doi:10.1093/comjnl/bxp034

Capsule Reviews

FAIROUZ KAMAREDDINE

The Capsule Reviews are intended to provide a short succinct review of each paper in the issue in order to bring to a wider readership. The Capsule Reviews were compiled by Fairouz Kamareddine. Professor Kamareddine is an Associate Editor of *The Computer Journal* and is based in the Department of Mathematical and Computer Sciences at Heriot-Watt University, Edinburgh, UK.

Heuristic-Guided Abstraction Refinement. HE, SONG, GU AND SUN

The paper starts from the observations that model-checking suffers from the state explosion problem when applied to large-scale systems and that abstraction can solve the state explosion problem. The paper concentrates on the counterexample-guided abstraction refinement (CEGAR) and its use in hardware verification. When the counterexample is spurious (i.e. it is not a real path in the original model), the abstract model should be refined to eliminate the spurious paths. In this case, state separation creates a problem for model refinement.

To solve this separation problem, the paper proposes a heuristic-guided abstraction refinement. First, the preliminaries are presented where the original non-abstracted model, the abstraction process and the counterexample-guided approach are given. Then, the state separation problem is formulated and followed by the heuristic-guided search approach that uses greedy heuristics. Two heuristics are given, both of which guide the search to the feasible solutions of the state separation problem, but one is dynamic and finds better solutions. Furthermore, an algorithm for inferring the separation set from selected samples instead of the entire set is incorporated into the heuristic approach to give an efficient sampling technique. Two heuristic solvers (one for each given heuristic) are implemented and used to run a number of Benchmarks and experiments in order to compare the two proposed heuristics and other approaches.

Metrics to Evaluate the Use of Object-Oriented Frameworks. SILVA AND FREIBERGER

Having a successful object-oriented (OO) framework that is widely used depends on the efforts required in learning how to use it and on an evaluation of whether or not it was adequately used. This paper makes available historical uses of OO frameworks in a quantitative presentation and gives an objective evaluation of the use of OO frameworks (based on general principles, design constraints and historical use). The main contributions of the paper are the development of a set of metrics that allow OO framework developers and users to obtain objective feedback. This is done through two tools developed by the authors, which provide the automatic obtaining, storing and handling of metrics. First, the paper discusses in detail the quality features of the OO framework approach explaining how an analysis of such quality features (inherent from the general OO frameworks as well as from design decisions) can be made in an objective way. This analysis is exploited to give the metrics that help OO framework reuse, OO framework general specialization and OO framework design decision adherence.

The paper shows how these metrics can be used to evaluate OO framework-based software. The two tools are then presented and a case study with four applications is given. The case study illustrates the strengths and weaknesses of the analysis based on the proposed metrics. Other experiments consolidate the results and conclusions of the paper.

On Formalizing and Normalizing Role-Based Access Control Systems. POWER, SLAYMAKER AND SIMPSON

Role-based access control (RBAC) supports flexible approaches to authorization. In this context, a role is considered to be a collection of permissions, and users are only granted permissions through roles they are assigned. In 2004, the American National Standards Institute (ANSI) approved an RBAC standard (referred to as the ANSI RBAC standard) which aimed to provide an accepted model of RBAC, which gives a general agreement on the definition of RBAC features. This standard suffers from drawbacks, which are addressed in this paper. More specifically, this paper uses the Z notation to give a formal model of RBAC to reason about normalization and equivalence.

This model is also shown to be suitable to handle well the semantics of inheritance as well as the treatment of the role of hierarchies.

The success of this approach is due to the fact that the schema language of Z allows for optional components such as role hierarchies to be added without changing the underlying model. The paper takes an earlier (informal) model of RBAC (by Li *et al.*) and creates a more formal model that uses the Z schema language. The paper defines the notion of equivalence of core

RBAC systems and the normalization process, which transforms all equivalent core RBAC systems into the same equivalent core RBAC system. This normalization approach is given as a sequence of precisely defined operation schemas where composition is used to form a single normalization operation. Then, the paper addresses the issue of adding role hierarchies to the schema-based model. The notions of equivalence and normalization are then revised to deal with hierarchical RBAC systems. Two examples of the normalization process for hierarchical RBAC systems are then given. Afterwards, the paper studies the modelling of exclusive role constraints and the role inheritance semantics.

Serum Proteomic Abnormality Predating Screen Detection of Ovarian Cancer. GAMMERMAN, VOCK, BURFORD, NOURETDINOV AND LUO

Recent studies, which bring together mass spectrometry and machine learning, have shown to be useful at distinguishing between healthy individuals and cancer patients. This paper aims to demonstrate that the information contained in mass spectra together with the level of an established tumour marker serum is helpful for early detection of ovarian cancer. The paper adds one extra dimension to earlier studies, by using samples that predate the diagnosis of cancer in the cancer patients. Women in this study were screened for ovarian cancer for up to seven years and hence have serial samples (and for those who developed ovarian cancer, between 2 and 11 serial samples are available per woman). Extensive measurements were carried out in this pilot study. The serum samples were collected as part of a pilot trial of ovarian screening involving 13 460 post-menopausal women. Of these, 6682 women were randomized to the screening group. An extensive and lengthy sample collection and handling process is described and followed by a discussion of the serum purification process and of the pre-processing of the raw data.

Test Selection for Hierarchical and Communicating Finite State Machines. FLORENTIN AND IPATE

Conformance testing tests whether for a finite state machine (FSM) specification whose transition diagram is known and for a black-box implementation, there is a set of test sequences that establish that the implementation conforms to the specification. In conformance testing, hierarchical and concurrent models are usually turned into behaviourally equivalent FSMs, which are used as the basis for test generation. This approach suffers from drawbacks including state explosion. To overcome the drawbacks, this paper proposes a method for generating tests for a hierarchical FSM by reusing and refining the tests for the FSM components of the hierarchy. In Section 2, the author introduces basic FSM concepts and results and briefly describes the W-method, a general approach to uncover possible errors in an implementation. In Section 3, the author moves from states as single entities to compound states that contain an internal behaviour that could be represented as an internal FSM. The resulting model is called a hierarchical FSM. The author explains in detail how test generation can be handled from a hierarchical FSM and how three test suites can be constructed. Since hierarchical FSMs can contain a history connector that remembers its last current internal state, the author studies in Section 4 testing hierarchies with history connectors. The paper only studies shallow history connectors (where the remembered last current internal state is at the same hierarchical level). In Section 5, the author details a case study where a hierarchical FSM is used to specify a word processor. This case study illustrates the advantages of the method, but also reveals some limitations, which are discussed in detail in Section 5. In Section 6, the approach is adapted to testing master-slave communicating FSMs (CFSMs). In Sections 7 and 8, the author gives a discussion of his proposal and of the related work.

UML Modelling of User and Database Interaction. ALMENDROS-JIMENEZ AND IRIBARNE

User interactions and database interactions are at the heart of domain-specific modelling (DSM). User interactions are achieved by the means of user interfaces. Database interactions are usually due to user interactions. Both user and database interactions involve client-side processes as well as server-side processes. In this context, a suitable software architecture should separate concerns (e.g. presentation/business logic, navigation/interaction, workflows/transactions, etc.). The model–view–controller (MVC) pattern is a design pattern proposed in the literature for the organization of user interface programs. This paper proposes that previous MVC architecture is complex enough to be considered as a UML profile and that user and database interactions should be modelled in UML in separate diagrams. The paper presents a design technique for user and database interactions based on UML where user interfaces and user interactions are modelled in UML by means of class and state diagrams, whereas databases and database interactions are modelled by means of class and sequence diagrams. Furthermore, user requirements are modelled using the use case diagram. This proposal accommodates the MVC architecture. After the discussion of the related literature and the earlier approach of the authors, which is extended here, the paper presents in Section 2 the proposed design technique for UML modelling of user and database interactions. Throughout, an Internet Book Shopping (IBS) system is used as a case study to illustrate the approach. The authors explain how each of the relevant diagrams is built (the use case diagram, user interaction diagrams, user interface diagrams, user interface class diagrams, database class diagrams and database–interaction diagrams). Since a number of rules and conventions and synchronizations are usually involved in this technique, the authors sketch the requirements for a tool that supports the design technique. The case study of the IBS system is presented in detail in Section 3,

where all the relevant diagrams are developed and followed by a sketch of the code generation.

Tournament Coding of Integer Sequences. TEUHOLA

Compression is an essential part of saving resources in the storage and transmission of data. Compression consists of two phases (modelling and coding). To ensure compression efficiency, statistical coding is recommended. However, this faces drawbacks when a large range of integers is to be encoded. Therefore, alternative, fast and non-statistical techniques for compact coding of arbitrarily large integers are needed. This paper aims to develop a compact representation for a sequence of integers from an arbitrary range $[0, u]$. In Section 2, the author sets his assumptions about the coding of bounded integers. In Section 3, the author reviews the so-called interpolative coding of sequences of integers (which has been shown to be superior to all other non-statistical coding methods) and which was the starting point for the tournament coding proposed in the paper. In Section 4, the author gives the basic version of tournament coding, which, like interpolation coding, starts with pairing numbers; however, instead of the sum, it uses the larger number to bound the smaller.

In a tournament tree, the winner always becomes the parent and the global winner is the root. In Section 5, tournament coding is elaborated in detail and an elementary analysis is given where n source integers are independent and uniformly distributed within $[0, 2k]$ with mean k (hence all nodes except the root can be encoded with at most an upper bound number of bits using some bounded semi-fixed length code for integers). This analysis concludes that for uniformly distributed independent integers, tournament coding is more efficient than interpolative coding. Sections 6 and 7 give further analysis of compression gains. In particular, Section 7 also studies experimentally the analysis of Section 5, specifically where the n source integers are non-uniformly distributed within $[0, 2k]$. Section 8 concludes.