ISSN 2186-7437

NII Shonan Meeting Report

No. 2012-10

Quantitative methods in security and safety critical applications

Annabelle McIver
Jin Song Dong
Carroll Morgan

November 9–12, 2012

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Quantitative methods in security and safety critical applications

Organizers:
Annabelle McIver (Macquarie University, Sydney, Australia)
Jin Song Dong (National University of Singapore)
Carroll Morgan (UNSW, Sydney, Australia)

November 9–12, 2012

Quantitative formal methods.

Quantitative Formal Methods deal with systems whose behaviour of interest is more than the traditional Boolean correct or incorrect judgment. That includes timing (whether discrete, continuous or hybrid), as well as probabilistic aspects of success or failure including cost and reward, and quantified information flow.

The major challenge for researchers is to develop quantitative techniques that are both supple and relevant: the former is important because theories that amplify our reasoning powers are key to understanding system behaviour; the latter is important because our ultimate goal is to improve the practice of developing, deploying and certifying actual running software in the field.

There has been intense research in the development of semantic methods and associated tools for the analysis of quantitative properties of systems, resulting in a number of mature tools having impressive portfolios of case studies.

Modelling and verification challenges

One of the major challenges in this area is that generic verification systems are difficult to apply. For example wireless protocols are large and complex and even if current algorithmic analysis (such as used in model checking) are able analyse them in principle, the modelling languages and approaches typically used in these tools do not include advanced structuring mechanisms to describe them nor the semantic detail to interpret them in a form to which algorithmic techniques can be directly applied, or even modelled. Moreover some of the semantic issues are yet to be resolved; this is the case for security systems where there is not yet agreement as to semantic structures that apply both to the specific characteristics of the systems (measurement of information flow) and the abstraction techniques that are typically used in verification techniques. Similarly in wireless applications and the smart grid, more tailored methods capturing their characteristics tend to have much greater traction than general formal methods.

From the practical verification perspective there is a great deal of untapped potential both in recent theoretical and developing tool capabilities. Researchers

have a better understanding of how to apply theoretical structures including categorical features, metric spaces and domain spaces which show how to take advantage of algorithmic techniques which similarly have been developing, but without necessarily with a particular semantics in mind. It is now within reach of researchers to bring together those advances to apply them to modern systems which rely on critical, quantified analysis.

Meeting goals

The broad aims of this proposed meeting are to explore effective modelling and analysis methods with which to tackle the above challenging problems. Through a focussed study of a selection of these case studies, the goal is to bring together existing and emerging techniques in semantics and algorithms, possibly combining them in new ways. The hope is to obtain powerful new verification techniques able to tackle the quantitative aspects of modern system designs.

The meeting

Twenty four participants met at Shonan to share ideas on the meeting's theme. Together they represented the following countries: Australia, China, Fiji, France, Germany, Japan, the Netherlands, Singapore, Switzerland, the United Kingdom and the United States.

The topics for discussion included Quantitative Information Flow, Markov Chains, Algebras, Parallel Computing, Languages, Semantics, Logics of Knowledge, and tool support for these, and Quantum Computation. These theories were applied to applications which included databases, security, geo-indistinguishability, wireless networks and train scheduling. Details of the presenters' talks appear in the next section.

Finally, a number of open questions were posed:

1. Is weak bisimulation between Markov Automata decidable in PTIME?
2. Is there a refinement relation on event structures that reduces to the probabilistic refinement in the absence of concurrency?
3. Please find a complete weak simulation (without breaking soundness) on FPSes (i. e. substochastic DTMCs).
4. Conduct an epistemic/knowledge-based program analysis of a new problem.
5. Formalise good measurements for protocol quality.
6. How can reachability probability in quantum Markov chains be computed?

Overview of the talks

Revisiting GSPNs: New Semantics and Analysis Algorithms

Speaker: Joost-Pieter Katoen (RWTH, Aachen, Germany)

Abstract: This talk shows how Markov Automata (MA) can be used to provide a truly simple semantics of Generalized Stochastic Petri Nets (GSPNs), a popular model in performance and dependability analysis that exists for more than 25 years. In fact, our approach works for all GSPNs. No restrictions are imposed on the concurrent/conflicting enabledness of immediate transitions. This contrasts with existing solutions for GSPNs.

We complement the semantics by novel analysis algorithms for expected time, long-run average time, and timed reachability objectives of MA, i.e., GSPNs. Two case studies indicate the feasibility of these algorithms and show that a classical analysis for confused GSPNs may lead to significant over-estimations of the true probabilities.

The key message is: nondeterminism is not a threat, treat it as is! This yields both a simple GSPN semantics and trustworthy analysis results.

Weak simulation is sound and complete. No! Yes! No! Yes! No!

Speaker: David Jansen (Radboud University, The Netherlands)

Abstract: Weak simulation for probabilistic systems was defined by Baier, Hermanns, Katoen and Wolf (2003, 2005), including systems involving substochastic distributions. A goal was to establish a simulation relation that is sound and complete w. r. t. the liveness fragment of the logic PCTL.

Unfortunately, the defined relation is only sound on Markov chains (without substochastic distributions) and is not complete. Together with Lijun Zhang and Lei Song, I tried several improvements of the definition, but (until now) to no avail. The presentation shows the current state of affairs.

Comments and new proposed definitions are explicitly solicited.

Algebraic Approach to Probabilistic, Nondeterministic and Concurrent Systems

Speaker: Tahiry Rabehaja (Macquarie University, Australia)

Abstract: The algebraic approach to the formal study of software has proven to be very useful for sequential as well as concurrent standard programs. It gives a simple yet powerful abstraction of complex interactions governing the behaviour of a multi-component system. There have been several generalisations of the existing algebras to account for the presence of probability that interacts with nondeterminism. In this talk, we discuss the construction of an algebra that captures most of these features in a single framework and provide the steps required toward its consistency. The structure yields a weak concurrent Kleene algebra which we apply to the formal verification of the structural properties of

a well known distributed protocol namely, Rabins solution to the choice coordination problem. This case study provides a witness to the applicability of our tool though the algebra is by no mean a panacea. Further studies are required to account for explicit probability as well as true-concurrency and the relationship between our approach and other existing tools are to be explored.

Quantitative Measurements based on Markov Models

Speaker: Lin Gui (National University of Singapore)

Abstract: Stochastic systems are abundant in the real work, whose systems are subject to various phenomena of stochastic nature, such as message loss or garbling with certain probability. In our work, we consider the problem of quantitative measurement for stochastic systems using numerical and statistical approaches. The former can obtain exact probability solution while the latter one is based on sampling, avoiding exploration of overall state space. Our objective is to develop integrative methods for quantitative measurements of those safety critical systems and financial systems those bring very high cost if there is any failure. We have done some preliminary work in following two aspects.

In the first part of the work, some preliminary work have been done in implementing some statistical approaches in verifying a model checker PAT, modeled and verified a case study of smart grids.

In the second part, we have proposed to combine numerical probabilistic model checking techniques with statistical approaches in the way that the subsystems are verified by statistical sampling methods and numerical approaches are applied for system level model checking based on MDP. This idea has been applied to the field of software reliability engineering, in terms of reliability prediction and distribution, and realized in a toolkit named RaPiD, which has been applied to investigate two real-world systems.

Future research will focus on exploring reduction techniques to improve existing methods, adapting our integrative methods in verifying general models with global properties, and performing quantitative measurements for the systems in different domains.

Automated Specification Discovery in a Combined Abstract Domain

Speaker: Shengchao Qin (University of Teesside, UK)

Abstract: Discovering program specifications automatically for heap- manipulating programs is a challenging task due to the complexity of aliasing and mutability of data structures used. This paper describes a compositional analysis framework for discovering program specications in a combined abstract domain with shape, numerical and bag (multi- set) information. Our framework analyses each method and derives its summary independently from its callers. We propose a novel abstrac- tion method with a bi-abduction technique in the combined domain to discover pre-/post-conditions which cannot be automatically inferred be- fore. The analysis does not only prove the memory safety properties, but also finds relationships between pure and shape domains

towards full functional correctness of programs. A prototype of the framework has been implemented and initial experiments have shown that our approach can discover interesting properties for non-trivial programs.

Past, Present and Future of Formal Methods for Wireless Mesh Networks

Speaker: Peter Höfner (National ICT Australia)

Abstract: Wireless Mesh Networks (WMNs) are a promising technology that is currently being used in a wide range of application areas, including Public Safety, Transportation, Mining, etc. Typically, these networks do not have a central component (router), but each node in the network acts as an independent router, regardless of whether it is connected to another node or not. They allow reconfiguration around broken or blocked paths by hopping from node to node until the destination is reached. Unfortunately, the performance of current systems often does not live up to the expectations of end users in terms of performance and reliability, as well as ease of deployment and management.The presentation will start with an overview of the (formal) methods for WMNs developed over the last two years. In particular, I will focus on a process-algebraic approach, which turned out to be very useful and powerful, and which can be complemented by model checking. Usability of our approach is illustrated by the analysis of the Ad-hoc On-demand Distance Vector (AODV) routing protocol, a popular routing protocol designed for WMNs, and one of the four protocols currently standardised by the IETF MANET working group.

The talk will then discuss possible directions for future work. In particular, the talk will discuss two topics. (a) the extension of the process-algebraic approach by probabilities to model packet loss and to allow quantitative reasoning (b) the quest of protocol comparison, i.e., under which circumstances is one protocol is better than another. Here new formalisms need to be created and evaluated.

Operational versus Weakest Precondition Semantics for the Probabilistic Guarded Command Language

Speaker: Friedrich Gretz (RWTH Aachen and Macquarie University)

Abstract: We propose a simple operational semantics of pGCL, Dijkstras guarded command language extended with probabilistic choice, and relate this to pGCLs wp-semantics by McIver and Morgan. Parameterised Markov decision processes whose state rewards depend on the post-expectation at hand are used as operational model. We show that the weakest preexpectation of a pGCL-program w.r.t. a post-expectation corresponds to the expected cumulative reward to reach a terminal state in the parameterised MDP associated to the program. Analogously, a correspondence between weakest liberal pre-expectations and liberal expected cumulative rewards can be established.

Model Checking and Synthesis for Knowledge and Probability

Speaker: Ron van der Meyden (University of New South Wales, Australia)

Abstract: There has been an intuition that notions of knowledge and probability provide a useful level of abstraction for reasoning about and designing distributed and multi-agent systems, by providing a focus on how an agents actions are related to its state of information. To support exploration of this approach, we have developed a model checker, MCK that automates analyses from this perspective. The talk will describe the status of the system and some of the problems to which it has been applied. In particular, some recent work on synthesis of implementations from knowledge-level descriptions will be covered.

Differential Privacy and Extensions Part I

Speaker: Catuscia Palamidessi (INRIA, France)

Abstract: In the first part of this talk we discuss the general problem of protecting private information and we present differential privacy, a framework which has been recently and quite successfully introduced in the area of statistical databases. We discuss the trade-off between privacy and utility, and present some fundamental result in the area. Finally, we discuss the relation between differential privacy and quantitative information flow in the min-entropy approach.

Differential Privacy and Extensions Part II

Speaker: Konstantinos Chatzikokolakis (INRIA, France)

Abstract: In the second part of this talk we generalize the notion of differential privacy so to make it applicable to domains other than databases. We start form the observation that the standard notion of differential privacy relies on the notion of Hamming distance on the set of databases, and we extend it to arbitrary metric spaces. We show various examples, and we revise some of the fundamental results of differential privacy in this extended setting. As a particular case study, we consider location-based applications, and the resulting notion of geo-indistinguishability. Finally, we present a mechanism to achieve geo-indistinguishability, and we discuss some practical applications.

Differential Privacy and Extensions Part II

Speaker: Mingsheng Ying (University of Technology Sydney, Australia)

Abstract: The probabilistic nature of quantum systems implies that quantitative formal methods will be very useful in modeling and verification of quantum systems. But the existing quantitative formal methods cannot be directly applied to quantum systems due to the essential difference between classical and quantum probability theories.

We introduce a Markov chain model of quantum systems. Some characterizations of the reachable space, uniformly repeatedly reachable space and

termination of a quantum system are derived. Based on these characterizations, algorithms for computing the reachable space and uniformly repeatedly reachable space and for deciding the termination are given.

Not all bits are created equal: taking the meaning and value of secret bits into consideration for quantitative information flow measures

Speaker: Mario S. Alvim (University of Pennsylvania, US)

Abstract: The established models for quantitative information flow (QIF) models are generally based on concepts of entropy to describe the information leakage in a system. The most used measures of entropy are Shannon entropy, min-entropy, and guessing entropy, and essentially they are a measure of, respectively, how much information flows, how likely it is that the secret be guessed in one try, and how long it takes to the secret to be guessed. All these measures, however, implicitly consider that every bit of the secret has the same value. In many practical scenarios, however, some bits carry more important information than others. For instance, in a bank system, the bits representing the clients account number and pin code are more sensitive (and valuable) than the bits representing the clients street address. Therefore the leakage of one field or the other should not be considered equivalent, even if they consist in the same number of bits.

In this talk we will discuss ongoing work on how to define good measures of value for QIF. Among the measures considered, we propose generalizations of min-entropy and guessing entropy that take the value of bits into account. We focus on deterministic systems and link the measures of value to the Lattice of Information. We also study how these measures behave when attacks can be composed by and adaptive adversary, and we give bound on the leakage of systems.

Writing Code for Large Heterogeneous Parallel Machines

Speaker: Yifeng Chen (Peking University, China)

Abstract: This talks introduces a programming interface called PARRAY (or Parallelizing ARRAYs) that supports system-level succinct programming for heterogeneous parallel systems like GPU clusters. The current practice of software development requires combining several low-level libraries like Pthread, OpenMP, CUDA and MPI. Achieving productivity and portability is hard. PARRAY extends mainstream C programming with novel array types of the following features: 1) the dimensions of an array type are nested in a tree structure, conceptually reflecting the memory hierarchy; 2) the definition of an array type may contain references to other array types, allowing sophisticated array types to be created for parallelization; 3) threads also form arrays that allow unification of various distributed communication patterns. This leads to shorter, more portable and maintainable parallel codes, while the programmer still has control over performance-related features necessary for deep manual optimization. Although the source-to-source code generator only faithfully generates low-level library calls according to the type information, higher-level programming and

automatic performance optimization are possible through building libraries of sub-programs.The techniques have been applied to large-scale applications on modern supercomputer systems.

Incorporating time to an integrated formal method

Speaker: Steve Schneider (University of Surrey, UK)

Abstract: CSP——B is a combination of CSP and B that enables the modelling of systems which are both data-rich and have complex control behaviour. The approach combines CSP processes for control with B (or Event-B) Machines for managing state. Their joint semantics uses CSP semantic models, obtained through Morgan's and Butler's wp semantics for action systems. Recent work has considered the use of CSP——B to model railway signalling systems. The models are suitable for automated safety analysis, but consideration of capacity issues needs the models to include timing behaviour. Timed CSP is a natural candidate for introducing time into the models. However, its combination with B is not so straightforward, since the wp semantics is not sufficient for Timed CSP semantics, and hence new challenges arise. This talk will consider the challenges of combining Timed CSP with B, and our approach to addressing them. This is an active and ongoing area of research.

Model Checking Hierarchical Probabilistic Systems

Speaker: Jun Sun (SUTD Singapore University of Technology)

Abstract: Probabilistic modeling is important for random distributed algorithms, bio-systems or decision processes. Probabilistic model checking is a systematic way of analyzing finite-state probabilistic models. Existing probabilistic model checkers have been designed for simple systems without hierarchy. In this talk, we show how the PAT toolkit supports probabilistic model checking of hierarchical complex systems. We propose to use PCSP#, a combination of Hoares CSP with data and probability, for system modeling. For verification, we improve standard probabilistic model checking with two methods. One is that probabilistic refinement checking (against a non-probabilistic specification) can be applied to verify probabilistic systems against safety/co-safety temporal logic properties efficiently. The other is a way of distributed probabilistic model checking. We demonstrate the usability and scalability of the extended PAT checker via automated verification of benchmark systems and comparison with state-of art probabilistic model checkers.

Compositional Verification of Timed Systems

Speaker: Liu Yang (Nanyang Technological University, Singapore)

Abstract: Compositional techniques such as assume-guarantee reasoning (AGR) can help to alleviate the state space explosion problem associated with model checking. However, compositional verification is difficult to be automated, especially for timed systems, because constructing appropriate assumptions for

AGR usually requires human creativity and experience. To automate compositional verification of timed systems, we propose a compositional verification framework using a learning algorithm for automatic construction of timed assumptions for AGR. We prove the correctness and termination of the proposed learning-based framework, and experimental results show that our method performs significantly better than traditional monolithic timed model checking.

Extending Abstract Interpretation to New Applicative Scenarios

Speaker: Raju Halder (Macquarie University, Australia)

Abstract: Due to incessant growth of the amount of data, the information systems are facing serious challenges while managing, processing, analyzing, or understanding large volume of data in restricted environments. As a result of this, the performance of the systems in terms of optimization issues are really under big threat. The gap between the advances in information technology and the amount of data with which systems are dealing is a major concern for scientists now-a-days. To cope with this situation, we extend and integrate the well-established mathematical framework Abstract Interpretation to the broader context of Information Systems. In particular, we formalize a complete denotational semantics, both at concrete and abstract level, of data-intensive applications embedding data manipulation language operations such as SELECT, UPDATE, INSERT and DELETE. This theoretical proposal serves as a formal foundation of several interesting practical applications, including persistent watermarking, fine grained access control, cooperative query answering, etc. We also address the issue of program slicing refinement, leading to an abstract program slicing algorithm that covers SQL data manipulation languages as well.