme. *7th International Workshop on Groupware*. CRIWG. IEEE. 2001.
- [8] R. Stamper Language and computer in organized behavior. *Riet and Meersman (eds) Linguistic Instruments in Knowledge Engineering*. Elsevier. p:143-163. 1992.

13. Especificação de Mecanismo para Anotação Semântica de Dados Geoespaciais Disponíveis na Web

Autores: Carla Geovana do Nascimento Macário, Claudia Maria Bauzer Medeiros

A agricultura é uma atividade que contribui significativamente para o PIB brasileiro. Segundo [1] em 2005 o PIB atingiu R\$ 1.929 bilhões, sendo R\$ 537 bilhões provenientes de atividades agrícolas, o que corresponde a quase 30% do montante brasileiro. Desse valor, grande parte é proveniente do agronegócio. Diversos fatores contribuem para estes números. Um deles é a disponibilidade cada vez maior de sistemas que auxiliam no planejamento e no gerenciamento da produção. Em qualquer uma das fases da produção – seja ela antes, dentro e depois da porteira – a tecnologia de informação está presente como uma importante ferramenta para o aumento da produtividade e da qualidade dos produtos gerados. A rapidez no acesso às informações e, conseqüentemente, na tomada de decisões, é determinante na competitividade de uma empresa e seu sucesso a longo prazo.

Aplicações geográficas são aquelas que podem utilizar os SIGs em seu desenvolvimento para produzir algum tipo de informação geográfica, como é o caso da agricultura de precisão, que permite construir bancos de dados espaciais e temporais visando a otimização na utilização dos insumos agrícolas e o uso racional da terra, com conseqüente redução de custos e impactos ambientais.

Atualmente a web desempenha um papel importante neste contexto, funcionando, para esses sistemas, como um grande repositório de informações distribuídas pelo mundo– imagens de satélites, repositórios de dados, registros de coletas, mapas, entre outras. Tal fato tem provocado mudanças radicais nos sistemas no que diz respeito ao acesso e ao uso de informação geoespacial. Há 10 anos esses sistemas processavam dados locais e, portanto a informação era suficientemente livre de ambigüidade para aqueles que as usavam. Nos dias de hoje, os dados e métodos distribuídos podem ser recuperados e combinados em uma maneira *ad hoc*, de repositórios em qualquer lugar do mundo, extrapolando o seu contexto local. Como a semântica dos dados é geralmente descrita em uma linguagem não processável, o seu compartilhamento torna-se difícil. Além disso, os dados são ainda coletados e armazenados por organizações diferentes, usando seus próprios dispositivos, para propósitos distintos, geralmente em formato não padrão. Como aumento da demanda e do número de aplicações e usuários, não se sabe a priori quais os usos que cada dado pode ter. Normalmente, a busca por métodos e dados faz-se pelo seu conteúdo sintático, focando apenas no casamento de palavras-chave. Isso leva a vários problemas semânticos quando se trata de aplicações geográficas [2]:

- em nível de metadado, impede a descoberta de informação geográfica;
- em nível de esquema, impede a recuperação de informação geográfica;

- em nível de dados, impede a interpretação e integração de informação geográfica.

A interoperabilidade semântica é um pré-requisito para o acesso e busca efetiva de dados relevantes em diferentes contextos de aplicação. As soluções que vem sendo amostradas passam pela definição de padrões (para facilitar o intercâmbio de dados) e metadados (para permitir encontrar o que se deseja). Tem-se então um desafio: como viabilizar o uso de dados geoespaciais distribuídos e heterogêneos? Esta não é uma atividade trivial. Primeiro, pelo volume de informação gerado a cada momento. Segundo, pela dificuldade de prover um conteúdo semântico a estes dados de maneira a proporcionar seu uso efetivo pelas diversas aplicações geográficas.

Como resposta parcial a esse desafio, surgiu a idéia da Web Semântica, que busca agregar conteúdo semântico à informação. Para isso, propõe o uso de anotações sobre os recursos de informação – páginas web, imagens, mapas, etc... – para descrição de seu conteúdo, por meio de ontologias de domínio. Como resultado espera-se que as páginas de informação sejam anotadas em uma linguagem processável permitindo aos diversos agentes de busca proceder consultas mais eficazes [3].

Quando se considera componente espacial, base em aplicações geográficas, o uso de anotações se torna mais complicado. Informação geográfica associa objetos e eventos a localidades, possuindo um vocabulário rico de nome de lugares e objetos geográficos, relacionamentos espaciais, padrões, fluxos e interações. Além disso, estes nomes variam com o tempo e até com o tipo de uso. A simples adoção de ontologias no processo de anotações não é suficiente. O uso de ontologias pode ser capaz de resolver questões de ambigüidades de nomes, mas é insuficiente para resolver relacionamentos espaciais, que podem ser especificados usando termos imprecisos como perto, próximo, longe, a $X km$. Por outro lado, coordenadas geográficas podem explicitar o local (onde), facilitando muitas consultas, mas não garante a semântica associada.

Esta limitação tem motivado pesquisa em Recuperação de Informação Geográfica (RIG), uma área em expansão que consiste principalmente em desenvolver ferramentas para recuperação de recursos de informação geográfica (dados, mapas, informação em geral) na Web para responder a questões do tipo *<tema, relacionamento espacial, localização>* [4]. Algumas das questões envolvidas são interpretar adequadamente os termos da consulta, resolvendo possíveis problemas de ambigüidade, e encontrar os documentos associados a relacionamentos geográficos estabelecidos, retornando-os segundo sua relevância.

O processo para criação totalmente automática de anotações representa um problema não resolvido. Várias abordagens semi-automáticas, que exigem a intervenção do usuário em algum momento do processo, vêm sendo propostas. A anotação semântica, ainda que semi-automatizada, aparece como uma boa solução para a interoperabilidade semântica, quando tratamos de informação textual. Mas quando se pensa em aplicações geográficas devemos considerar vários tipos de recurso além de texto - imagens de satélite, mapas, gráficos, bases de dados. Mecanismos para

anotação semântica desses recursos são praticamente inexistentes e com isso, grande parte de informação útil aos sistemas geográficos deixam de ser recuperadas.

O objetivo deste projeto de pesquisa de doutorado é propor um serviço de anotações semânticas geoespaciais – anotações para diferentes dados espaciais (textos, bases de dados, mapas, gráficos, imagens, workflows) – permitindo a recuperação de informação geográfica estratégica para a agricultura. Seguindo a área de RIG, pretende-se a adoção de ontologias como base para este processo, focando também na descoberta de conhecimento. Os principais pontos de pesquisa a serem cobertos são: uso de ontologias e componentes espaciais para anotação de recursos não textuais; uso combinado de diferentes fontes de dados para geração de anotações; documentação deste processo de geração, usando por exemplo, *workflows* científicos; armazenamento das anotações e formato a ser utilizado nas anotações. Espera-se que ao final as anotações geradas por meio deste serviço permitam aos usuários obter respostas a questões do tipo *Quais as culturas são adequadas para plantio na região X?* ou *Existem áreas de desmatamento na região Y?*.

Pretende-se que o serviço a ser desenvolvido seja integrado ao WebMAPS [5] - Sistema baseado na Web Semântica para Monitoramento Agrícola e Previsão de Safras, contribuindo para um de seus objetivos específicos: facilitar o gerenciamento e previsão de safras, e a tomada de decisão com base em dados confiáveis e em ferramentas computacionais que incorporem metodologias científicas ajustadas às condições geofísicas e climatológicas brasileiras.

Referências

- [1] CEPEA. *Centro de Estudos Avançados em Economia Aplicada (CEPEA)*. Esalq/USP, 2007. <<http://www.cepea.esalq.usp.br/pib/>>. Acesso em: 19 jun. 2007.
- [2] M. Lutz and E. Klien. Ontology-based retrieval of geographic information. *International Journal of Geographical Information Science*, 20(3):233–260, 2006.
- [3] V. Uren, P. Cimiano, J. Iria, S. Handschuh, M. Vargas-Vera, E. Motta, and F. Ciravegna. Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Web Semantics: Science, Services and Agents on the World Wide Web*, 4(1):14–28, january 2006.
- [4] C.B. Jones, A.I. Abdelmoty, D. Finch, G. Fu, and S. Vaid. The spirit spatial search engine: Architecture, ontologies and spatial indexing. In *Geographic Information Science: Third International Conference, Gi Science 2004*, pages 125 – 139, Adelphi, Md, USA, October 2004.
- [5] Claudia Maria Bauzer Medeiros. *The WebMaps Project*. LIS/IC/UNICAMP, 2003. Disponível em: <<http://www.ic.unicamp.br/~cmbm/webmaps.html>>. Acesso em: 8 dez. 2006.

14. Metodologias de teste para descrições de hardware de alto nível

Autores: Bruno Albertini, Sandro Rigo e Guido Araujo

Com a crescente complexidade dos sistemas embarcados, os projetistas de hardware estão sendo forçados a aumentar o nível de abstração das descrições. Segundo o *roadmap* do ITRS [2], a redução do *time-to-market* tornará descrições de baixo nível, como VHDL, inviáveis nos próximos anos. A presença de componentes de alta complexidade, como NoCs (*Network on Chip* [3]) já pode ser notada em dispositivos embarcados de uso geral, como telefones celulares. Outro ponto forte que alavanca a mudança de nível de abstração é a integração hardware-software. Utilizando-se modelos de alto nível, é possível fornecer um simulador funcional aos desenvolvedores de software logo depois da especificação funcional do hardware. Isto permite que o desenvolvimento do software ocorra em paralelo com o desenvolvimento do hardware, economizando tempo e identificando falhas de projeto nos estágios iniciais do desenvolvimento, permitindo correções menos custosas.

A migração para níveis de abstração mais altos trouxe diversos problemas, entre eles a depuração. O SystemC [6] é uma biblioteca que estende C++ para que o mesmo suporte descrições de hardware, fornecendo tipos de dados como *bit vector* e um núcleo de simulação baseado em eventos. Na atualidade é a ferramenta mais utilizada no nicho de desenvolvimento de hardware de alto nível, ganhando cada vez mais destaque na indústria, tendo sido recentemente aceito como padrão internacional pela IEEE. A característica mais notável é a possibilidade de utilizar todo o arcabouço já existente para o desenvolvimento de software em C++ para o desenvolvimento de hardware. O hardware é projetado como módulos em SystemC, representados como classes de C++. O núcleo de simulação é capaz de aglutinar os módulos, interligá-los por meio de primitivas especificadas pelo usuário (extensões da biblioteca) e gerar um arquivo executável que simula o comportamento paralelo inerente aos dispositivos de hardware.

Para diminuir o tempo de desenvolvimento, os projetistas naturalmente adotaram uma abordagem baseada em plataformas. Nesta abordagem o desenvolvedor mantém uma biblioteca de componentes reutilizáveis e faz uso dos mesmos sempre que possível. A plataforma é literalmente montada a partir destes componentes ou especializações dos mesmos. O problema que surgiu e que pretendemos nos dedicar é a verificação destes módulos. Depurar os módulos descritos em SystemC têm se mostrado uma tarefa árdua e tediosa. A biblioteca é complexa e o fluxo de execução do núcleo de simulação cruza com o do módulo diversas vezes durante sua simulação. Ferramentas como o GDB, baseada em execução monitorada de instruções, estão sendo abandonadas devido a estes motivos.

O presente trabalho visa prover os desenvolvedores que utilizam SystemC de uma metodologia capaz de prover *breakpoints*, introspecção de dados, anotações de registradores e variáveis internas e injeção de estímulos. A reflexão computacional

pode cumprir estes requisitos e é perfeitamente aplicável, dado que os módulos nada mais são que classes em C++. Com a reflexão, podemos extrair informações sobre a classe, inspecionar e modificar atributos, listar e executar métodos. Um dos problemas que encontramos no início do trabalho foi a falta de reflexão em C++. Ao contrário de linguagens mais recentes que possuem reflexão nativamente, a reflexão em C++ só pode ser obtida através de bibliotecas. Como não conseguimos encontrar nenhuma biblioteca em um estágio de desenvolvimento utilizável, partimos da biblioteca Root-SEAL [5], desenvolvida pelo CERN e complementamos com nossas necessidades.

A metodologia que estamos desenvolvendo envolve a criação de um modelo em SystemC capaz de controlar o núcleo de simulação da biblioteca e de utilizar a reflexão para injeção e introspecção dos módulos. Até o presente momento, podemos instrumentar os tipos de dados mais comuns de C/C++ e da biblioteca SystemC. O usuário acessa o módulo supervisor por meio de uma interface telnet, de onde pode controlar a execução por meio de uma interface similar a do GDB. O módulo também é plugável em sistemas compatíveis com o protocolo desenvolvido, como o PDesigner [4], um *plugin* para Eclipse que permite a captura gráfica da interconexão entre módulos SystemC.

Um ganho notório da utilização da reflexão como ferramenta de depuração é a possibilidade de se obter introspecção sem o código fonte. Para refletir uma classe, basta o binário do módulo e o arquivo *header*. Em todas as abordagens que avaliamos é necessário o fornecimento do código fonte para se obter a mesma funcionalidade. Algumas abordagens comerciais conseguem a mesma funcionalidade utilizando um núcleo de simulação SystemC ou um compilador modificado especificamente para tal tarefa. Nossa abordagem utiliza a biblioteca SystemC sem modificações e o compilador GCC padrão. Podemos também registrar alterações no estado de um módulo para análise posterior ou injetar estímulos nos atributos em tempo de execução para forçar a simulação a atingir casos de teste específicos.

Os resultados obtidos até o momento estão servindo como norteadores das áreas de pesquisas para o doutorado do aluno. Conseguimos montar uma metodologia minimamente intrusiva, com nenhuma modificação das ferramentas usuais e com muito pouco *overhead* na simulação. Já obtivemos sucesso com introspecção de código para depuração e injeção de estímulos. Pretendemos estender esta metodologia para abranger todos os tipos de dados do domínio do hardware e possivelmente tipos compostos, com alguma restrição. Nos planos futuros, incluímos teste e adaptação das técnicas de verificação utilizadas em engenharia de software como uma forma de explorar o potencial das bases desenvolvidas. Há outros trabalhos que já utilizam a metodologia desenvolvida em pelo menos duas instituições de pesquisa. O trabalho e os resultados iniciais foram publicados [1].

Referências

- [1] Bruno Albertini, Sandro Rigo, Guido Araujo, Cristiano Araujo, Edna Barros e Willians Azevedo. A computational reflection mechanism to support platform debugging in systemc. Em *CODES+ISSS '07: Proceedings of the 5th IEEE/ACM international conference on Hardware/software codesign and system synthesis*, pp. 81–86, New York, NY, USA, 2007. ACM, Salzburg, Austria.
- [2] ITRS (International Technology Roadmap for Semiconductors). ITRS. Website, setembro de 2007. <http://www.itrs.net>.
- [3] Marcello Coppola, Stephane Curaba, Miltos D. Grammatikakis, Giuseppe Maruccia e Francesco Papariello. OCCN: A Network-on-Chip Modeling and Simulation Framework. Em *Proceedings of the conference on Design, Automation and Test in Europe (DATE'04)*, p. 30174, Washington, DC, USA, 2004. IEEE Computer Society.
- [4] CIn. PDesigner. Website, setembro de 2007. <http://www.pdesigner.org>.
- [5] S. Roiser e P. Mato. The SEAL C++ reflection system. Em *CHEP '04: Presented in the Computing in High Energy and Nuclear Physics congress (CHEP'04)*. CERN, setembro de 2004.
- [6] OSCI. SystemC library. Website, setembro de 2007. <http://www.systemc.org>.

15. Tempo de Convergência para Equilíbrio de Nash em Jogos Não-Cooperativos

Autores: André Vignatti e Flávio Miyazawa

O surgimento, de certa forma inesperado, da internet trouxe consigo vários novos desafios para a ciência da computação. Rapidamente surgiram várias áreas com abordagens práticas propondo soluções para esses desafios. No entanto, somente nos últimos anos é que áreas mais teóricas da computação começaram uma tentativa de formalizar esses aspectos como teoria e estudar melhor esses novos problemas.

A internet é formada por várias entidades, onde cada entidade tem um objetivo próprio, e essas entidades se relacionam com outras de várias maneiras, às vezes seus relacionamentos são cooperativos, às vezes competitivos, ou também se relacionam de maneira egoísta, e neste caso, não se preocupam com as conseqüências que suas ações podem acarretar para o sistema como um todo. Esse cenário, que envolve entidades (chamados também de usuários, agentes ou jogadores), cada qual com um conjunto de estratégias para escolher, onde cada entidade tem preferências por determinadas configurações do sistema e preferências pelas estratégias que ela deve escolher, é ideal para ser modelado formalmente na *teoria dos jogos*. Mas, além da teoria dos jogos, os aspectos computacionais inerentes a esses problemas necessitam das ferramentas da computação para serem analisados. Essa nova área teórica tem sido chamada de *Teoria Algorítmica dos Jogos* e, como o nome sugere, mistura idéias de computação com a área de teoria dos jogos.

Nossa pesquisa está baseada no caso em que o relacionamento entre as entidades é egoísta, não-cooperativo. Uma estratégia egoísta de um agente pode influenciar na decisão dos outros agentes, fazendo com que esses mudem de estratégia também. Será que a seqüência de estratégias escolhidas pelos agentes pode levar a um estado do sistema onde ninguém deseja mudar sua estratégia? Em outras palavras, será que alcançaremos um estado de equilíbrio de Nash nesse sistema? Se a resposta for sim, então em quantos passos alcançaremos tal equilíbrio?

Dentro trabalhos pioneiros na área, podemos citar [1], que forneceu a idéia inicial para o estudo do equilíbrio de Nash, e a idéia de usar a razão entre uma solução em equilíbrio e uma solução ótima para medir o “quão boa” é uma situação de equilíbrio, desde então essa razão foi chamada de *preço da anarquia*. Uma técnica bastante usada, tanto para provar a existência do equilíbrio quanto para limitar o número de passos são as *funções potenciais*, inicialmente propostas por [2]. Em alguns casos a função potencial foi também usada para medir o “preço da anarquia”. Grandes avanços foram obtidos também por [3], que consideram jogos onde recursos são usados por um número infinitesimal de jogadores, ideal para modelar grandes problemas e situações onde o ambiente é descrito de forma contínua, ao invés de forma discreta.

Nosso foco está na versão de teoria dos jogos dos problemas clássicos de *balanceamento de carga* e *empacotamento* quando os *jogadores são egoístas*, mais especifica-

mente, estamos interessados em analisar o tempo de convergência para o equilíbrio de Nash nesses problemas.

O problema do *balanceamento de carga com jogadores egoístas* é definido da seguinte forma. Temos n tarefas, cada tarefa controlada por um jogador distinto, e m máquinas. Se denotarmos o peso de uma tarefa i por a_i , a j -ésima máquina por M_j e a soma dos pesos das tarefas em M_j por H_j , então o custo pago por i quando está em M_j é igual a H_j (H_j inclui o peso a_i). Suponha que a tarefa i está em M_j . Como i quer minimizar seu custo, ele irá migrar de M_j cada vez que encontrar outra máquina $M_{j'}$ tal que $H_{j'} + a_i < H_j$.

O problema do *empacotamento com jogadores egoístas* é definido da seguinte forma. Temos n itens, cada item controlado por um jogador distinto, e m bins. Todos os bins tem o mesmo custo fixo, e o custo de um bin é dividido entre todos os itens contidos no bin de acordo com a fração normalizada do bin que eles usam. Mais formalmente, se denotarmos o peso de um item i por a_i , o j -ésimo bin por B_j e a soma dos pesos dos itens empacotados em B_j por H_j , temos que o custo pago por i ao usar B_j é $\text{custo}(i, j) = a_i/H_j$. Suponha que o item i é empacotado em B_j . Como i quer minimizar seu custo, ele irá migrar de B_j cada vez que encontrar outro bin $B_{j'}$ tal que $\text{custo}(i, j') < \text{custo}(i, j)$. Essa desigualdade vale para qualquer j' tal que $H_{j'} + a_i > H_j$, portanto um item irá migrar cada vez que encontrar um bin no qual ele se encaixa melhor de acordo com o espaço não ocupado.

Vários resultados para o problema de *balanceamento de carga* são conhecidos, citaremos alguns a seguir. Em 2003, [4] deram o ponta-pé inicial no estudo do tempo de convergência de vários casos do balanceamento de carga. Em 2004, [5] estudou sobre o tempo esperado de convergência caso em que mudanças no sistema são feitas de maneira aleatória. Em 2006, [6], estudaram o cenário onde as tarefas não necessariamente migram em instantes de tempos distintos, aproximando-se mais de uma situação real de um sistema distribuído. No caso do *empacotamento*, [7] obteve alguns resultados semelhantes ao de [4], adaptando técnicas do *balanceamento de carga* para este problema. Até onde sabemos, esse é o único artigo (até a presente data, outubro/2007) que trata do problema de empacotamento egoísta.

Os dois problemas citados são de fato bastante parecidos, no entanto, não há uma tradução direta dos resultados. O esforço de nossa pesquisa atualmente está em adaptar resultados existentes do problema de balanceamento para o problema de empacotamento, usando idéias e técnicas parecidas com as que já foram utilizadas no problema de balanceamento. Mais especificamente, conseguimos demonstrar a equivalência entre o empacotamento no caso de itens com tamanho iguais e o problema de balanceamento com tarefas de pesos iguais. Além disso, estamos trabalhando em adaptações da análise probabilística de [5] para o problema do empacotamento.

Referências

- [1] E. Koutsoupias and C. Papadimitriou. *Worst-Case Equilibria Lecture Notes in Computer Science*, 1563:404–413, 1999.
- [2] R. Rosenthal. *A Class of Games possessing pure-strategy Nash equilibria International Journal of Game Theory*, 2, 65–67, 1973.
- [3] T. Roughgarden and E. Tardos. *How bad is selfish routing? Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pp. 93, 2000.
- [4] E. Even-Dar and A. Kesselman and Y. Mansour. *Convergence time to Nash equilibrium in load balancing ACM Trans. Algorithms*,3,3, pp 32,2007.
- [5] P.W. Goldberg. *Bounds for the convergence rate of randomized local search in a multiplayer load-balancing game, Proceedings of the twenty-third annual ACM symposium on Principles of distributed computing*, 131–140, 2004
- [6] P. Berenbrink and T. Friedetzky and L.A. Goldberg and P. Goldberg and Z. Hu and R. Martin. *Distributed selfish load balancing, Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*,354–363, 2006.
- [7] V. Biló. *On the packing of selfish items, Proceedings of the 20th Parallel and Distributed Processing Symposium*, pp 9–, 2006.

16. Uplink Scheduling with Quality of Service in IEEE 802.16 Networks

Authors: Juliana Freitag Borin and Nelson L. S. da Fonseca

To support a wide variety of multimedia applications, the IEEE 802.16 standard [1] defines five types of service flows, each one with different QoS requirements and a corresponding uplink scheduling policy. The Unsolicited Grant Service (UGS) supports constant bit rate (CBR) or CBR-like applications such as VoIP. This service receives fixed size data grants periodically without explicit requests from the subscriber stations (SSs). The real-time Polling Service (rtPS) is designed to support real-time service flows that generate variable size data packets on a periodic basis, e.g. MPEG video streams. This service offers periodic unicast request opportunities and QoS guarantees in terms of bounded delay and minimum bandwidth. The extended real time Polling Service (ertPS) uses a grant mechanism similar to the one used to support UGS connections. The difference is that periodic allocated grants can be used to send bandwidth requests to inform the required grant size. The non-real-time Polling Service (nrtPS) supports flows which require a better-than-best-effort service such as FTP applications. This service offers periodic unicast request opportunities, but using more spaced intervals than rtPS, and minimum bandwidth guarantee. The Best Effort service (BE) is used for best-effort traffic such as HTTP. The nrtPS and the BE service share contention request opportunities.

The MAC layer assigns all the arriving packets to connections according to their QoS needs. Each connection is associated to a service flow. The resource allocation for SSs is on demand. When an SS has data to transmit, it sends a bandwidth request to the base station (BS). The BS, in turn, allocates time slots for the SS based on the bandwidth request and on the QoS requirements of the connection. The time slots are organized in frames, and each frame is divided in two parts: the downlink subframe, used by the BS to send data and control information to the SSs, and the uplink subframe, shared among all the SSs to send their data to the BS.

Although the five service levels furnish the basis for QoS provision, the core of this task resides in the resource allocation, i.e., in the scheduling mechanism. An efficient scheduling algorithm is an essential condition to support QoS requirements, and it has a great influence on the network performance. Different scheduling mechanisms [2, 4–6] have been proposed in the literature, however, not all of them comply with the IEEE 802.16 standard. Moreover, most of them are based on scheduling policies, such as *Weighted Fair Queueing*, which are not so simple to implement.

In this paper, we introduce a BS uplink scheduling algorithm which allocates bandwidth for the SSs based on the bandwidth requests and on the QoS requirements of each service flow. The proposed scheme is full standard-compliant and can be easily implemented in the BS. Results obtained through simulation experiments using ns-2 tool show that the proposed scheme is able to support QoS guarantees.

The Proposed Scheduling Approach

The proposed BS uplink scheduler uses three priority queues, which we will refer to as low priority queue, intermediate queue and high priority queue. The scheduler always serves a queue with higher priority before one with lower priority as long as there is a packet in the higher-priority queue. The low priority queue stores BE bandwidth requests. The intermediate queue holds bandwidth requests sent by rtPS and nrtPS connections. rtPS and nrtPS requests can migrate to the high priority queue to guarantee that the QoS requirements are met. Besides the requests migrated from the intermediate queue, the high priority queue stores periodic grants and unicast request opportunities that must be scheduled in the next frame. The BS executes the uplink scheduler in each frame, and it broadcasts the scheduling result to the SSs through the UL-MAP message. Fig. 5 shows the architecture of the proposed BS uplink scheduler.

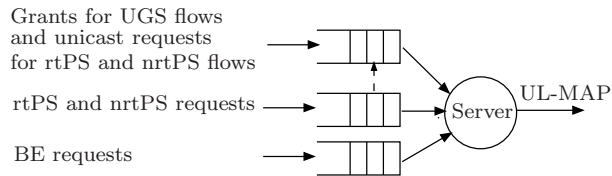


Figure 5: Architecture of the proposed uplink scheduler

At each frame, the scheduler generates and inserts the periodic grants in the high priority queue according to a pre-defined interval. The interval between UGS and ertPS grants is informed by SSs during the connection establishment, while the intervals for request opportunities are defined by the BS. In this way, we guarantee that UGS and ertPS grants, and rtPS and nrtPS request opportunities are provided as specified by the standard. To guarantee the maximum delay requirement, the BS assigns a deadline for each rtPS bandwidth request in the intermediate queue. Each time the scheduler is executed, the requests whose deadline expires in the frame that follows the next one migrate from the intermediate queue to the high priority queue. To calculate the deadline, it would be necessary to know the arriving time of the packets in the SSs' queues. Since the BS has no access to this information, we consider the worst case, i.e., that the packet arrived in the queue immediately after the corresponding connection has sent a bandwidth request. Hence, the deadline of a request is equal to the sum of the arriving time of the last request sent by the corresponding connection and the connection maximum delay requirement.

The scheduler guarantees the minimum bandwidth requirement of rtPS and nrtPS traffic over a window of duration T . Each time the scheduler is executed, it calculates a priority value for each request in the intermediate queue considering the following information of the corresponding connection: minimum bandwidth requirement, amount of backlogged requests (in bytes), and amount of received bandwidth in the window T . Requests of connections which has already received

the minimum bandwidth in the current window acquire low priority values. For the remaining requests, the less bandwidth the connection has received, the higher is the priority of its requests.

To conduct this study, we have implemented a ns-2 module for IEEE 802.16 networks [3]. The purpose of this study is to explore the scheduler capabilities and to evaluate its behavior in ideal channel conditions.

The topology of the simulated network consisted of a BS with the SSs uniformly distributed around it. The frame duration was 5 ms and the capacity of the channel was 40 Mbps, assuming a 1:1 downlink-to-uplink TDD split. In order to prevent that packet scheduling in the SSs affects the results, each SS has only one traffic flow. We consider four types of traffic: voice, voice with silence suppression, video, FTP and WEB, which are associated to UGS, ertPS, rtPS, nrtPS, and BE services, respectively.

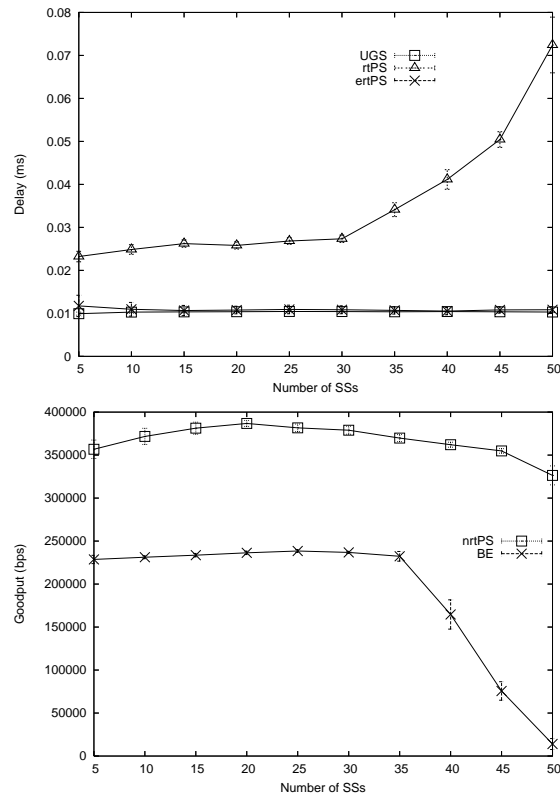


Figure 6: Delay and goodput of the service flows

Fig. 6 (top graphic) shows the average delay of UGS, rtPS, and ertPS uplink connections versus the number of SSs. The delay of UGS and ertPS connections was not affected by the load increase, which shows that the uplink scheduler is able to provide data grants at fixed intervals as required by these services. Conversely, the delay of the rtPS connections increases with the offered load, however, it does not surpass the required one.

As can be seen in the bottom graphic of Fig. 6, the goodput of the nrtPS connections decreased slightly as the offered load increased. Nonetheless, all the nrtPS connections had the minimum bandwidth requirement guaranteed. On a different trend, the throughput of the BE connections decreased sharply when the system is overloaded, which was expected due to the load increase of the higher priority service flows.

The simulations experiments show the efficacy of the proposed scheme. If there are connections of different service levels in the network, then the scheduler allocates enough slots for each connection such that the QoS requirements are provided. Furthermore, the proposed solution uses a simple approach based on three priority queues and the scheduling algorithm does not require extensive calculations. In fact, the complexity of the proposed algorithm is $O(k + n + r \log r)$, where k is the number of slots in the uplink subframe, n is the number of connections, and r is the number of bandwidth requests in the intermediate queue.

Acknowledgements. This research was sponsored by UOL (www.uol.com.br), through its UOL Bolsa Pesquisa program, process number 20060511022200a.

Referências

- [1] *IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems.* IEEE Std., Rev. IEEE Std802.16-2004, 2004.
- [2] J. Chen, W. Jiao, and H. Wang, *A Service Flow Management Strategy for IEEE 802.16 Broadband Wireless Access Systems in TDD Mode.* In Proceedings of the IEEE International Conference on Communications, pp. 3422-3426, 2005.
- [3] J. Freitag and N. L. S. da Fonseca, *WiMAX Module for the ns-2 Simulator.* In Proceedings of the IEEE PIMRC'07, pp. 1-6, 2007.
- [4] G. Chu, D. Wang, and S. Mei, *A QoS Architecture for the MAC Protocol of IEEE 802.16 BWA System.* In Proceedings of the IEEE International Conference on Communications, Circuits and Systems and West Sino Expositions, pp. 435-439, 2002.
- [5] M. Hawa and D. W. Petr, *Quality of Service Scheduling in Cable and Broadband Wireless Access Systems.* In Proceedings of the 10th IEEE International Workshop on Quality of Service, pp. 247-255, 2002.
- [6] K. Wongthavarawat and A. Ganz, *IEEE 802.16 based last mile broadband wireless military networks with Quality of Service support.* In Proceedings of the IEEE MILCOM'03, pp. 779-784, 2003.

17. Testes baseados em modelos e injeção de falhas na validação da robustez de protocolos de comunicação

Autores: Thaise Yano, Eliane Martins e Fabiano Luís de Sousa

Robustez é “o grau ao qual um sistema ou componente pode funcionar corretamente na presença de entradas inválidas ou condições ambientais estressantes” [7]. Recentemente, a SEI (*Software Engineering Institute*) produziu uma nota técnica para guiar o teste de robustez de sistemas [6]. Esta nota diz que é essencial determinar se o sistema pode lidar com entradas inesperadas de usuários ou falhas de sistema para validação do software. Teste de robustez é imprescindível para a revelação de vulnerabilidades do sistema que poderia aparecer muito tardiamente no ciclo de desenvolvimento ou após sua entrega. Além disso, dada a tendência atual no desenvolvimento de software para reuso de componentes ou mesmo de aplicações inteiras em diferentes contextos, também explica o crescimento na importância de validação do comportamento do software na presença de condições não previstas.

Teste de conformidade visa a estabelecer se uma dada implementação está conforme a sua especificação. Este tipo de teste tem sido utilizado tradicionalmente para o teste de protocolos. Embora seja um passo necessário na validação de protocolos, o teste de conformidade é limitado no que diz respeito ao teste na presença de falhas que afetam *links* ou *host* remotos, por exemplo. Portanto, teste de robustez é um complemento útil para os testes de conformidade.

Há várias maneiras de aplicar teste de robustez. Duas abordagens que podem ser consideradas são: formal e experimental. A abordagem formal inspira-se nas técnicas de teste de conformidade, considerando então a derivação de modelos para representar o comportamento do sistema no caso normal como também em presença de falhas. A desvantagem dessa abordagem é que as falhas que podem ser consideradas são limitadas, devido normalmente ao software em teste não ter controle sobre o ambiente de execução. A fim de compensar esta limitação, propomos complementar esses testes com uma abordagem experimental, baseada em injeção de falhas.

Em abordagens de injeção de falhas, falhas são deliberadamente introduzidas no sistema para observar seu comportamento. Esta abordagem é útil em muitos aspectos: (i) entender os efeitos de falhas reais; (ii) verificar o comportamento do sistema em presença de falhas, obtendo *feedback* para a correção ou melhorias do sistema; (iii) estimar experimentalmente parâmetros úteis para medir a dependência do sistema. Entretanto, com a exceção de poucos trabalhos [4,9], a maioria da injeção de falhas para a validação de protocolos não considera modelos de seu comportamento.

Assim, propomos uma abordagem híbrida para o teste de robustez, que combina técnicas de conformidade e injeção de falhas. Em trabalhos anteriores essas técnicas foram aplicadas juntas na metodologia de teste CoFI (*Conformance and Fault Injection*) proposta para validar aplicações espaciais [1,3]. Adaptamos a CoFI para o teste de protocolo, a fim de permitir a geração automática de casos de teste

para exercitar cenários de erro como também as falhas que ativam os erros nesses cenários.

O comportamento do protocolo é representado por modelos baseado em estados utilizando Máquinas de Estados Finitos Estendida (MEFE), que é definida pela 6-tupla $M = \langle S, s_0, V, v_0, P, I, O, F, T \rangle$ em que: S é o conjunto finito de estados, s_0 é o estado inicial, V é o conjunto finito de variáveis, v_0 são os valores iniciais das variáveis, P é o conjunto de parâmetros de entrada, I é o conjunto de entradas (eventos externos), O é o conjunto de saídas, $F \subseteq S$ é um conjunto de estados finais, T é o conjunto finito de transições. Os elementos de T são uma 4-tupla $t = \langle s_s, g, act, s_f \rangle$ que representam respectivamente: o estado de início (atual) de t ; guarda, i.e., uma expressão lógica envolvendo condições sobre parâmetros e variáveis; ação executada quando a transição é ativada; e o estado de término (próximo) de t . Além da ocorrência de um evento externo, para disparar uma transição é necessário satisfazer a guarda associada. Quando uma transição é disparada, a ação correspondente é executada, podendo mudar valores de variáveis e parâmetros ou também produzir uma saída. Quando a ação é finalizada, a máquina vai ao estado s_f , que pode ser o mesmo de s_s .

Em relação ao teste de robustez, consideramos que o conjunto de entradas I é composto por: conjunto de falhas F e conjunto de eventos externos E . O último contém os eventos externos aceitos pelo sistema que ativarão a implementação em teste (IeT) e o primeiro é definido de acordo com o modelo de falhas considerado. Em nosso caso, o interesse é determinar quão robusto é a IeT em presença de falhas de comunicação, isto é, falhas que podem afetar *link* ou *host* remotos. Como descrito comumente na literatura, falhas afetam as mensagens para IeT, que podem ser suprimidas, atrasadas, replicadas ou corrompidas. Em outras palavras, apenas eventos externos correspondentes às mensagens trocadas entre a IeT e seu *peer* podem ser afetadas pelas falhas. Cada falha é caracterizada por um conjunto de atributos, baseado em [2]: (i) tipo, i.e., *qual* tipo de falha deve ser injetada. As falhas representam modos de falhas de comunicação; (ii) instante (tempo ou identificação da mensagem) *quando* a falha deve ser injetada; (iii) máscara indica *como* a mensagem será corrompida; (iv) localização da falha designa *onde* as falhas serão injetadas. Dado que as falhas afetam mensagens do fluxo do canal de comunicação, a localização representa campos das mensagens; (v) número indica *quantas vezes* uma falha deve ser injetada.

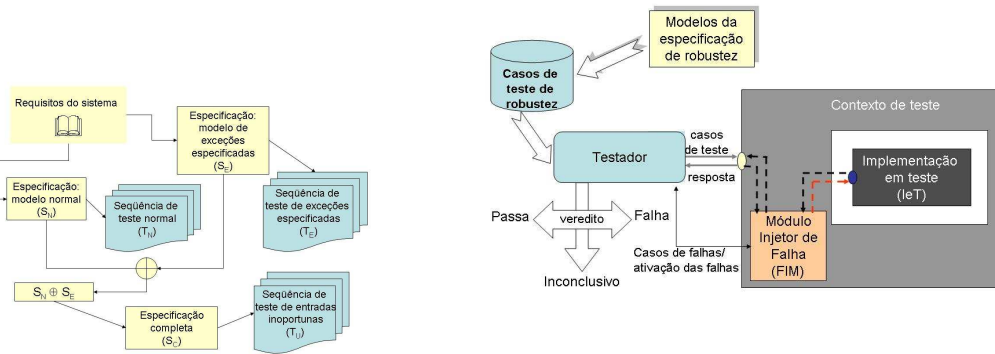
O conjunto de saída O também pode ser considerado como dividido em dois conjuntos: O_N , composto por saídas produzidas em situações normais, e O_F , conjunto de saídas produzidas em presença de falhas. Em geral, O_F contém saídas extras tais como mensagens de erros, exceções ou códigos de saída de erro.

O comportamento do sistema é representado em diversos modelos baseados em estados, que representam o comportamento parcial do sistema. Ao invés de criar um único modelo, diferentes modelos são construídos para representar o comportamento de um protocolo: especificação normal (S_N), especificação de exceção (S_E) e especificação completa (S_C). Isso evita lidar com um grande número de diferentes classes

de entradas de uma só vez. Os modelos S_N e S_E representam, respectivamente, o comportamento normal da IeT e em presença de falhas prevista na especificação. Para gerar S_C , constrói-se um modelo contendo tanto o comportamento normal e de exceções especificadas. Em seguida completa-se o modelo conforme proposto por Binder para o teste de caminhos furtivos [5].

A abordagem de geração de casos de teste é mostrada na Figura 1.a. A seqüência de teste T_N gerada de S_N constitui o teste de conformidade. Já a seqüência de teste T_E gerada de S_E visa ao teste das exceções especificadas. A seqüência de teste T_U para entradas inoportunas são obtidas de S_C , a fim de determinar se um evento ilegal pode se aceitar pela IeT. O teste de robustez é constituído por $(T_E \cup T_U)$. Portanto determina-se se o protocolo é robusto ou não de acordo com os modelos de comportamento.

Uma visão abstrata do teste de robustez encontra-se na Figura 1.b. Embora bastante similar ao teste de conformidade, no teste de robustez há a adição de um injetor de falhas no contexto, para interceptar as mensagens enviadas do testador para a IeT. Além do controle e observação da IeT, o testador também controla o módulo de injetor de falhas (FIM) com os caso de falhas.



(a) Arquitetura do teste de robustez

(b) Geração de teste

Figura 7: Abordagem de teste

Um estudo de caso da metodologia proposta foi realizado com o protocolo WTP (*Wireless Transaction Protocol*), que é uma das camadas do WAP (*Wireless Application Protocol*). Um dos principais objetivos do WAP é oferecer acesso a Internet para dispositivos móveis, como telefones celulares. Uma vez construído os modelos, esses foram submetidos à ferramenta Condado [8] para a geração de casos de teste.

Uma vantagem de nossa proposta é que as falhas consideradas não são limitadas a aquelas controladas por um testador clássico. Em contraste à maioria dos trabalhos de injeção de falhas para validação de protocolos, mostramos como integrar as falhas nos modelos de comportamento do protocolo. Como um resultado, os casos de teste e casos de falha podem ser gerados juntos através da derivação de teste dos modelos

para o teste do protocolo. Além disso, o veredito de robustez é dado baseando-se nos modelos de comportamento.

Futuros trabalhos estão direcionados à geração automática de dados para parâmetros e condições de guarda das transições. Em geral a geração de dados é aleatória e métodos mais efetivos são necessários, tal como o uso de algoritmos evolutivos.

Referências

- [1] A. M. Ambrosio, E. Martins, N. L. Vijaykumar, and S. V. Carvalho. A conformance testing process for space applications software services. *Journal of Aerospace Computing, Information, and Communication*, 3(4):146–158, April 2006.
- [2] A. M. Ambrosio, E. Martins, N. L. Vijaykumar, and S.V. Carvalho. A methodology for designing fault injection experiments as an addition to communication systems conformance testing. In *1st Workshop on Dependable Software - Tools and Methods in the IEEE Conference on Dependable System and Network*, 2005.
- [3] A. M. Ambrosio, F. Mattiello-Francisco, V. A. Santiago, W. P. Silva, and E. Martins. Designing fault injection experiments using state-based model to test a space software. In *3rd Latin-American Symposium on Dependable Computing (LADC)*, pages 170–178, Morelia, Mexico, September 2007.
- [4] D. Avresky, J. Arlat, J.-C. Laprie, and Y. Crouzet. Fault injection for the formal testing of fault tolerance. In *2nd International Symposium on Fault-Tolerant Computing*, pages 345–354, 1992.
- [5] R. V. Binder. *Testing object-oriented systems: models, patterns, and tools*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.
- [6] J. Cohen, D. Plakosh, and K. Keeler. Robustness testing of software intensive systems: explanation and guide. Technical Report CMU/SEI-2005-TN-015, CMU/SEI, 2005.
- [7] IEEE. IEEE standard glossary of software engineering terminology, padrão 610.12, 1990. IEEE Press.
- [8] E. Martins, S. B. Sabião, and A. M. Ambrosio. ConData: a tool for automating specification-based test case generation for communication systems. *Software Quality Journal*, 8(4):303–319, December 1999.
- [9] N.Suri and P.Sinha. On the use of formal techniques for validation. In *28th Annual International Symposium on Fault-Tolerant Computing*, pages 390–399, 1998.

18. Um modelo para suporte à avaliação da qualidade da educação à distância no ensino superior via Internet

Autores: Ricardo Luís Lachi e Heloísa Vieira da Rocha

Avaliar a qualidade de um curso a distância não é uma tarefa simples. Na literatura existem vários trabalhos apresentando diferentes abordagens sobre esse tema sendo, a principal linha de pesquisa, a avaliação de um curso por meio da definição de indicadores de qualidade.

O objetivo deste trabalho é justamente apresentar um modelo que dê suporte a avaliação de cursos superiores à distância via Internet por meio de indicadores. As duas principais dificuldades envolvidas neste tema dizem respeito aos conjuntos de indicadores de qualidade existentes na literatura, sendo essas: 1) a compreensão do que deve ser analisado sobre um determinado indicador proposto por algum trabalho da literatura e, uma vez adotado um determinado conjunto de indicadores, 2) a obtenção de informações relevantes sobre cada um destes, especificamente.

No trabalho publicado em [1] são descritas essas duas dificuldades mencionadas, as formas como elas podem ser superadas, além de todo um levantamento feito da literatura de trabalhos correlacionados para dar respaldo à abordagem proposta. Uma representação esquemática do modelo planejado para dar suporte à avaliação por meio de indicadores pode ser visto na **Figura 1**.

Nessa **Figura 1**, o modelo começa a partir dos 24 indicadores considerados relevantes no estudo IHEP [2]. Para cada um desses indicadores foram elaborados conjuntos de perguntas a partir de referências importantes existentes na literatura [3–7]. Essas perguntas são organizadas em questionários específicos que devem ser aplicados nas entidades envolvidas no curso à distância sob avaliação (professor, aluno, instituição).

Uma vez obtidas todas as respostas aos questionários, é aplicada a técnica estatística multivariada denominada Análise Fatorial. O resultado da aplicação dessa técnica permite apontar objetivamente quais perguntas podem ser agregadas em um mesmo grupo, ou como é o caso, em um mesmo indicador [8]. Em palavras mais simples, quais as perguntas que explicam de fato um determinado indicador.

Na seqüência, será determinado o coeficiente Alfa de Cronbach para cada um dos 24 indicadores. Isso permitirá definir se as perguntas agrupadas em um determinado indicador têm consistência interna entre si, ou seja, se elas apresentam uma relação coerente entre si no esforço de medir o indicador [8]. Por meio dessa análise estatística, é possível eliminar as perguntas que tem o menor peso para o valor de

Alfa de Cronbach.

Concluído esse processo inicial, é apresentado na **Figura 1** o Modelo de Avaliação Final. Esse deve aplicado e as respostas coletadas servirão como base para a avaliação do curso à distância. Com base nessas respostas, é prevista a apresentação de 3 relatórios.

O primeiro envolve a apresentação das respostas que indicam se o curso à distância cumpre ou não aspectos exclusivamente técnicos - como por exemplo: a necessidade da senha ter no mínimo 8 caracteres, no caso em que há a disposição um ambiente de suporte a aprendizagem via Internet. A definição de quais são os requisitos técnicos mínimos a serem cumpridos, foram extraídos de literatura técnica específica. Um ponto interessante a ser observado é que, caso haja a disposição um ambiente de educação à distância (EaD) disponível, essas informações técnicas podem estar armazenadas nele de forma que possam ser coletadas automaticamente nesta etapa. O segundo relatório a ser apresentado envolve a compilação de inúmeras estatísticas descritivas sobre as respostas coletadas, tais como, média, mediana, desvio padrão, variâncias, etc. Por fim, o último relatório procura relacionar as respostas obtidas nos questionários, tentando apontar aquelas com maior correlação.

Finalizando, esse modelo proposto é robusto e flexível o suficiente para a construção de indicadores de qualidade relevantes para cursos à distância. Robusto porque se baseia em questões estatisticamente significativas a respeito de um indicador. Flexível porque permite a inclusão de novas questões e/ou novos indicadores desde que sejam devidamente validados. Além disso, também inova porque foge da atual tradição existente na literatura da fundamentação baseada em estudos de casos puramente empíricos.

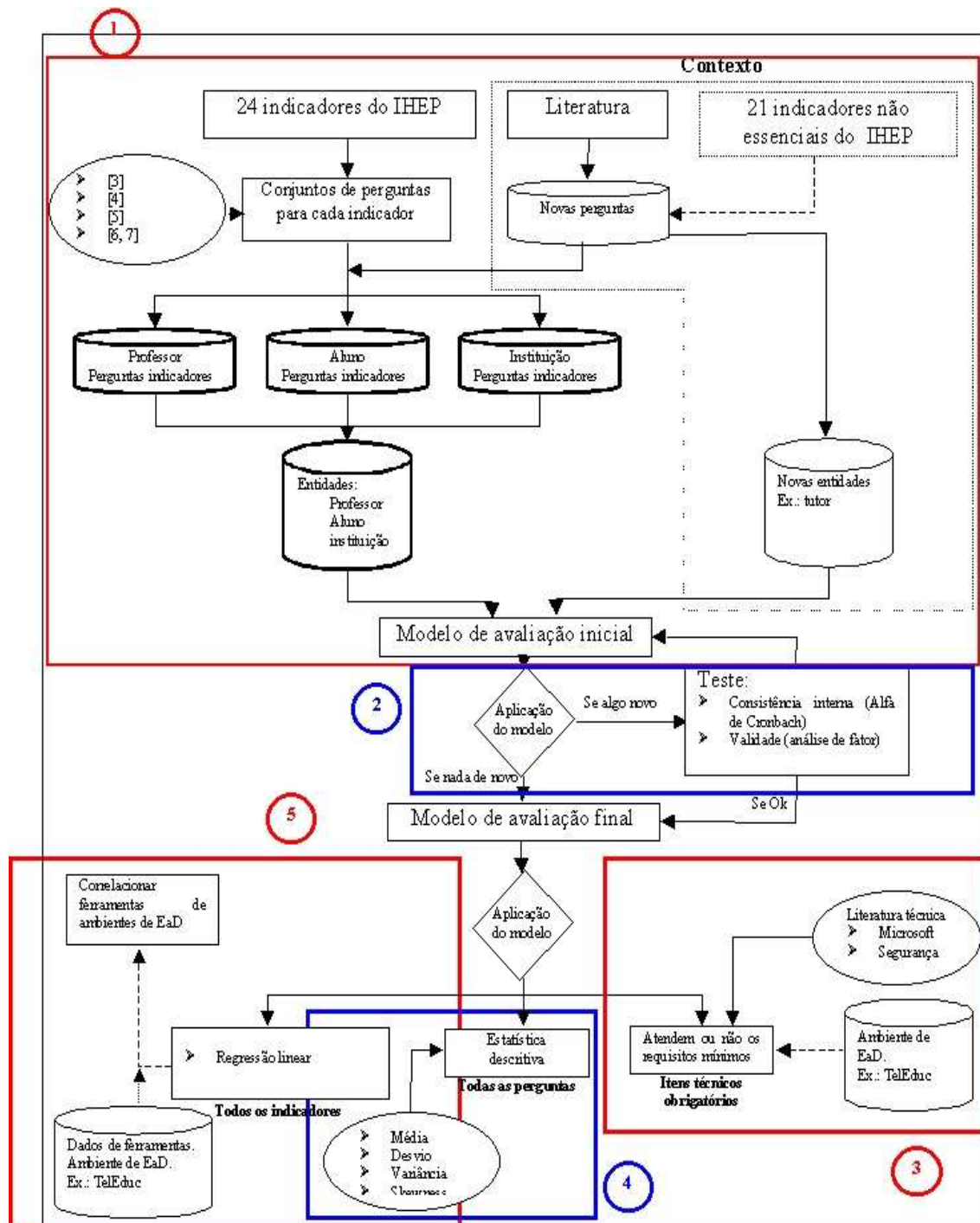


Figura 1. Modelo de avaliação de um curso à distância de nível superior via Internet com destaque para as 5 ferramentas computacionais necessárias para dar suporte a esse modelo: 1) construção dos questionários; 2) Validação do modelo; 3) Ferramenta integrada com um ambiente de EaD para coleta dos dados técnicos; 4) Ferramenta para a geração das estatísticas descritivas; 5) Ferramenta que relaciona as respostas obtidas explicitando aquelas com maior correlação estatística.

Referências

- [1] LACHI, R. L.; ROCHA, H. V. da; OEIRAS, J. Y. Y. *Avaliação de cursos a distância: uso de indicadores para assegurar qualidade*. Simpósio Brasileiro de Informática na Educação. 2006.
- [2] PHIPPS, R.; MERISOTIS, J. *Quality on the Line - benchmarks for success in Internet-based distance education*. Washington, DC: The Institute for Higher Education Policy, 2000. 42p (Relatório).
- [3] MEC. Ministério da Educação. Secretaria de Educação a Distância. *Indicadores de qualidade para cursos de graduação à distância*. Brasília. 2001.
- [4] DEMO, Pedro. *Educação e Qualidade*. 10.ed. Campinas: Papyrus, 1996. v.1. 160 p.
- [5] SINAES. *Avaliação externa de instituições de educação superior - Diretrizes e Instrumento*. MINISTÉRIO DA EDUCAÇÃO. Comissão Nacional de Avaliação da Educação Superior. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. 2006. Disponível em: "http://www.inep.gov.br/download/superior/2005/avaliacao_institucional/Avaliacao_Institucional_Externa.pdf". Acesso em: 13 jan. 2007.
- [6] EATON, J. S. *Maintaining the delicate balance: distance learning, higher education accreditation, and the politics of self-regulation*. ACE - American Council on Education, Washington, DC. 2002.
- [7] HOPE, A. *Quality Assurance*. In: FARRELL, G. (ed.). *The Changing Face of Virtual Education*. Vancouver: Commonwealth of Learning, 2001.
- [8] PEREIRA, J. C. R. *Análise de Dados Qualitativos: Estratégias Metodológicas para as Ciências da Saúde, Humanas e Sociais*. 3. ed. - São Paulo: Editora da Universidade de São Paulo, 2001.
- [9] SILVA, C. G. da; ROCHA, H. V. da. *Exploração de bases de dados de ambientes de Educação a Distância por meio de ferramentas de consulta apoiadas por Visualização de Informação*. Tese de doutorado. Instituto de Computação - UNICAMP. 2006.

19. Redes Mesh Autônômicas

Autores: Neumar Malheiros, Edmundo Madeira

Introdução

O controle e gerência de Redes Mesh Sem Fio [1] são tarefas bastante complexas devido à natureza dinâmica e heterogênea de sua arquitetura, à mobilidade e heterogeneidade dos elementos de rede e aos requisitos de auto-configuração. As tradicionais abordagens de gerência centralizada, baseadas em um modelo cliente e servidor, não são viáveis neste contexto.

Para lidar com esse problema, este projeto de pesquisa propõe investigar uma abordagem para gerência de Redes Mesh baseada nos princípios de um novo paradigma de sistemas computacionais, denominado *Autonomic Computing* [2]. Este paradigma consiste em sistemas computacionais que são capazes de gerenciar a si mesmos, de acordo com objetivos de alto nível.

O objetivo da presente proposta de pesquisa é projetar uma arquitetura distribuída para prover funções de auto-gerência em Redes Mesh. Em princípio, as propriedades autônômicas consideradas são: auto-configuração e auto-otimização.

Os principais desafios desta pesquisa envolvem questões como: Como prover a auto-gerência do sistema a partir da interação coordenada entre elementos de rede autônomos? Como garantir a convergência de mecanismos distribuídos de auto-gerência regulados por políticas de alto-nível? No que diz respeito à estabilidade, até que ponto é possível manter a qualidade de serviço diante da dinâmica da rede? Como um dado parâmetro/evento (local) afeta o comportamento do sistema como um todo? Como coordenar a interação entre elementos de rede autônomos, a fim de otimizar o desempenho e compartilhar de forma justa os recursos da rede?

Metodologia

A abordagem para o projeto de Redes Mesh Auto-Gerenciáveis consiste em projetar uma arquitetura na qual elementos de rede auto-gerenciáveis cooperam entre si. Desta maneira, a capacidade de auto-gerência de rede é resultado da interação coordenada entre os seus elementos auto-gerenciáveis.

O projeto da arquitetura envolve definir suas funções e como elas se relacionam. As funções da arquitetura são divididas em duas categorias: **funções de auto-gerência** (são as funções de auto-configuração e auto-otimização através dos quais um elemento de rede gerencia a si mesmo) e **função básicas** (são as funções de suporte às funções de auto-gerência). A Figura 8 apresenta uma visão geral da arquitetura incluindo as principais funções que serão desenvolvidas neste projeto.

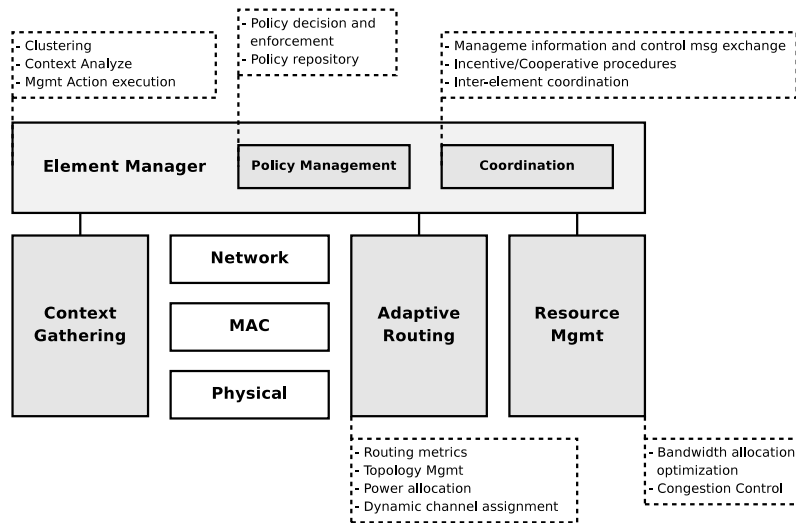


Figura 8: Visão geral da arquitetura para Redes Mesh Autônomicas.

Funções Básicas

Coordenação entre elementos A arquitetura deve incluir mecanismos de comunicação e coordenação que suportem a transferência de informações de gerência e a troca de mensagens de controle entre os elementos de rede. As mensagens de controle viabilizam a interação (coordenação) entre os elementos. Assim, os elementos podem compartilhar informações de controle e disponibilidade de recursos em suporte às funções de gerência distribuídas.

Monitoramento de contexto Qualquer operação autônoma exige que o elemento autônomo conheça a si mesmo e o seu ambiente de execução. A arquitetura deve possuir suporte a conhecimento de contexto (*context-aware*). Primeiro, cada elemento de rede precisa monitorar seu próprio estado (contexto interno). Segundo, cada elemento deve ser capaz de monitorar as condições da rede (contexto externo). É importante obter as informações de contexto em várias camadas da rede (*cross-layer context gathering*).

Gerência de políticas Outra função básica da arquitetura é o suporte a políticas que representem objetivos de alto nível definidos pela administração da rede. Esses objetivos devem refletir o comportamento e o desempenho desejados, mas não precisam, necessariamente, definir como alcançá-los. As políticas são de fundamental importância nesta arquitetura, pois elas servem para regular e direcionar os mecanismos de auto-configuração e auto-otimização.

Funções de Auto-Gerência

Roteamento “Adaptativo” Os elementos de rede devem ser capazes de cooperar entre si para otimizar a alocação de canais e o cálculo de rotas. Objetivo aqui

é desenvolver um framework capaz de adaptar o mecanismo de roteamento em face a alterações no estado da rede a fim de maximizar a vazão média. O desenvolvimento dessa função envolverá a definição de novas métricas de roteamento e algoritmos para auto-configuração do mecanismo de roteamento.

Gerência de Recursos A arquitetura deve incluir mecanismos para gerenciar e otimizar a alocação de recursos entre os elementos de rede de forma distribuída. Neste caso, será necessário um algoritmo distribuído para alocação de largura de banda a fim de: otimizar o desempenho do sistema (em termos de vazão média); garantir compartilhamento justo (fair) dos recursos entre os elementos de rede; e controlar congestionamento.

Avaliação da Proposta

A avaliação da proposta será realizada a partir de simulações utilizando o simulador ns-2. Diversos trabalhos utilizaram este simulador para avaliar soluções para Redes Mesh. A avaliação envolve estender o simulador (implementar no simulador os mecanismos e algoritmos propostos). Os fatores críticos na simulação são: definir corretamente o modelo de mobilidade e de tráfego e definir cenários realistas.

O trabalho de Conti e Giordano [3] discute graves problemas na pesquisa em redes ad hoc multihop e enfatiza o seguinte: a pesquisa nesta área alcançou muitos resultados teóricos, mas a falta de cenários realistas não permitiu avanços no desenvolvimento dessas redes. A pesquisa nesta área está focada no desenvolvimento de redes complexas de larga escala sem infra-estrutura, mas não existe uma demanda que justifica isso (exceto no setor militar). Seria mais importante pesquisar e desenvolver redes de propósito geral de pequena escala como Redes Mesh para aplicações como: redes de acesso comunitárias; compartilhamento de arquivos; e acesso à Internet.

Trabalhos Relacionados

O trabalho de Wu et al [4] propõe uma solução entre a camada MAC e a camada de rede considerando em conjunto o problema de alocação da canal e cálculo de rota. Nesta proposta foi definida um nova métrica baseada na interferência entre os canais. Outro importante trabalho [5] confirma as tendência de se considerar em conjunto alocação de canal e roteamento.

Em [6], foi proposta uma nova métrica de roteamento baseada no Expected Transmission Time (ETT). O mecanismo proposto define pesos para cada enlace de acordo com o ETT e calcula o peso acumulado de um caminho, considerando a interferência entre enlaces que usam o mesmo canal.

Todos esses trabalhos apresentam limitações no que diz respeito à avaliação das propostas. As simulações realizadas não consideram cenários do mundo real. Algumas suposições definidas são distantes do mundo real (por exemplo, disponibilidade

em tempo real de informações de estado). Os cenários de simulação consideram apenas alguns poucos fluxos de tráfego entre nós específicos da rede (as caracterização do tráfego não é realista).

Além dos trabalhos sobre roteamento discutidos anteriormente, serão considerado os resultados na pesquisa sobre gerência de recursos. No trabalho de He et al. [7], foi proposto um mecanismo baseado na Teoria de Controle para reajustar dinamicamente a taxa de transmissão para um fluxo.

A proposta apresentada em [8] identifica regiões de contenção definidas como cliques maximais no grafo de conexão da rede e calcula um preço para cada região. A taxa de transmissão para um fluxo é ajustada de acordo com o preço das regiões de contenção pelas quais o fluxo passa, de acordo com um lei de preço x demanda (maior o preço, menor a taxa de transmissão).

Referências

- [1] I. F. Akyildiz e X. Wang. A Survey on Wireless Mesh Networks. *IEEE Communications Magazine*, 43(9):S23–S30, 2005.
- [2] S. Dobson; S. Denazis; A. Fernández; D. Gaïti; E. Gelenbe; F. Massacci; P. Nixon; F. Saffre; N. Schmidt; e F. Zambonelli. A Survey of Autonomic Communications. *ACM Transactions on Autonomous and Adaptive Systems*, 1(2):223–259, 2006.
- [3] M. Conti e S. Giordano. Multihop ad hoc networking: The theory. *IEEE Communications Magazine*, 45(4):78–86, 2007.
- [4] H. Wu; F. Yang; K. Tan; J. Chen; Q. Zhang; e Z. Zhang. Distributed channel assignment and routing in multiradio multichannel multihop wireless networks. *IEEE Journal on Selected Areas in Communications*, 24(11):1972–1983, 2006.
- [5] M. Alicherry; R. Bhatia; e L. E. Li. Joint Channel Assignment and Routing for Throughput Optimization in Multiradio Wireless Mesh Networks. *IEEE Journal on Selected Areas in Communications*, 24(11):1960–1971, 2006.
- [6] R. Draves; J. Padhye; e B. Zill. Routing in multi-radio, multi-hop wireless mesh networks, 2004.
- [7] W. He; X. Liu; e K. Nahrstedt. A feedback control scheme for resource allocation in wireless multi-hop ad hoc networks. Em *The 2nd International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2005. MobiQuitous 2005*, p. 432–439. ACM Press New York, NY, USA, 2004.
- [8] Y. Xue; B. Li; e K. Nahrstedt. Optimal Resource Allocation in Wireless Ad Hoc Networks: A Price-Based Approach. *IEEE Transactions on Mobile Computing*, 5(4):347–364, 2006.

20. Políticas Semânticas para Serviços Web

Autores: Diego Zuquim Guimarães Garcia e Maria Beatriz Felgar de Toledo

A tecnologia de serviços Web possui propriedades importantes para o desenvolvimento e a execução de aplicações distribuídas e para a realização de operações de negócio eletrônico dentro e entre fronteiras organizacionais [1].

Serviços Web representam uma tentativa de criar um conjunto de padrões abertos para a interoperabilidade entre sistemas [1]. A tecnologia de serviços Web realiza a aplicação da Arquitetura Orientada a Serviços na Web [9]. A definição de serviço Web incluída no documento do World Wide Web Consortium (W3C) [4] consiste de: (i) sistema de *software* projetado para apoiar interações entre máquinas em uma rede; (ii) interface descrita em um formato processável por máquina (especificamente Web Services Description Language - WSDL [5]); (iii) outros sistemas que interagem com um serviço conforme sua descrição utilizando mensagens SOAP (inicialmente um acrônimo para Simple Object Access Protocol) [8]. Essas mensagens são conduzidas utilizando tipicamente o protocolo HyperText Transfer Protocol (HTTP), com uma serialização eXtensible Markup Language (XML) em conjunto com outros padrões Web relacionados.

Um modelo para serviços Web já está consolidado [4]. Nele, provedores publicam seus serviços em repositórios Universal Description Discovery & Integration (UDDI) [6]. Consumidores procuram por serviços em tempo de desenvolvimento ou em tempo de execução, ligando-se a serviços estática ou dinamicamente. Então, os consumidores podem invocar esses serviços. Entretanto, a insuficiência desse modelo para permitir a ampla aceitação da tecnologia é tipicamente indicada na literatura [9]. Atualmente, trabalhos estão sendo desenvolvidos em descoberta e composição de serviços Web, dentre outras áreas. O grande número de serviços e o acoplamento fraco entre provedores e consumidores de serviços indicam a necessidade de aprimorar a descoberta de serviços Web. Especificamente, é necessário incrementar a descrição de serviços para que, além das características funcionais atualmente utilizadas nas buscas, características não-funcionais também possam ser consideradas [7].

A consideração de características não-funcionais em serviços Web proporciona diversos benefícios. As descrições de características não-funcionais publicadas possibilitam a composição de serviços Web baseada em critérios de qualidade. Elas também tornam possível a avaliação de caminhos de execução alternativos para a adaptação de processos. Além disso, atributos de qualidade podem ser utilizados como base para modelos de custo empregados na otimização de processos. O monitoramento de atributos de qualidade ajuda a alcançar os níveis de qualidade esperados por consumidores. Ele também permite que provedores detectem problemas, acomodem mudanças e balanceiem relações de custo/desempenho [11]. Exemplos de características não-funcionais incluem: disponibilidade (a porcentagem de tempo em que um serviço está operante) e segurança (define se um serviço oferece mecanismos de confidencialidade, integridade, autenticação e autorização).

O padrão Web Services Policy Framework (WS-Policy) [2], um *framework* para o processamento de políticas para serviços Web, pode ser utilizado para incrementar a descrição de serviços. WS-Policy compreende um modelo para definir propriedades de serviços Web e uma sintaxe XML correspondente para expressar essas definições como políticas. Referências a políticas podem ser incluídas em documentos XML arbitrários. Políticas podem ser aplicadas em diferentes níveis. Em WS-Policy, uma política é uma coleção de alternativas. Cada alternativa de política é uma coleção de asserções. Uma asserção de política é definida como uma propriedade individual de um comportamento [2]. Tipicamente, asserções de política especificam características que são críticas para a seleção e utilização apropriadas de serviços Web, por exemplo, características não-funcionais.

É possível especificar políticas para provedores e consumidores utilizando WS-Policy. Entretanto, a especificação de políticas é baseada em aspectos sintáticos de propriedades de serviços. Portanto, não é possível representar o significado de asserções de políticas, o que restringe a verificação de compatibilidade entre políticas. Padrões da Web Semântica [3] provêm tal representação. Um exemplo de padrão largamente aceito é o Web Ontology Language (OWL) [10]. Ele facilita a representação de conteúdo interpretável por máquina na Web e pode ser utilizado para representar o significado de termos em vocabulários e seus relacionamentos. Essa representação é chamada ontologia [10]. A utilização de ontologias para apoiar o compartilhamento de informação e a interoperabilidade semântica vem sendo pesquisada há algum tempo. Recentemente, alguns métodos para raciocínio automático foram desenvolvidos baseados no padrão OWL. Eles podem ser aplicados para processar conteúdo em OWL. O conhecimento de domínio oferecido por ontologias pode apoiar a análise de documentos com o objetivo de compreender o conteúdo.

Para aprimorar a descoberta de serviços, este trabalho propõe estender o *framework* WS-Policy para complementar descrições de serviços com políticas para características não-funcionais enriquecidas com semântica. Uma abordagem baseada em ontologias definidas utilizando a linguagem OWL é empregada. O trabalho envolve uma revisão do estado da arte relativo a serviços Web semânticos. Além disso, mecanismos são desenvolvidos para a especificação de políticas utilizando conceitos ontológicos e para a descoberta de serviços considerando o conhecimento de domínio de políticas. O trabalho é voltado principalmente para o apoio de interações com serviços em atividades de negócio, possibilitando a adoção combinada da tecnologia de serviços Web básica e a extensão proposta. As contribuições esperadas do trabalho estão relacionadas com a descrição de aspectos semânticos de políticas para serviços Web e com a verificação de compatibilidade entre políticas semânticas. Além disso, outro diferencial do trabalho é a preocupação com a especificação de características não-funcionais de serviços Web, um fator determinante para o estabelecimento de parcerias entre provedores e consumidores de serviços Web em ambientes caracterizados pelo fraco acoplamento entre os participantes.

Referências

- [1] G. Alonso, F. Casati, H. Kuno e V. Machiraju. *Web Services: Concepts, Architectures and Applications*. Springer-Verlag, 2004.
- [2] S. Bajaj et al. Web Services Policy Framework (WS-Policy) Version 1.2 Specification. BEA, IBM, Microsoft, SAP, Sonic e VeriSign, 01-03-2006. <http://download.boulder.ibm.com/ibmdl/pub/software/dw/specs/ws-polfram/ws-policy-2006-03-01.pdf>, acessado em 08/2006.
- [3] T. Berners-Lee, J. Hendler e O. Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, 2001.
- [4] D. Booth, H. Haas, F. McCabe, M. Champion, C. Ferris e D. Orchard. Web Services Architecture. W3C, 11-02-2004. <http://www.w3.org/TR/2004/NOTE-ws-arch-20040211/>, acessado em 08/2006.
- [5] R. Chinnici, J.-J. Moreau, A. Ryman e S. Weerawarana. Web Services Description Language Part 1 Core Language Version 2.0 Specification. W3C, 10-05-2005. <http://www.w3.org/TR/2005/WD-wsdl20-20050510/>, acessado em 08/2006.
- [6] L. Clement, A. Hatley e T. Rogers. Universal Description Discovery & Integration Version 3.0.2 Specification. OASIS, 19-10-2004. <http://uddi.org/pubs/uddi-v3.0.2-20041019.htm>, acessado em 08/2006.
- [7] K.-C. Lee, J.-H. Jeon, W.-S. Lee, S.-H. Jeong e S.-W. Park. QoS for Web Services. W3C, 25-11-2003. <http://www.w3c.or.kr/kroffice/TR/2003/NOTEwsqos20031125/>, acessado em 08/2006.
- [8] N. Mitra. SOAP Part 0 Primer Version 1.2 Specification. W3C, 24-06-2003. <http://www.w3.org/TR/2003/REC-soap12-part0-20030624/>, acessado em 08/2006.
- [9] M. P. Papazoglou. Extending the Service-Oriented Architecture. *Business Integration Journal*, 40(2):18–21, 2005.
- [10] P. F. Patel-Schneider, P. Hayes e I. Horrocks. OWL Web Ontology Language Semantics and Abstract Syntax. W3C, 02-2004. <http://w3.org/TR/owl-semantics/>, acessado em 08/2006.
- [11] V. Tasic, A. van Moorsel e R. Wong. Quality of Service Middleware for Web Services: Achieved Results and Challenges for the Future - Panel. Em *Anais EDOC Middleware for Web Services Workshop*, páginas 65–67, 2005.

21. CoGPlat: Uma Plataforma para e-Government Orientada a Serviços na Web Semântica

Autores: Ivo José Garcia dos Santos e Edmundo Roberto Mauro Madeira

Introdução

As Tecnologias de Informação e Comunicação (TICs) têm sido aplicadas de maneira cada vez mais intensa por entidades governamentais através do mundo com o objetivo de melhor servir à sociedade [3]. Este fenômeno, conhecido por *Governo Eletrônico* (ou *e-Government*), que inicialmente se limitava ao uso da informática para agilizar os processos burocráticos internos dos órgãos públicos, tem caminhado rumo a uma grande transformação. Os primeiros passos nesta direção foram dados com o início do oferecimento de serviços públicos tradicionais também por canais eletrônicos - no Brasil temos como exemplos de sucesso a *Declaração de Imposto de Renda via Internet*, a *Delegacia Eletrônica do Estado de São Paulo* e o projeto *e-Poupatempo* [6]. Mas isto representa somente o início da mudança: o preceito que estabelece que uma democracia requer cidadãos bem informados [1], de forma a aumentar a credibilidade dos administradores públicos e legitimar suas decisões, exige mecanismos que permitam a participação ativa e autêntica dos cidadãos nos mais diversos processos governamentais [7]. De forma a aumentar a eficiência, transparência e legitimidade destes processos, novos serviços e aplicações (em geral complexos e multi-organizacionais) precisam ser fornecidos. Neste contexto, questões como a heterogeneidade dos sistemas já existentes, a autonomia das diferentes entidades que participam nos processos, aspectos legais, e a privacidade dos dados sendo tratados, entre outras, criam uma série de novos desafios do ponto de vista computacional.

Contribuições

A primeira contribuição deste trabalho é a proposição de um modelo de plataforma (*CoGPlat :: Collaborative e-Government Platform* [5]) que oferece suporte para serviços e aplicações de Governo Eletrônico (*e-Government*) inseridos no contexto da Web Semântica. O principal objetivo da plataforma é facilitar a construção e a operação de aplicações que habilitem a interação e colaboração entre entidades governamentais, organizações e cidadãos em diferentes cenários da administração pública. Uma segunda importante contribuição do trabalho está relacionada com as técnicas utilizadas para implementar Composições de Serviços. O processo proposto inclui um mecanismo de seleção dinâmica de serviços considerando suas descrições semânticas e a mediação das interações entre os serviços de acordo com diversos tipos de políticas que tratam aspectos como privacidade, autonomia e anonimato.

Arquitetura

A plataforma *CoGPlat* foi idealizada como um *middleware* orientado a serviços e sua infra-estrutura (ilustrada na Figura 9) é composta por uma camada de *interface* que oferece serviços para as aplicações clientes da plataforma, pelo *núcleo* do middleware propriamente dito e também por uma *camada de descoberta e execução de serviços*.

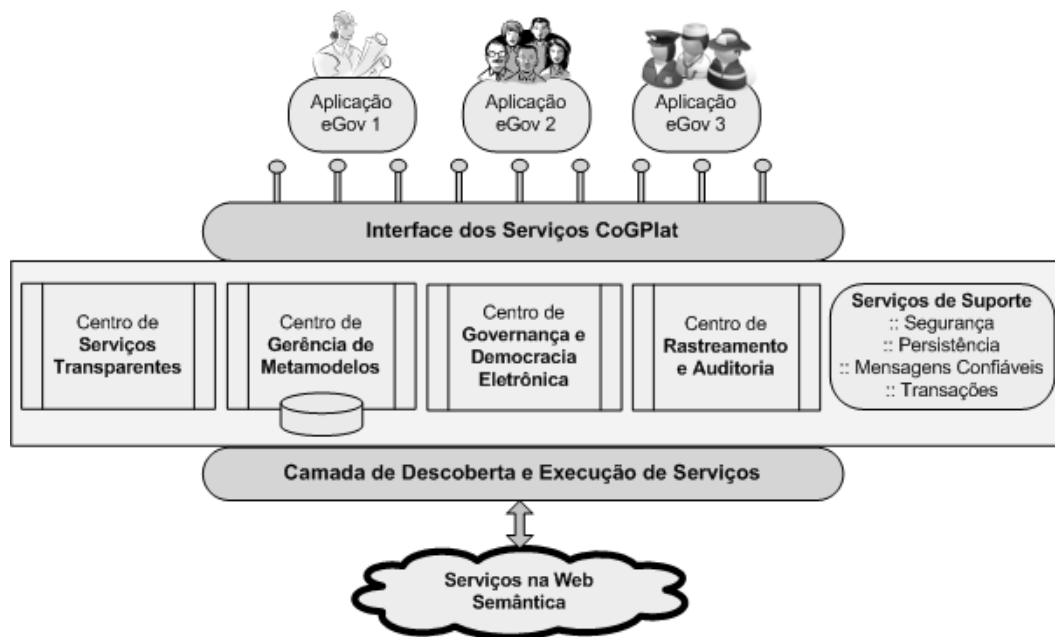


Figura 9: CoGPlat :: Infra-estrutura

O núcleo da plataforma é composto pelos seguintes elementos:

- **Centro de Serviços Transparentes:** responsável por preparar e executar a composição de serviços solicitados pelas aplicações de maneira transparente e dinâmica;
- **Centro de Gerência de Metamodelos:** oferece mecanismos para gerenciar os modelos, metamodelos e ontologias utilizados na descrição dos serviços, composições (abstratas e concretas), processos e entidades que participam da plataforma;
- **Centro de Governança e Democracia Eletrônica:** oferece serviços que têm como principais objetivos facilitar a tomada de decisão por parte dos administradores públicos e aumentar o grau de participação dos cidadãos nos processos públicos;
- **Centro de Rastreamento e Auditoria:** oferece a possibilidade de monitorar o andamento dos processos em execução na plataforma, além de permitir a auditoria em processos já concluídos;

- **Serviços de Suporte:** oferecem mecanismos de Segurança, Persistência, Mensagens Confiáveis e Transações aos processos que rodam sobre o a plataforma.

Implementação

O protótipo da plataforma *CoGPlat* está sendo implementado com o apoio das seguintes tecnologias:

- **Interface de Serviços:** *Web Services* tradicionais com descrição em WSDL e comunicação em SOAP/HTTP;
- **Núcleo:** baseado em Microsoft .Net (C#);
- **Descoberta de Serviços:** utiliza o OWLS-MX [2], um *matchmaker* baseado em JAVA e Jena, que realiza o casamento tanto semântico (OWL-S Profiles) quanto sintático entre requisições e serviços;
- **Execução e Validação das Composições:** o motor de orquestração é baseado na tecnologia *Windows Workflow Foundation* (WF). Existe a possibilidade futura de oferecer suporte a execução de processos WS-BPEL e a validação em tempo real de coreografias WS-CDL.

Aplicações

As facilidades oferecidas pela plataforma *CoGPlat* permitem a proposição de diversas aplicações em diferentes cenários da administração pública. Com o objetivo de demonstrar o protótipo do *middleware* as seguintes aplicações estão sendo implementadas sobre *CoGPlat*:

- **Construção Civil:** automatização do processo burocrático completo de construção de edificações, desde o pedido inicial de aprovação da planta até entrega do habite-se;
- **Gerenciamento de Emergências:** visa facilitar a coleta e o gerenciamento de informações fundamentais para o trabalho de times de gerenciamento de emergências, os quais são normalmente compostos por diferentes órgãos públicos e privados (bombeiros, polícia, defesa civil, cruz vermelha, forças armadas, empresas de energia e saneamento etc);
- **Ouvidoria:** visa implantar um canal interativo para reclamações e sugestões que esteja integrado com todos os órgãos da administração pública, gerenciando de maneira inteligente e transparente todo o processo, desde o instante da requisição inicial do cidadão, passando pelo acompanhamento das soluções sendo tomadas e terminando com uma posterior análise da satisfação do solicitante com a resposta/solução dada pelo gestor público.

Considerações Finais

O trabalho encontra-se atualmente em sua última etapa, estando a defesa prevista para o primeiro trimestre de 2008. Trabalhos futuros incluem a proposição e implementação de um conjunto maior de aplicações que explorem de diferentes maneiras as funcionalidades da plataforma e permitam a realização de testes de desempenho do *middleware* levando em conta diferentes configurações de rede e de carga (variando parâmetros como o número de entidades participantes, a quantidade de serviços disponíveis e o tipo de distribuição das facilidades). Uma interessante extensão a este trabalho seria a implementação da infra-estrutura sobre uma grade computacional orientada a serviços [4].

Agradecimentos

Este projeto de doutorado teve o apoio financeiro do CNPq (bolsa de doutorado) e da CAPES/DAAD (bolsa de doutorado-sanduiche). Gostaríamos de agradecer também ao Instituto Fraunhofer-FOKUS (Berlim, Alemanha) e à Microsoft Research.

Referências

- [1] Thomas Jefferson. *Personal communication to R. Price*, 1789.
- [2] M. Klusch, B. Fries, and K. Sycara. Automated semantic web service discovery with owls-mx. In *Proc. of the 5th Intl. joint conference on Autonomous agents and multiagent systems (AAMAS)*, pages 915–922, USA, 2006. ACM.
- [3] Gary Marchionini, Hanan Samet, and Larry Brandt. Digital government. *Communications of the ACM*, 46(1):25–27, Jan 2003.
- [4] Ivo J. Garcia Santos and Edmundo R. M. Madeira. E-government and grid computing: Potentials and challenges towards citizen-centric services. In *Proc. of the 9th Intl. Conf. on Enterprise Information Systems (ICEIS)*, pages 144–148, Portugal, 2007.
- [5] Ivo J. Garcia Santos, Edmundo R. M. Madeira, and Volker Tschammer. Towards dynamic composition of e-government services - a policy-based approach. In *Proc. of the 5th IFIP Intl. Conf. on e-Commerce, e-Business and e-Government (I3E)*, volume 189 of *IFIP*, pages 173–185. Springer, Out 2005.
- [6] Vera Tokairim and Agnaldo do Carmo Lopes. E-poupatempo: ampliando os limites da prestação de serviços públicos. In *VIII Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública*, 2003.
- [7] Richard T. Watson and Bryan Mundy. A strategic perspective of electronic democracy. *Communications of the ACM*, 44(1):27–30, Jan 2001.

22. Uma Arquitetura para gerência de mobilidade em redes de acesso sem fio heterogêneas

Autores: Jorge Lima de Oliveira Filho, Edmundo R. M. Madeira

Atualmente existe uma grande quantidade de dispositivos aptos a acessar redes sem fio, através de uma grande diversidade de tecnologias de acesso. Cada tecnologia tem suas vantagens e desvantagens de acordo com o cenário utilizado, por exemplo, o uso da rede Wi-Fi (*Wireless Fidelity*) é ideal para ambientes internos e inadequado para o uso externo. Já a tecnologia WiMAX (*Worldwide Interoperability for Microwave Access*) é adequada para ambientes externos com alcance relativamente longo. A tendência mundial é que todas as redes sem fio sejam baseadas no protocolo IP e que estas novas redes suportem múltiplas tecnologias de acesso. As redes com estas características também são conhecidas como sistemas sem fio de próxima geração (*Next-generation wireless system - NGWS*) [1].

Este trabalho propõe a criação de uma arquitetura para redes multi-acesso sem fio que considera a necessidade das pessoas se deslocarem, enquanto fazem uso de suas aplicações, sem perder a conexão ao longo do percurso (continuidade de sessão), ainda que alternem entre diferentes redes de acesso. A arquitetura utiliza dentre outros critérios, os parâmetros de QoS para decidir efetuar uma troca de rede. Se por um lado as redes móveis atuais têm limitações de cobertura, por outro, as diversas tecnologias de acesso não se integram. A arquitetura proposta tem como objetivo resolver esses aspectos criando uma infra-estrutura comum aos usuários fazendo uso de várias tecnologias de acesso de forma transparente.

Vários trabalhos propostos pela comunidade científica tratam da gerência de *handoff* nas redes de acesso sem fio heterogêneas. Muitos trabalhos relacionados com esta proposta são modificações do IP móvel e são implementados na camada de rede. Alguns trabalhos também se baseiam no protocolo SIP (*Session Initiation Protocol*) [5] e são implementados na camada de aplicação. Em [4] é proposto um conjunto de protocolos e mecanismos que proporcionam mobilidade na Internet, conhecido como IP móvel. O IP móvel permite roteamento de datagramas de forma transparente para nós móveis na Internet. Cada nó é conhecido por um identificador (*Home Address*) independentemente da rede que ele estiver conectado na Internet. Enquanto estiver situado fora de sua rede nativa (*Home Network*), um nó é também associado a um "endereço aos cuidados" (*care-of-address - CoA*). Ao receber datagramas destinados ao nó, o agente nativo reenvia-os através de um túnel para o CoA. Ao fim do túnel, cada datagrama é então entregue ao nó móvel. O protocolo HOPOVER [2] (*HandOff Protocol for OVERlay networks*), foi projetado para redes sobrepostas (*Overlay networks*). Pode realizar tanto o *handoff* horizontal quanto o vertical. Este protocolo tem a desvantagem da necessidade de utilização de estações bases para manter um grande número de informações sobre os nós móveis.

Em [3] é proposto um protocolo de gerência de *handoff* implementado nas camadas de rede e de enlace (*cross-layer*). Neste protocolo é suportado tanto a gerência

de *handoff* inter-sistema quanto a intra-sistema. O protocolo usa informações de velocidade do dispositivo móvel e a qualidade do sinal do rádio como critério de decisão de *handoff*. Informações específicas da camada de enlace são passadas para a camada de rede, melhorando o desempenho do *handoff*.

A gerência de mobilidade é uma tarefa muito complexa, mais especificamente a gerência de *handoff* inter-sistema, onde a heterogeneidade das redes envolvidas é muito grande. Novas técnicas no desenvolvimento de protocolos de *handoff* são necessárias para diminuir o atraso na troca de rede. A maioria das arquiteturas propostas consistem em implementações na camada de rede e é baseada no IP móvel e, devido a uma alta troca de mensagens entre seus componentes, impossibilita que o *handoff* seja eficiente. Para lidar com esses problemas, este trabalho propõe investigar a melhor forma de executar um *handoff* entre redes sem fio heterogêneas utilizando dentre outros critérios de decisão, os parâmetros de QoS através de uma abordagem *cross-layer* entre as camadas de enlace e de rede no projeto de sistemas sem fio de próxima geração (*Next-generation wireless system* - NGWS).

O objetivo precípua desta proposta é a criação de uma arquitetura para redes multi-acesso sem fio, utilizando as tecnologias Wi-Fi e WiMax, com suporte à continuidade de sessão de forma transparente para o usuário.

Referências

- [1] I.F. Akyildiz, X. Jiang, and S. Mohanty. A survey of mobility management in next-generation all-ip-based wireless systems. *IEEE Wireless Communications*, 11(4):16 – 28, August 2004.
- [2] Fan Du, Ni L.M, and A.H. Esfahanian. Hopover: a new handoff protocol for overlay networks. *IEEE International Conference on Communications*, 5:3234–3239, 2002.
- [3] S. Mohanty and I.F. Akyildiz. A cross-layer (layer 2 + 3) handoff management protocol for next-generation wireless systems. *IEEE Transactions on Mobile Computing*, 5(10):1347–1360, 2006.
- [4] C. Perkins. Ip mobility support for ipv4. RFC 3220, 2002.
- [5] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. Sip: Session initiation protocol. RFC 3261, 2002.

23. Métodos Geométricos e Fotométricos para Visão Computacional 3d

Autores: Rafael Felipe Veiga Saracchini, e Jorge Stolfi

A captura de modelos tridimensionais de objetos reais é um problema computacional com inúmeras aplicações. Podemos citar a documentação e restauração de monumentos, a engenharia reversa de peças ou prédios, a operação autônoma de robôs, a fabricação de próteses, o monitoramento de animais ou ambientes, etc..

Na última década, foram desenvolvidos vários tipos de *scanners tridimensionais*, aparelhos que capturam a geometria 3D de um objeto diretamente — por exemplo, medindo o tempo de trânsito de pulsos *laser* dirigidos ao objeto. Entretanto, tais equipamentos ainda são caros e volumosos, e seu campo de aplicação ainda é relativamente restrito. Assim, ainda há interesse em métodos que permitem obter a terceira dimensão a partir de fotos digitais comuns, dispensando equipamentos especializados.

As técnicas que podem ser usadas para esse fim podem ser divididas em duas classes principais: *estéreo fotométrico* e *estéreo geométrico*.

Métodos estéreo-fotométricos utilizam-se de três ou mais fotos do objeto, todas tiradas do mesmo ângulo e distância, mas com iluminações diferentes. Em princípio, os valores de pixels correspondentes nessas fotos permitem determinar a inclinação desse ponto da superfície. Este *mapa de inclinações* pode ser então ser integrado para se obter a profundidade de cada pixel.

Trabalhos como os de [2], [5] e [6] assumem o modelo Lambertiano para recuperar as informações relativas à inclinação da superfície - as *normais*. Segundo [1] são necessárias três ou mais imagens para eliminar a ambiguidade de algumas normais recuperadas por este método, adicionado o problema causado por sombras projetadas e *highlights*, que não são cobertos pelo modelo acima. Diversas abordagens como [3] tentam solucionar este problema utilizando uma grande quantidade de imagens para assim obter informações suficientes. Tal abordagem porém requer grande poder computacional dado o fato do problema de se calcular estas normais se tornar uma busca em um espaço N-Dimensional. Neste projeto foi desenvolvido um método em [4] extremamente eficiente na redução nesta busca, embora o mesmo ainda possa ser ainda mais aprimorado.

Métodos estereo-geométricos utilizam-se duas ou mais fotos do objeto, todas tiradas em condições semelhantes de iluminação, mas de posições diferentes. O deslocamento da câmera entre uma foto e outra dá origem a deslocamentos aparentes (*paralaxes*) das várias partes da cena na imagem, que variam conforme a profundidade dessas partes. Assim, é possível recuperar a terceira dimensão da cena identificando os pontos das imagens dadas que correspondem ao mesmo ponto na cena, e analisando os deslocamentos relativos desses pontos.

As duas abordagens, fotométrica e geométrica, foram extensamente estudadas nas últimas décadas. Há entretanto relativamente poucos trabalhos que tentam

combinar as duas abordagens. Possivelmente, isso tem a ver com a natureza das informações obtidas. O estéreo geométrico produz diretamente a profundidade de cada ponto, enquanto que o estéreo fotométrico essencialmente produz as derivadas dessa informação. Conciliando estes dois métodos pode-se obter informações que normalmente seriam perdidas utilizando-se apenas um método ou outro.

Referências

- [1] Robert J. Woodham. Determining Surface Curvature with Photometric Stereo. *Proceedings of the 1989 IEEE International Conference on Robotics and Automation*, pages 36–42, 1989.
- [2] Osamu Ikeda. Shape Reconstruction for Color Objects using Segmentation and Photometric Stereo. *Proceedings of ICIP 2004*, pages 1365–1368, 2004.
- [3] Svetlana Barsky and Maria Petrou. The 4-Source Photometric Stereo Technique for Three-Dimensional Surfaces in Presence of Highlights and Shadows. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pages 1239–1252, 2003.
- [4] Helena C. Gama Leitão, Jorge Stolfi and Rafael Saracchini. A Bucket Grid Structure to Speed Up Table Lookup in Gauge-Based Photometric Stereo. *XX Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2007)*, pages 221–230, 2007.
- [5] Aaron Hertzmann and Steven M. Seitz. Shape and Materials by Example: A Photometric Stereo Approach. *Proceedings IEEE CVPR 2003*, pages 533–540, 2003.
- [6] Aaron Hertzmann and Steven M. Seitz. Example-Based Photometric Stereo: Shape Reconstruction with General, Varying BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1254–1264, 2005.

24. Geração Automática de Storyboards para Vídeos de Futebol Televisivos

Autores: Nielsen Cassiano Simões, Neucimar J. Leite

A análise automática de vídeo é um importante passo para o entendimento semântico de seu conteúdo. Em especial, vídeos de esportes, por possuírem características específicas, permitem a determinação de um subconjunto com os principais eventos que podem ser utilizados de diferentes formas, tais como em programas esportivos, noticiários, televisão digital ou telefonia móvel. Uma boa alternativa para a navegação em vídeos é a utilização de storyboards, que constituem quadros representativos do vídeo de forma a fornecer um resumo visual e estático dos principais acontecimentos. Essa alternativa é de grande interesse para a navegação do vídeo [6], permitindo ao usuário identificar a região de interesse, podendo ser utilizado inclusive para recuperação por conteúdo de vídeos [2].

Um segmento de vídeo digital é composto por uma seqüência de quadros. Um subconjunto ininterrupto de quadros, obtidos de uma mesma câmera, é denominado tomada. Em geral, o processo de análise de vídeo está associado à detecção de transições entre tomadas [1], decorrentes da edição do vídeo. Pode-se também definir alguns quadros para representar cada tomada, denominados quadros-chave. São esses quadros a base para a geração de storyboards. A detecção de tomadas representa o primeiro e principal passo para a análise semântica automática de um vídeo digital.

Este trabalho fornece um modelo para geração automática de storyboards para vídeos de futebol televisivos que, diferentes de outros vídeos, possuem uma cinemática (dinâmica de câmeras entre tomadas) associada aos eventos esportivos. No caso do futebol, os principais eventos são: gol, cobrança de falta com barreira, escanteios, pênaltis, aplicações de cartões e substituições. A geração de storyboards está, primeiramente, associada a detecção de tomadas e, posteriormente, a seleção de quadros chaves.

A detecção de tomadas é baseada no uso de uma imagem representativa do vídeo, denominada Ritmo Visual [3, 4, 7], constituindo uma boa alternativa em relação aos métodos tradicionais por métricas de dissimilaridade entre quadros [1]. O Ritmo Visual utilizado foi apresentado em [5], o qual baseia-se na utilização da Moda (pixel mais representativo) por linha, explorando a característica da presença do gramado em vídeos de futebol.

Após a detecção de tomadas, a seleção de tomadas foi realizada para eliminar tomadas de falhas de transmissão, comuns em transmissões televisivas. Considerando as possíveis tomadas de câmera utilizadas em jogos esportivos, relativas à abertura da câmera (zoom, visão média ou visão longa), uma pré-classificação é realizada para a identificação dos eventos conforme a cinemática do padrão televisivo. A imagem de Ritmo Visual da Moda também foi utilizada para a determinação da seleção de tomadas e a pré-classificação.

Utilizando 3 níveis semânticos, a geração de storyboard (Figura 10) oferece alternativas para a navegação do vídeo. A Figura 10 a) apresenta o mais baixo nível semântico, associado apenas às transições de tomadas, semelhante aos storyboards automáticos geralmente utilizados. Assumindo que o usuário prefira uma imagem mais explicativa, com maiores detalhes do vídeo, os quadros cuja imagem oferece uma visão média do gramado (um jogador é visto, por exemplo de corpo inteiro na imagem), um segundo nível semântico é oferecido para storyboard, conforme Figura 10 b). Considerando a cinemática do jogo, oferecida pelos padrões de transmissão, a Figura 10 c) ilustra o terceiro nível semântico para o storyboard automático, associado à ocorrência dos principais eventos do jogo.

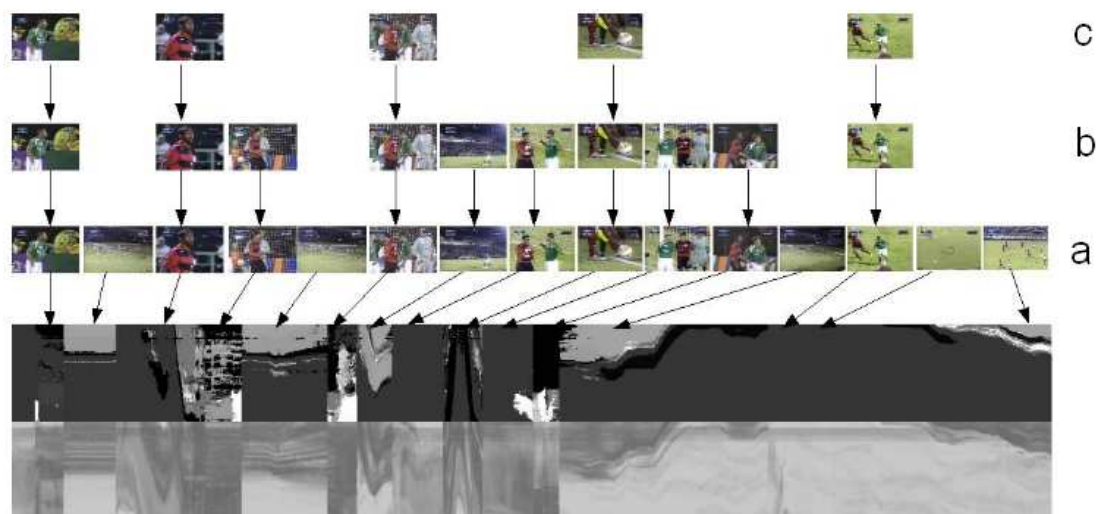


Figura 10: Níveis semânticos de Storyboard para um segmento de vídeo de futebol: a) baixo nível, b) médio e c) alto.

Como resultado, os 3 níveis de storyboard oferecidos fornecem boas alternativas para a navegação de vídeos de futebol televisivos, conforme o nível semântico desejado pelo usuário.

Referências

- [1] R. Brunelli, O. Mich, and C. M. Modena. A survey on the automatic indexing of video data. *Journal of Visual Communication and Image Representation*, 10(2):78–112+, 1999.
- [2] Michael G. Christel and Rong Yan. Merging storyboard strategies and automatic retrieval for improving interactive video search. In Nicu Sebe and Marcel Worring, editors, *CIVR*, pages 486–493. ACM, 2007.
- [3] Silvio Jamil Ferzoli Guimarães, Michel Couprie, Neucimar Jerônimo Leite, and Arnaldo de Albuquerque Araújo. A method for cut detection based on visual

rhythm. In *SIBGRAPI XIV Brazilian Symposium on Computer Graphics and Image Processing*, pages 297–304, 2001.

- [4] C. W. Ngo, T. C. Pong, and R. T. Chin. Survey of video parsing and image indexing techniques in compressed domain. *Symposium on Image, Speech, Signal Processing, and Robotics (Workshop on Computer Vision)*, 1:231–236, 1998.
- [5] Nielsen Cassiano Simões, Neucimar Jerônimo Leite, and Beatriz Marcotegui. Sumarização automática de eventos de vídeo de futebol televisivos. In *C. M. B. Medeiros, A. M. Melo, C. G. N. Macario, F. Castor Filho, N. C. Malheiros, and V. P. A. Neris (Eds.), Anais do 2o workshop de teses de doutorado em andamento do IC-UNICAMP*. Technical Report IC-06-24, Instituto de Computação – UNICAMP, 2006.
- [6] Hari Sundaram and Shih-Fu Chang. *Video Analysis and Summarization at Structural and Semantic Levels*, chapter 4, pages 75–94. Springer, 2003.
- [7] HongJiang Zhang, Atreyi Kankanhalli, and Stephen W. Smoliar. Automatic partitioning of full-motion video. *ACM Multimedia Systems*, 1(1):10–28, 1993.

25. An Approach for Development of Reactive Systems with Distributed Behavior

Authors: Carlos Neves Júnior and Hans Kurt Edmund Liesenberg

A great number of studies for modeling behavior of computer systems were done over the last decades. These studies raised on one more dichotomy for computer systems: transformational versus reactive systems. A transformational system has the following lifecycle: accepts user inputs, performs computation (transformation) on them and produces outputs as result. A reactive system, in general, does not compute a function and it continuously reacts to external inputs that change its internal state and behavior. Event-driven systems, software controllers of any kind, user interface dialog managers are some examples of reactive systems. If reactive systems with complex behavior are difficult to specify, distributed reactive systems are even more [6]. The modeling of complex behavior in presence of distributed requirements is not well addressed by great part of the available methods for specifying reactive systems.

One of the most relevant studies on the reactive systems area is *Statecharts* [1], an approach that uses a visual formalism for specification and design of complex behavior. Statecharts is also used on the scope of object oriented design where a subset of it is part of the *Unified Modeling Language* (UML) [2] notation. UML Statecharts, as it is known, is used to define stateful behavior for objects. There are some other studies that target the scope of the distributed reactive systems [5]. The current Doctorate effort is particularly interested on this problem, the modeling and execution of reactive systems with distributed behavior.

The Xchart technology [3], is one proposal leveraged by the Institute of Computing-UNICAMP that originally was created to reduce the difficulties encountered on the development of complex user interface dialogue managers, in contrast to other user interface researches that mainly focus the development cycle of the presentation layer. Xchart is one variant of *Statecharts* and can also be positioned in a wider domain of systems: the niche of distributed reactive systems.

Xchart proposes a visual language and one support environment for the specification and execution of complex reactive systems. The specification of reactive behavior using the Xchart language allows the design of the concurrent execution of control elements, therefore, their potential distribution over the network. The Xchart support environment is composed of tools for edition and execution of specifications described in the Xchart language. The available Xchart tools are developed on top of the Java technology and part of the development environment is packaged as an Eclipse plug-in. As basis for the current Doctorate cycle, there were good outcomes during the Master piece of this work [4]: the implementation of a prototype for the Xchart execution environment. The Xchart execution environment is the middleware that enable the execution of Xchart specifications also called instances. Instances can be potentially executed in a distributed environment and so

it is responsibility of the Xchart execution environment to provide distribution services for them. These distribution services were currently implemented using JEE technology. Another recent work in the scope of the Xchart support environment [8] implemented one editor for design of Xchart specifications and will also be the basis for some of the present research.

The aim of the current work is to extend the current Xchart technology in order to offer a real alternative for the specification and execution of distributed reactive systems. This work will consist on a complete revision of the requirements for the development of one complete environment for design, simulation, tests and deployment of distributed reactive systems. These extensions on the previous works will also address the language formalism since new elements may be included for better support of the distributed requirements. There are some proposals that target the distribution of reactive systems, but they count on the developer to generate code in order to fulfill distribution needs of the application [5] [7]. The idea is to differentiate from these available works that are based on *Statecharts*, extending the Xchart formalism that is already more adequate for distributed development in comparison with *Statecharts*. Xchart already has mechanisms for concurrency control, shared memory (global variables) and also services for message communication among distributed instances (broadcast, multicast and peer-to-peer) that might be extended.

Transparency related with the distribution of the elements is a very important requirement to meet in the proposed development environment, once it avoids the developer to be overloaded with some low level activities. Other requirements like security and fault-tolerance will also be evaluated. The algorithms and the technologies used in the previous experiments will also be revised, so new and more efficient solutions might be used. The decision related with what technologies will be used are still subject of research, but there are some trends that will be certainly leveraged like *SOA (Service Oriented Architecture)*. SOA can allow a better interoperability and integration of the reactive system components with other application layers of the solution. More detailed information about the Xchart project can be found at the official project web site [9].

Referências

- [1] David Harel. *Statecharts: A Visual Formalism for Complex Systems*. *Science of Computer Programming*, 8(3): 231-274, June 1987.
- [2] Booch, Rumbaugh, and Jacobson. *The Unified Modeling Language User Guide*. *Addison-Wesley*, 1998.
- [3] Fábio Nogueira de Lucena e Hans Kurt Edmund Liesenberg. *Um Modelo de Especificação e Implementação de Gerenciadores de Diálogo*. *Ph.D's Thesis, IC/UNICAMP, Campinas/SP, December 1997*.

- [4] Carlos Neves Júnior e Hans Kurt Edmund Liesenberg. Gerente de Distribuição do Ambiente Xchart em J2EE: Distribuindo Sistemas Reativos com a Tecnologia Xchart. *Master's dissertation, IC/UNICAMP, Campinas/SP, September 2005.*
- [5] Hassan Gomaa. Designing Concurrent, Distributed Real-Time Applications with UML. *Addison-Wesley, 2000.*
- [6] Pnueli, Amir e Rosner, Roni. Distributed Reactive Systems are Hard to Synthesize. *Proceedings of 31st IEEE Symp. on Foundations of Computer Science. 1990. pp. 746-757.*
- [7] Mark Mahoney and Tzilla Elrad. Distributing Statecharts to Handle Pervasive Crosscutting Concerns. *Building Software for Pervasive Computing Workshop at OOPSLA, October 2005.*
- [8] Diogo Kollross e Hans Kurt Edmund Liesenberg. Sins: Um Editor Xchart na forma de Plugin para o Ambiente Eclipse. *Master's dissertation, IC/UNICAMP, Campinas/SP, October 2007.*
- [9] Xchart Project web site. <http://xchart.incubadora.fapesp.br> *November, 2007.*

26. Dynamic Task Scheduling on Computational Grids

Authors: Luiz Fernando Bittencourt and Edmundo R. M. Madeira

Grid Computing [6] emerged in the middle of 1990's as being the use of a non-dedicated distributed heterogeneous environment to compute jobs. This distributed system, called *computational grid*, is usually composed of multiple organisations with independent administrative domains. Nowadays grids are spreading as a high-performance environment for data storage and parallel processing. With the emergence of e-Science, grids are being used to compute jobs which demand high processing power and jobs which make use of large data sets. Another concept that matches with these requirements is the workflow programming paradigm. This model is useful to represent flows of control and to facilitate the complexity management of processes that have multiple dependent tasks. With this, the workflow paradigm is becoming a standard for management of scientific processes with massive data sets. Thus, the combination of grids and workflows is a promising way of doing science [9].

Within the context of workflow execution on distributed systems, scheduling of tasks is primordial to provide efficiency and to speed up the process results arrival. More difficulties arise when scheduling tasks on the grid, namely heterogeneity and dynamic behaviour. Whilst heterogeneity means complexity in selecting resources to execute tasks, the dynamic behaviour (variations in resources performance and availability) means complexity in performance prediction and rescheduling. Scheduling, performance prediction and rescheduling are receiving substantial attention from the grid researchers nowadays [4, 7, 8]. However, some important points when scheduling tasks on grids, like concurrent execution and scheduling of processes with dependent tasks, are not receiving the deserved attention.

The task scheduling problem consists in, given a set of dependent tasks and its dependencies, choosing in which resource each task will execute. A scheduler, in general, has an objective function. For example, a scheduler can try to minimise the execution time (makespan) of the whole process, maximise the resource utilisation or maximise the overall system throughput. We focus on the former. A workflow is usually represented by a directed acyclic graph (DAG) $G = (V, E)$, where V is the set of tasks to be executed and E is the set of edges that represent data dependencies between tasks. Thus, an edge $e_{i,j}$, with its source on task t_i and its target on task t_j , means that t_j can start its execution only after t_i ends its execution and sends all necessary data to t_j . Dependent task scheduling is an NP-Complete problem [5], hence the associated optimization problem is NP-Hard.

In our work we study the aspects cited above to have better schedules for processes on a grid environment. We mix a scheduling heuristic [3], a dynamic scheduling strategy [1], and a performance oriented adaptive task allocation algorithm [2]. Whilst the developed heuristic tries to minimise the communication costs between

tasks, the dynamic scheduling and the adaptive task allocation deal with performance losses in the grid shared resources. Besides the NP-Completeness of the problem, we have to deal with the heterogeneity and the dynamic behaviour of the resources, since it is a shared environment. In our work we combine some techniques to deal with these characteristics, with the objective of minimizing the makespan of the processes. A four-step scheme representing the scheduling strategy is shown in Figure 11.

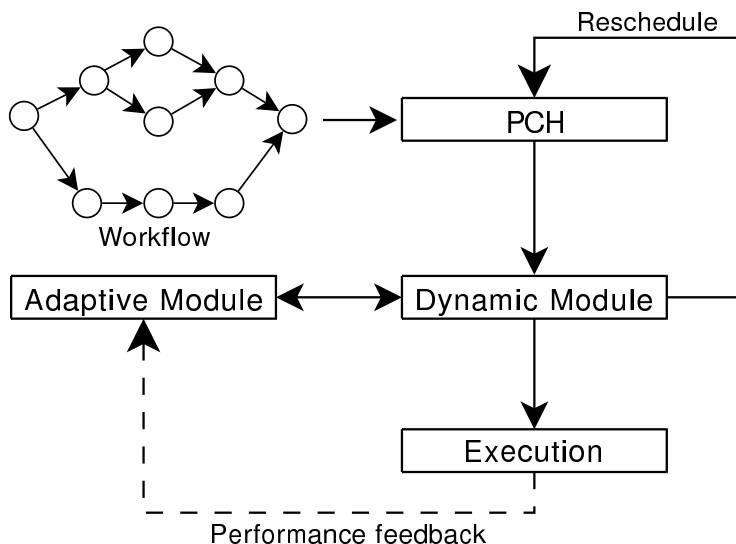


Figure 11: Four-step scheduling scheme.

In the first step, we use the *Path Clustering Heuristic* (PCH) [3] to make the initial schedule of tasks. The *round-based* technique [2] is used in the second step, where the algorithm decides dynamically on each *round* which scheduled tasks will be sent to execution. With this, the algorithm can have performance feedback from the resources to make rescheduling decisions based on the initial schedule given by the PCH algorithm, avoiding to send tasks to execute in resources with poor performance. The dynamic module chooses which tasks will be sent to execution on each round supported by the adaptive module, which regulates the size of each round. This adaptive round size regulation tries to avoid sending too many tasks to execute in resources that had strong performance variations in the past. The round size is measured in terms of execution time of the tasks on a given resource, thus it depends on each resource capacity. The execution is the third step, and the rescheduling is the fourth step.

Simulations have shown that the four-step scheduling scheme with the proposed algorithms can give better schedules by enhancing the resources usage, with the PCH giving a good initial schedule and the dynamic round-based algorithm adapting the schedule to the measured resources performance.

We are now developing an algorithm to fill gaps in the schedule. A schedule of a process composed of dependent tasks have gaps between execution of tasks, since

each task must wait for the data from its predecessors to start its execution. When scheduling more than one process in the same set of resources, the algorithm can consider these gaps to execute tasks of processes that arrive. As a consequence, the complexity increases and there are more combinations of possible schedules, whilst the process execution times can be decreased and the overall machine utilisation maximised. Preliminary results suggest that the gap searching algorithm in conjunction with a simple rescheduling algorithm can provide good schedules in situations inherent to the grid. Whilst the gap searching algorithm, supported by the dynamic and adaptive modules, tries to make good schedules when resources have good performance (with little or no variation in expected performance), the rescheduling deals with heavier performance variations. Also, using queue and gaps to execute tasks, the algorithm can estimate the finish time of the process more precisely than when executing processes at the same time on the same resource. With many processes arriving, the first process can delay in an unpredictable way if it shares resources with the arriving processes. This way, one cannot have a good estimation of the execution time of his/her process when it is submitted, what can be achieved using queues.

Referências

- [1] L. F. Bittencourt and E. R. M. Madeira. A dynamic approach for scheduling dependent tasks on the xavantes grid middleware. In *4th ACM International Workshop on Middleware for Grid Computing*, Melbourne, Australia. nov 2006.
- [2] L. F. Bittencourt and E. R. M. Madeira. A performance oriented adaptive scheduler for dependent tasks on grids. *Concurrency and Computation: Practice and Experience (to appear)*, 2007.
- [3] L. F. Bittencourt, E. R. M. Madeira, F. R. L. Cicerre, and L. E. Buzato. Uma heurística de agrupamento de caminhos para escalonamento de tarefas em grades computacionais. In *XXIV Simpósio Brasileiro de Redes de Computadores*, pages 83–98, may 2006.
- [4] H. Chen and M. Maheswaran. Distributed dynamic scheduling of composite tasks on grid computing systems. In *11th IEEE Heterogeneous Computing Workshop*, pages 119–128, Washington, DC, USA, 2002. IEEE Computer Society.
- [5] H. El-Rewini, H. H. Ali, and T. G. Lewis. Task scheduling in multiprocessing systems. *IEEE Computer*, 28(12):27–37, 1995.
- [6] I. Foster, C. Kesselman, and S. Tuecke. The anatomy of the Grid: Enabling scalable virtual organization. *The International Journal of High Performance Computing Applications*, 15(3):200–222, 2001.

- [7] R. Prodan and T. Fahringer. Dynamic scheduling of scientific workflow applications on the grid: a case study. In *SAC '05: Proceedings of the 2005 ACM symposium on Applied computing*, pages 687–694, New York, NY, USA, 2005. ACM Press.
- [8] X.-H. Sun and M. Wu. Grid harvest service: A system for long-term, application-level task scheduling. In *Proceedings of the 17th International Symposium on Parallel and Distributed Processing*, page 25.1, Washington, DC, USA, 2003. IEEE Computer Society.
- [9] I. J. Taylor, E. Deelman, D. B. Gannon, and M. Shields, editors. *Workflows for e-Science: Scientific Workflows for Grids*. Springer Verlag, 2007.

27. Balanceamento de Carga para Serviços Web Altamente Disponíveis

Autores: Alan Massaru Nakai, Edmundo Madeira, Luiz Eduardo Buzato

Com a proliferação de companhias que realizam porções significativas de seus negócios *online*, cresce o interesse por alta disponibilidade de serviços Web. Um motivo evidente para este crescente interesse é o fator financeiro [1]. Considerando que uma companhia obtém uma receita de \$1 bilhão anuais com suas transações *online*, pode-se estimar, a grosso modo, que 0,1% de indisponibilidade dos seus serviços geraria um prejuízo de \$1 milhão anuais. Na mesma linha de raciocínio, a indisponibilidade dos serviços pode gerar descontentamento de clientes, influenciando negativamente na imagem da companhia e, conseqüentemente, gerando perdas de receita. Além disso, a falta de disponibilidade do sistema em aplicações críticas, como aplicações militares e da medicina, pode ter conseqüências catastróficas, levando risco ao bem estar humano.

Muitas razões podem levar à indisponibilidade de um serviço Web, como: falha ou sobrecarga do servidor, problemas de roteamento e congestionamento de vias de comunicação. Uma solução comum para o aumento da disponibilidade de um serviço Web é a adição de redundância ao sistema, via replicação dos servidores. Esta solução aumenta a probabilidade de que um cliente possa conectar-se a um servidor mesmo na presença de falhas parciais. Além disso, a replicação potencializa a escalabilidade do serviço Web, diminuindo a chance de sobrecargas. O sucesso da replicação depende do êxito obtido na solução de dois sub-problemas [2]: (i) a gerência da consistência dos dados replicados entre os múltiplos servidores [3, 4] e (ii) a distribuição da carga de trabalho entre eles.

Este trabalho concentra-se no problema da distribuição de carga para servidores Web distribuídos. Este problema consiste em prover a ligação dinâmica entre o cliente e um dos nós replicados, visando minimizar a carga máxima de um nó servidor, dentre todos os nós, em um dado instante. A atribuição de um nó servidor a um cliente pode levar em consideração informação dinâmica de ambas as partes. Além disso, técnicas de balanceamento de carga podem ser aplicadas em diversos níveis da arquitetura de *software* do sistema.

De forma geral, as técnicas para balanceamento de carga para servidores Web replicados encontradas na literatura podem ser divididas em duas categorias: técnicas para a servidores Web geograficamente distribuídos (ex. [5–7]) e técnicas para a servidores Web replicados e hospedados em aglomerados (ex. [8–10]). Incluem-se na primeira categoria, as soluções aplicáveis a servidores Web distribuídos nos quais todo nó replicado possui um endereço IP visível aos clientes. A segunda categoria engloba as abordagens aplicáveis aos servidores Web distribuídos cujos nós compõem um aglomerado, apresentando um único endereço IP visível, geralmente atribuído ao roteador de borda do aglomerado.

O objetivo deste trabalho é especificar um mecanismo para o balanceamento de

carga de servidores Web compostos por múltiplos aglomerados distribuídos geograficamente, visando manter alta disponibilidade dos serviços providos. Este mecanismo englobará técnicas e políticas de balanceamento de carga aplicáveis em todos os níveis da arquitetura do serviço Web, incluindo o DNS (*Domain Name System*), dispositivos de redirecionamento e os nós servidores. A metodologia para o desenvolvimento do trabalho inclui quatro grandes tarefas: (i) análise das técnicas existentes e da possibilidade de sua integração, (ii) desenvolvimento de uma arquitetura de software para implantação do mecanismo proposto e (iii) avaliação experimental do mecanismo.

O atual estágio do trabalho é a definição de uma infra-estrutura para realização dos experimentos para avaliação de mecanismos de balanceamento. As abordagens possíveis são: experimentos reais, simulação e emulação. Em experimentos reais, a arquitetura é implantada e usuários reais produzem a carga de trabalho executando aplicações de testes. Embora esta abordagem forneça resultados precisos, demanda elevado custo de recursos computacionais, pois exige a disponibilidade de múltiplos aglomerados espalhados pela Internet. Em simulações, tanto a arquitetura quanto a carga de trabalho são simulados por *software*. A simulação demanda menos recursos computacionais do que experimentos reais, porém, a complexidade da simulação de aplicações distribuídas na Internet pode adicionar erros aos resultados dos experimentos. Já na emulação, a arquitetura é implantada e a carga de trabalho é emulada com base em históricos (*traces*) de aplicações reais, permitindo resultados mais fiéis a realidade do que em simulações. Entretanto, neste trabalho, uma questão importante para os experimentos é a configuração dos servidores DNS utilizados nas resoluções de endereço durante as requisições. Esta questão não é refletida nos resultados dos experimentos via emulação.

Para avaliar o mecanismo de balanceamento proposto, pretende-se utilizar duas métricas: *utilização máxima dos nós servidores* e a *disponibilidade dos serviços Web*. A utilização máxima mede, em um dado instante, a utilização máxima observada de um nó servidor, dentre todos os nós servidores. A disponibilidade dos serviços pode ser estimada por meio da medição do número de requisições atendidas em um limite de tempo considerado aceitável pelo usuário.

As principais contribuições esperadas são:

- Uma análise comparativa das técnicas para balanceamento de carga para servidores Web distribuídos e de possíveis combinações;
- A especificação de um mecanismo para balanceamento de carga voltado para servidores Web compostos por aglomerados geograficamente distribuídos;
- A especificação e a implementação de uma arquitetura de *software* para implantação do mecanismo proposto;
- A validação experimental da solução.

Referências

- [1] Haifeng Yu and Amin Vahdat. The costs and limits of availability for replicated services. In *SOSP '01: Proceedings of the eighteenth ACM symposium on Operating systems principles*, pages 29–42, New York, NY, USA, 2001. ACM Press.
- [2] David B. Ingham, Santosh K. Shrivastava, and Fabio Panzieri. Constructing dependable web services. *IEEE Internet Computing*, 4(1):25–33, 2000.
- [3] Jim Gray, Pat Helland, Patrick O’Neil, and Dennis Shasha. The dangers of replication and a solution. pages 372–381, 1998.
- [4] Rachid Guerraoui and Andre; Schiper. Software-based replication for fault tolerance. *Computer*, 30(4):68–74, 1997.
- [5] M. Colajanni and P. S. Yu. A performance study of robust load sharing strategies for distributed heterogeneous web server systems. *IEEE Transactions on Knowledge and Data Engineering*, 14(2):398–414, 2002.
- [6] Devarshi Chatterjee, Zahir Tari, and Albert Y. Zomaya. A task-based adaptive ttl approach for web server load balancing. In *Proceedings. 10th IEEE Symposium on Computers and Communications, 2005. ISCC 2005.*, pages 877–884. IEEE Computer Society, 2005.
- [7] Liming Liu and Yumao Lu. Dynamic traffic controls for web-server networks. *Comput. Networks*, 45(4):523–536, 2004.
- [8] Emiliano Casalicchio, Valeria Cardellini, and Michele Colajanni. Content-aware dispatching algorithms for cluster-based web servers. *Cluster Computing*, 5(1):65–74, 2002.
- [9] D. M. Dias, W. Kish, R. Mukherjee, and R. Tewari. A scalable and highly available web server. In *COMPCON '96: Proceedings of the 41st IEEE International Computer Conference*, page 85, Washington, DC, USA, 1996. IEEE Computer Society.
- [10] Ariel Cohen, Sampath Rangarajan, and Hamilton Slye. On the performance of tcp splicing for url-aware redirection. In *USITS'99: Proceedings of the 2nd conference on USENIX Symposium on Internet Technologies and Systems*, pages 11–11, Berkeley, CA, USA, 1999. USENIX Association.