Preface

Tho Manh Nguyen

Institute of Software Technology and Interactive Systems,
Vienna University of Technology,
Favoriten Strasse 9–11/188,
A1040 Vienna, Austria
E-mail: tho@ifs.tuwien.ac.at

Biographical notes: Tho Manh Nguyen received his PhD in Information Systems from the Vienna University of Technology, Austria and worked as a Senior Researcher at the Institute of Software Technology and Interactive Systems. He has been awarded the Microsoft Student Travel Award, IBM Europe Student Event Recognition and the Outstanding Students Award. He is a PC member, PC chair and organiser of several international conferences and workshops and has a variety of publications in international conferences and journals in the fields of data warehousing and knowledge discovery. His research areas of interest include data warehousing, data mining and knowledge discovery, business intelligence systems, grid-based knowledge discovery, service-oriented computing, ontology and semantic management.

The rapid growth of information technologies has brought tremendous opportunities for data sharing, integration, and analysis across multiple distributed, heterogeneous data sources. In the past decade, data warehousing and mining are the well-known technologies used for data analysis and knowledge discovery in vast domain of applications. Data mining technology has emerged as a means of identifying patterns and trends from large quantities of data. Data mining has used a data warehousing model of gathering all data into a central site, then running an algorithm against that data.

A growing attention has been paid to the study, development and application of data warehousing and mining. Nevertheless, dependability aspects in these applications such as availability, reliability, integrity, privacy, and security issues are still being investigated. For example, in data warehousing applications, privacy considerations may prevent the approach of collecting data into the centralised warehouse because each data source has different privacy policy. Furthermore, the complexity of security increases as different sources of information are combined. Reliable, consistent and trustworthy of information are also significant requirements in data warehousing applications. Data mining has been shown to be beneficial in confronting various types of attacks to computer systems such as fraud detection, intrusion prevention. In some applications, e.g., clinic information system, government management, business competitive information, it is required to apply the mining algorithms without observing the confidential data values thus demands the privacy preservation. There are also many challenging issues that need further investigation in the context of data mining from both privacy and security perspectives such as mining of imbalanced data, bioinformatics data, streaming data, ubiquitous computing data, grid computing data etc.

Starting from the beginning of the ARES conference in 2006, the DAWAM workshop (workshop on "Dependability Aspects on Data Warehousing and Mining applications"), like ARES, reach its 4th year in 2009. Previous DAWAM workshops were held at Vienna University of Technology, Austria on April 20–22, 2006 (DAWAM 2006), April 10–13, 2007 (DAWAM 2007), Polytechnic University of Catalonia, Barcelona, Spain, March 4–7, 2008 (DAWAM 2008). This year, DAWAM 2009 is held at Fukuoka, Japan, March 16–19, 2009.

The goals of this workshop are to bring together users, engineers and researchers (from industry and academy) alike to present their recent work, discuss and identify problems, synergise different views of techniques and policies, and brainstorm future research directions on various dependability aspects of data warehousing and data mining applications. We strongly encourage researchers and practitioners with interest in the areas of reliability, availability, privacy and security, databases, data warehousing, data mining, and statistics to submit their experience, and/or research results. The workshop has attracted several researchers and practitioners with interests in the areas of reliability, availability, privacy and security, databases, data warehousing, data mining and statistics to discuss and share their experience and research results. We received 31 submissions from 19 countries and the Programme Committee finally selected ten papers, making the acceptance rate of 32.25%.

Amongst the 10 accepted papers, the authors of the four best papers were invited to extend their papers and resubmit them for this special issue. These extended papers had two more rounds of reviews, where reviewers made strong revisions and paid special attention to the new material. In this special issue of DAWAM 2009, the following papers were selected.

The first paper 'Defining and transforming security rules in an MDA approach for DWs' by Blanco et al. includes support to complex security rules into an MDA approach to develop secure DWs. Since DWs manage vital enterprise's information used to take strategic decisions, security measures have to be considered in the development of DWs in order to avoid unauthorised accesses to the stored information. This proposal for developing secure DWs is focused on confidentiality and provides security models at different abstraction levels (business, conceptual and logical levels). It has been aligned with an MDA architecture including CIM, PIM and PSM models and the automatic transformations between them. Nevertheless, complex security rules that can be defined in this approach by using OCL expressions are not well supported by models and cannot been automatically transformed. The paper includes support to security rules by extending the conceptual metamodel and also defines sets of transformation rules to automatically transform these models towards the final implementation.

In the second paper 'Practical algorithms for subgroup detection in covert networks', Memon et al. demonstrate how specialised data mining techniques can be used to destabilise terrorist networks. Law enforcement and intelligence agencies across the globe have begun to focus on innovative knowledge discovery techniques to aid in the analysis of terrorist information. The use of such techniques in intelligence tools can help combat terrorism by predicting terrorism activity. While traditional data mining aims at extracting knowledge from data, mining for investigative analysis (investigative data mining) aims at discovering hidden instances of patterns of interest, such as patterns indicating organised crime activity. The approach is demonstrated by the *iMiner* prototype which can help to find associations between terrorist and terrorist organisations and is capable of determining links between terrorism plots that occurred in the past, their affiliation with terrorist camps, travel records, and transfers of funds, etc. These findings are represented by a network in the form of an attributed relational graph. Paths from a node to any other node in the network indicate the relationships between individuals and organisations.

Failure phenomena of web server systems are considered to depend on their workload characteristics. In the third paper 'Identifying statistical failure mechanism of web server systems: measurement and reliability analysis', Fujii and Dohi illustrate how to apply their statistical failure mechanism analysis on the Apache server system and analyse the real access/error logs. Based on parametric and non-parametric statistics, they characterise the web server failure from both theoretical and empirical points of view. As the result, it can be shown that the number of sessions strongly affects to the failure rate property of the Apache server.

The fourth and the last paper 'Experimenting with an Intrusion Detection System for Encrypted Networks' by Goh et al. presents an implementation and performance evaluation of their Network-based Intrusion Detection Systems (NIDSs) proposal based on Shamir's secret-sharing scheme and randomised network proxies, that allows a traditional NIDS to function normally in a VPN environment without any modifications and without compromising the confidentiality afforded by the VPN. They demonstrate that their approach indeed allows a standard NIDS to detect real attacks carried out over encrypted VPN channels, while minimising the probability of an attacker evading detection by the NIDS down to approximately 3.4% in the worst case scenario.

We would like to express our gratitude to all of the Programme Committee members and the external reviewers who reviewed the papers very profoundly and in a timely manner. We would also like to thank to all of the authors who submitted their papers to DAWAM 2009, as their high-quality contributions formed the basis of this year's workshop's excellent technical programme.