

# Basic research in computer science and software engineering at SKLCS

Jian ZHANG (✉), Wenhui ZHANG, Naijun ZHAN, Yidong SHEN, Haiming CHEN, Yunquan ZHANG, Yongji WANG, Enhua WU, Hongan WANG, Xueyang ZHU

State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing 100080, China

© Higher Education Press and Springer-Verlag 2008

**Abstract** The State Key Laboratory of Computer Science (SKLCS) is committed to basic research in computer science and software engineering. The research topics of the laboratory include: concurrency theory, theory and algorithms for real-time systems, formal specifications based on context-free grammars, semantics of programming languages, model checking, automated reasoning, logic programming, software testing, software process improvement, middleware technology, parallel algorithms and parallel software, computer graphics and human-computer interaction. This paper describes these topics in some detail and summarizes some results obtained in recent years.

**Keywords** concurrency theory, real-time system, automated reasoning, formal specification, parallel algorithms, software process, middleware, computer graphics, human-computer interaction

## 1 Introduction

In China, there are about 200 state key laboratories. The State Key Laboratory of Computer Science (SKLCS) focuses on basic research in computer science. Its predecessor is the Laboratory of Computer Science, a key laboratory of the Chinese Academy of Sciences. This paper is a survey of the research activities carried out at SKLCS. It covers the laboratory's main research directions as discussed below:

— Theory and application of concurrent and real-time systems

We study various computing phenomena with the characteristics of concurrency, timeliness and mobility. On the theoretical side, we focus on different approaches to the modeling, analysis and design of concurrent and real-time systems, including process

algebras, temporal logics, timed automata, and so on. On the technology side, we try to invent new techniques for solving practical problems arising from distributed software, networking, life sciences, etc.

— Principles of software technology

Research in this area focuses on theories and methodologies for the analysis and development of software, especially techniques for improving software quality. It encompasses a broad range of subjects from formal specification, program analysis and testing, middleware, to software process improvement, software quality measurement and assurance.

— Parallel algorithms and parallel software

Research in this area focuses on large-scale parallel numerical simulation, parallel computation model, non-numerical parallel computing, parallel processing of huge volume of data, high performance algorithms, performance optimization, performance evaluation and benchmarking.

— Computer graphics and human-computer interaction

In computer graphics, we focus on realistic image synthesis, real time rendering of large and complex virtual environments, information representation and imaging, and animation. In human-computer interaction, we are interested in investigating the natural interaction between user and computer, especially ink understanding, multi-modal and context-aware interaction.

Currently, SKLCS has more than 30 staff members, including four academicians (members of the Chinese Academy of Sciences). We have won numerous awards, the most prominent being: National Natural Science Awards, one of first-class (1989) and two of second-class (1987, 1999); National Science and Technology Progress Awards, four of second-class (1992, 2000, 2002, 2005).

In the following sections, we describe some of our recent contributions in detail.

## 2 Concurrency and model checking

Some examples of concurrent systems are computer networks and communicating networks. The difficulties associated with concurrency have been tackled both through the construction of languages and concepts to make the complexity of concurrent execution manageable, and through the development of theories for reasoning about interacting concurrent processes. Theories of concurrency include process algebras, Petri nets and temporal logics. For verification of concurrent systems, one of the most practical methods is model checking.

- Process algebras

Research on process algebras conducted in SKLCS covers a wide range of topics, such as operational semantics, axiomatisation, verification algorithms and tools. One emphasis is on the “symbolic approach” to message-passing processes. Works along this line include:

Extending “symbolic transition graphs with assignment” to incorporate complex data types as well as “void actions”, which provides a theoretical framework for the implementation of a model checker for value-passing concurrent systems.

Completing proof systems for weak bisimulations in the  $\pi$ -calculus [1].

Completing axiomatization of timed-bisimulation for timed automata (joint work with Wang Yi at Uppsala University, Sweden) [2].

Proposing a tableau method for deciding weak bisimulation equivalence of context-free processes.

- Petri nets

Research on Petri nets conducted in SKLCS focuses on component-based methodology for the creation and verification of design specifications. Research results include (joint work with To-Yat Cheung at City University of Hong Kong and other researchers):

1) Proposing a property-preserving place-merging approach for the synthesis of components [3].

2) Investigating a kind of decompositions in terms of places, and providing conditions, under which if the large system possesses some properties, then the resulting components also preserve the desirable properties without the need of further verification [3].

3) Suggesting a property-preserving transition-reduction transformation to handle the synchronization problem in Petri-net-based system design, by first designing correct components without taking transition reduction into consideration and then introducing synchronizations by merging transitions of components [4].

4) Introducing three kinds of property-preserving subnet reduction methods for verifying system properties [5].

5) Extending WF-nets (for workflow analysis) by considering resources and allowing for multi-cases in a process, and providing a compositional method to construct and verify sound WF-nets [6].

- Model checking

Research on model checking conducted in SKLCS focuses on techniques for alleviating the state explosion problem and for the verification of parameterized infinite state systems. Research results include:

An approach for reducing the peak memory usage in model checking by first doing an analysis of a model with non-deterministic choices and then dividing the verification problem into sub-problems according to the underpinning LTL formulas that characterize different patterns of the executions of the model [7].

Automating the “guard strengthening” method for parameterized model checking of cache coherence protocols. During case studies, a data inconsistency error was identified (and fixed) in the German 2004 protocol [8].

Implementing a model checker for value-passing concurrent systems. A number of case studies have been carried out using this tool.

Applying model checking techniques to the testing of concurrent programs. A first-order extension of Computation Tree Logic has been proposed to describe testing purposes. Test cases are generated by slicing a system model against given formulas in the proposed logic [9].

- Modal logics

Modal logics are extensions of classical logics. Their application to the specification and verification of concurrent systems relies on adding temporal and spatial operators for the description of the execution and location of individual processes. In this area, we have obtained the following results:

A predicate-based approach to spatial logics for the Ambient calculus and the  $\pi$ -calculus, in which fix-point formulas are formed using predicate variables (instead of propositional variables) [10].

A model checking algorithm for finite-control mobile processes against formulas in the spatial logics, which is the first decidability result for model checking finite-control mobile processes against spatial properties, where both the processes and the logics admit recursions [11].

Showing that the non-deterministic choice  $+$  is definable in the mu-calculus, so that we can compare the expressiveness between the mu-calculus with process algebra-like modal logics, such as STL (Synchronization Tree Logic). For example, it was shown that the mu-calculus is as expressive as STL [12].

Extending the definability of  $+$  to FLC (Fixpoint Logic with Chop), which is an extension of the mu-calculus, that

is strictly more expressive than the mu-calculus, as FLC can express non-regular properties, and accordingly obtaining the compositionality of FLC. As an application of the compositionality of FLC, an algorithm is given to construct characteristic formulae of BPA (Basic Process Algebra) up to strong (weak) bisimulation directly from the syntax of processes in a compositional manner [13].

A bounded semantics for ACTL and a characterization of ACTL properties by propositional formulas [14].

### 3 Real-time and hybrid systems

A real-time system consists of several independent processes, which interact with each other and cooperate via communication in order to complete a common task in a given time. The correctness of such systems depends not only on the logical results of computing, but also on the time at which the computing was completed. Most real-time systems are embedded in a complex system or physical device as a subsystem, called embedded systems. Such kind of systems, normally containing two parts, i.e. continuous part and discrete part, are also called hybrid systems.

How to design and develop embedded systems, in particular safety critical systems, is a great challenge for computer scientists and experts of control theory. In the past several years, members of our lab have contributed much to advance the state of the art of attacking the challenge, including formal approaches to design real-time and hybrid systems, real-time scheduling theories and real-time applications.

- Formal approaches to the design of real-time and hybrid systems

Duration Calculus (DC) [15], due to Zhou, Hoare and Ravn, is an extension of Interval Temporal Logic (ITL) by introducing integral. It has been proved by numerous case studies that it is suitable for specifying and reasoning about real-time properties of systems. A most recent story about applications of DC is from the project AVACS (see <http://www.avacs.org>), in which DC was used to specify and verify the real-time properties of several real-world examples [16], for instance, the European Train Control System (ETCS), the Traffic Collision Avoidance System (TCAS) in Avionics and Automatic Cruise Control (ACC) in Automotive. In the past few years, we further developed the approach both in theory and in application. On the theoretical side, Zhou and Hansen summarized and generalized the well-established work related to DC and published a monograph on DC by Springer [17]. Regarding the applications of DC, we made lots of attempts to formalize scheduling theories with DC and obtained some exciting results. For example, Xu and

Zhan [18] pointed out and fixed a mistake on the Rate Monotonic Scheduler in Liu and Layland's seminal work [19].

In addition, Li and Tang [20] proposed a continuous linear temporal logic, called LTLC, which is an extension of XYZ/E [21]. A real-time system and its properties can be uniformly represented in the logical frame. The decidability issues about LTLC are investigated in Ref. [22].

Moreover, we also presented another formal semantics for Timed Automata (TA), which is a formal model of real-time systems, by combining its continuous semantics with discrete semantics [23]. The new formalism is called Finite Precision Timed Automata (FPTA). Compared with continuous and discrete TA, on the one hand, the reachability of FPTA is equivalent to that of continuous TA, if only timing constraints with left close and right open are considered; on the other hand, the language accepted by a FPTA is decidable, the same as discrete TA.

- Real-time scheduling theories

In the past years, we proposed lots of real-time scheduling algorithms to solve problems in practice and conducted their schedulability analysis. Some of them are listed as follows: DUC-LB (Distributed Utilization Control for real-time clusters with Load Balancing) [24] is a novel distributed utilization control algorithm for cluster-based soft real-time applications. Compared with earlier work on utilization control, a distinguished feature of DUC-LB is its capability to handle system dynamics caused by load balancing, which is a common and essential component of most clusters today. To measure internet bottlenecks in terms of location, capacity and available bandwidth, we proposed the BNeck [25] algorithm. Compared with existing algorithms, BNeck allows end users to measure the three bottleneck issues efficiently and accurately. In addition, several algorithms to resolve network blocking in soft real-time systems were invented. For example, Ref. [26] presented a congestion control algorithm based on an improved model in large-delay networks to design PID and PID-like controller; PSO-PID (Particle Swarm Optimization-Proportional-Integral-Differential) [27] is a novel controller for AQM (Active Queue Management) routers; and EVPSO (model of Particle Swarm Optimization with Escape Velocity) [28] equips particles with the escape velocity to avoid them trapped into local minima and to increase the diversity of population, which outperforms the basic PSO, especially for high dimension function, and can be used to handle multi-modal optimization problems.

- Real-time applications

In the past few years, we also applied well-established theories and techniques of real-time and hybrid systems to solve practical problems, such as real-time monitoring

of industrial processes, computer forensics, etc. In particular, we developed a large-scale real-time database system called Agilor.

## 4 Automated reasoning and logic programming

In the past five years, we have done extensive research on automated reasoning and logic programming with emphasis on the termination of logic programming, tabulated resolution, decidability of a class of theory, satisfiability problems in the propositional logic and the first-order logic, and so on. In this section, we describe our major achievements in these fields briefly.

- Termination of logic programs

For a program in any computer language, in addition to having to be correct logically, it should also terminate in a finite amount of time. Due to frequent use of recursion in logic programming, however, a logic program may more likely be non-terminating than a procedural program. Termination of logic programs then becomes an important topic in logic programming research. Because the problem is extremely hard (undecidable in general), it has been considered as a never-ending story. Most existing termination analysis approaches rely on some static information about the structure of the source code of a logic program, such as modes/types, norms/level mappings, models/interargument relations, and so on. In contrast, we proposed a dynamic approach which employs some key dynamic features of an infinite (generalized) SLDNF-derivation, such as repetition of selected subgoals and recursive increase in term size [29]. We also introduced a new formulation of SLDNF-trees, called generalized SLDNF-trees. Generalized SLDNF-trees deal with negative subgoals in the same way as Prolog and exist for any general logic programs.

- SLTNF-resolution

Global SLS-resolution is a well-known procedural semantics for top-down computation of queries under the well-founded model. It inherits from SLDNF-resolution the linearity property of derivations, which makes it easy and efficient to implement using a simple stack-based memory structure. However, like SLDNF-resolution, it suffers from the problem of infinite loops and redundant computations. To resolve this problem, we developed a new procedural semantics, called SLTNF-resolution, by enhancing Global SLS-resolution with loop cutting and tabling mechanisms [30]. SLTNF-resolution is sound and complete w.r.t. the well-founded semantics for logic programs with the bounded-term-size property, and is

superior to existing linear tabling procedural semantics, such as SLT-resolution.

- Decidability of fragments of predicate logic

In cooperation with Gilles Dowek of l'ecole polytechnique and INRIA, we give a new proof of a theorem of Mints that the positive fragment of minimal predicate logic is decidable. The idea of the proof is to replace the eigenvariable condition of sequent calculus by an appropriate scoping mechanism. The algorithm given by this proof seems to be more practical than that given by the original proof. A naive implementation is given in Objective Caml, version 3.08. Another contribution is to show that this result extends to a large class of theories, including simple type theory (higher-order logic) and second-order propositional logic. In this way, we obtain a new proof of the decidability of the inhabitation problem for positive types in system F [31].

- Satisfiability of logical formulas

Many problems can be regarded as deciding the satisfiability of certain logical formulas. In fact, the satisfiability problem in the propositional logic, known as SAT, is a fundamental problem in computer science and artificial intelligence. Since 1960, many researchers have been working on SAT, and many algorithms have been proposed to solve the problem. Generally speaking, there are two classes of SAT solving methods, namely, those based on backtracking search and those based on local search. We have proposed a new way for combining the two methods [32].

In practice, many problems are more naturally described by a set of first-order formulas rather than propositional formulas. However, the satisfiability problem in the first-order logic is undecidable in general. To alleviate this difficulty, we can choose to check the satisfiability in finite domains. In other words, we may only look for finite models of first-order formulas. Given a set of first-order formulas and a positive number  $n$ , it is always possible to decide whether there is an  $n$ -element model of the formulas. This is usually done through backtracking search. But a straightforward search procedure is almost useless, because of its low efficiency.

An important way to increase the efficiency of finite model searching is to employ the symmetry in the search space. In the early 1990s, we have proposed an isomorphism elimination method called the Least Number Heuristic (LNH). The LNH is very effective on many problems. But it does not help much if the problem has many predicate symbols. To remedy this weakness, we suggested and experimented with some techniques for eliminating isomorphic subspaces, based on Ramsey numbers [33]. More recently, we proposed a general-purpose technique, called Decision Assignment Scheme

Heuristic (DASH), which can eliminate isomorphism completely during the search [34].

## 5 Specification acquisition and reuse

MLIRF method [35] is a formal method for representation, acquisition, and reuse of formal specifications. The aim is to study how to assist human users by computer, through human-machine cooperation, to develop precise, complete and consistent formal specifications. Specifications are developed from human users' vague, incomplete and inconsistent informal statement of needs about target problems, together with, and making full use of, known specification knowledge. These specifications are approved by human users through verification and then used as the basis of software design and implementation (shown in Fig. 1). An ultimate solution to the problem still needs much research effort. Here we present our current achievements. The essentials of MLIRF method are introduced as follows.

- Representation of formal specifications

A specification is split into two parts: a problem space and operations upon it, which respectively correspond to the syntactic and semantic aspects of the specification. The problem space or the syntactic part of a specification consists of syntactic descriptions of the problem class to be solved and its solution class. Each such description, called a concept, describes a set of all instances belonging to a class, and is represented as a context-free language. Operations or semantics of specifications are expressed by recursive functions on context-free languages (CFRF) [36,37].

- Acquisition and reuse of specifications

It is possible to employ machine aided techniques in the acquisition of the specifications. It is known that context-free grammars can be inferred from samples, which is

called grammatical identification. To make the acquisition of specifications effective, we need the algorithm with some features that existing algorithms do not have. Therefore, a new method for grammatical identification is proposed [38]. Acquisition of concepts is an interactive process which employs an integration of measures. It can either perform inductive learning to get grammar rules from representative samples of concepts, which produces grammars that can reflect naturally structures of the concepts and works efficiently, or reuse known concepts in the learning process, or allow the specifier to give some pieces of concepts, or edit the intermediate result of acquisition.

It is also possible to infer recursive functions, or assist in the construction of recursive functions. Since recursive function learning techniques are still premature in practice, we take a more practical way, in which the machine assists users to inductively construct definitions of functions according to grammars of the concepts. Reuse of specification knowledge is fused in the acquisition process.

- Validation of specifications

The objective of specification validation is to assure that the specifications are in accordance with the intention of the specifiers. This can not be done automatically. What we can do is to provide techniques to help specifiers determine the validity of specifications. The measures provided for concept validation are as follows. The user may offer new samples conforming to their intention and check them against the obtained grammar. On the other hand, new samples of a grammar can be generated from the grammar as well for users to check if they are desired. The measure provided for the validation of operations is to perform the operations on samples to see if they conform to the user's original intention.

An experimental system SAQ (Specification AcQuisition system) has been implemented to support the MLIRF

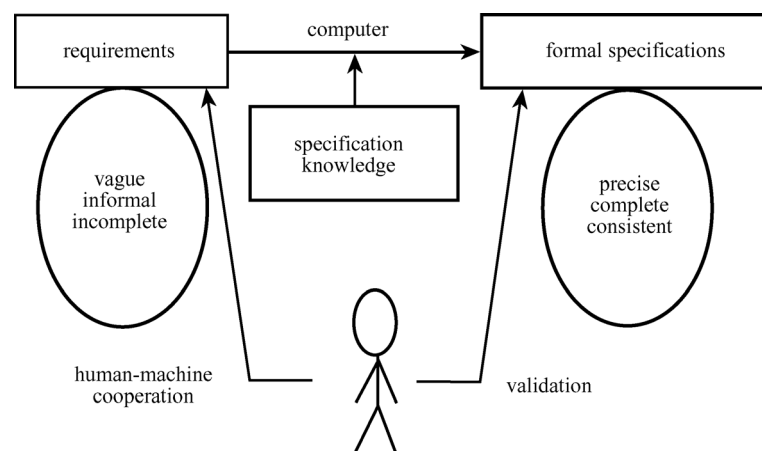


Fig. 1 Specification acquisition, reuse and validation

method [39]. Both a Solaris version and a Microsoft Windows version of SAQ have been developed. Several examples have been tried using SAQ, including formal specifications for classical algorithms on common data structures (such as list, stack, queue, tree, forest, graph, etc), a Java to C++ convertor, a Basic to C convertor, a Lisp interpreter, XML processing, a music notation system, and formal differentiation of elementary functions.

---

## 6 Parallel algorithms and parallel software

In recent years, the parallel computation group focuses on non-traditional structured grids and parallel computational model.

- Non-traditional structured grids

This concept is used in the research of the generalized Fast Fourier Transform (FFT) theories. By combination of the Approximation Theory and the Numerical Solution Methods for Partial Differential Equations (PDEs), we generalized the Fourier Methods to a class of irregular domains, such as arbitrary triangles and parallel hexagons in two dimensions [40–42], tetrahedrons and rhombic-dodecahedrons in three dimensions [42,43]. The computational complexity is reduced from  $O(N^4)$  to  $O(N^2 \log N)$  for two dimension FFTs, and from  $O(N^6)$  to  $O(N^3 \log N)$  for three dimension FFTs, respectively. The general algorithm framework (HFFT [44]) of the generalized FFTs for arbitrary simplices and super-simplices are set up and a public domain software package, FFTH, is developed.

Recently, we found the internal relationship between the multivariate Fourier Transforms and the algebraic orthogonal polynomials in Numerical PDEs. Based on second order self-adjoint PDE eigen-problem, the definitions of orthogonal polynomials over many typical domains, such as the cubes, the spheres, the simplices, the super-simplices and the curved-boundary simplices, are established. And many interesting facts such as the three layer recurrence formulas, the Chebyshev-like polynomials, the Gaussian cubatures, etc, are revealed.

- Computational model RAM( $h,k$ ) [45–47]

It is a computational model with considerations of instruction/thread level parallelism ( $k$  ways) and memory hierarchy ( $h$  levels) based on the memory access complexity concept [48]<sup>†</sup>. We argue that the complexity analysis of

algorithms should include traditional computation complexity (time and space complexity) and new memory access complexity (to reflect the different memory access behavior on memory hierarchy of different implementations). Unlike the UMH model [49] that gives detailed architectural model and scheduling of data transfer, RAM( $h,k$ ) accounts for the different memory hierarchy level by their different real measurable memory access cost under different memory access patterns of an algorithm including temporal and spatial locality, contiguous and non-contiguous accesses, short message and long message, etc. So different implementations of one algorithm (usually with the same computation complexity) will have different memory access complexity under RAM( $h,k$ ) model. Thus, we model the computer as one RAM with different memory access cost on each level under different memory access patterns.

- Parallel computational model DRAM( $h,k$ ) model [45–47]

It is the parallel version of RAM( $h,k$ ) model. It consists of  $p$  independent RAM( $h,k$ ) processors, each with its own local memory. They are connected at the memory level and communicate through message passing. It views message passing as another level of memory access to unify the analysis. The short message is viewed as non-contiguous memory access and the long message is viewed as contiguous memory access. The cost of message passing is analyzed using the LogP model [50]. Thus, sometimes it can be viewed as the combination of RAM( $h,k$ ) model and LogP model. DRAM( $h,k$ ) model pays more attention to the real measurable non-uniform access cost of local memory hierarchy and message passing.

---

## 7 Software process

The expanding role of software and information systems in the world has focused increasing attention upon the need for assurances that software systems must be developed at an acceptable speed and cost, on a predictable schedule, and in such a way that the resulting systems are of acceptably high quality and can be evolved surely and rapidly as usage contexts change. This sharpened focus is creating new challenges and opportunities for software process technology. To address these challenges and opportunities, we have developed a series of innovative methods and techniques for software process. Through Software Process Workshop (SPW2005, SPW/ProSim2006), the 1st International Conference on Software Process (ICSP2007) and the 19th International

---

<sup>†</sup> In the original paper, it is called memory complexity. To avoid confusion with space complexity, we changed it to memory access complexity later.

COCOMO (Constructive Cost Model) Forum, they have spread widely. The results are summarized below.

- Software process modeling

TRidimensional Integrated SOftware development Model (TRISO-Model) integrates process management, development technology and human beings for software development processes. Based on it, a process assessment and improvement approach has also been proposed to evaluate software process models, software organizations, software projects and the software development processes in general [51–54]. Software processes are highly people-dependent. According to this character, an organization-entity holding certain capabilities is defined as the process agent, and an organizational entity capabilities based software process modeling method has been presented to construct the optimized software process model [55–60]. A graphical modeling language based on the process algebra has also been developed, through which the operational semantics of the software processes can be described precisely and the inherent mobility in the execution of software processes can be modeled effectively. Those theories have been successfully applied to the verification of software processes against specified properties and to the conformance checking between the process performance and the pre-defined standardized process [61,62]. Based on the model driven architecture (MDA), the meta-model of different types of software processes, rules and the engine of transforming and merging among different meta-models have also been developed to merge different types of software processes [63,64].

- Empirical software engineering

To cope with the uncertainties during software cost estimation, the COCOMO-U cost estimation model extending COCOMO II has been proposed. COCOMO-U is good at cost estimation and decision making in the early phase of software development when the project has various uncertainties [65,66]. To improve cost estimation and assessment practices in Chinese government contract pricing, the COConstructive GOovernment cost Model (COGOMO) has been developed, which can provide guidances and insights for formal cost estimation [67]. Software measurement is an important facility to support effective and reasonable process management. Some process management and measurement methods have been proposed for controlling the process and improving the maturity level of an organization [68,69]. An active measurement model (AMM) is developed for supporting software process improvement [70]. Based on AMM, software organizations can both design the appropriate measurement process according to the goals of the processes they focus, and identify the opportunity and roadmap of improvement according to the result of

measurement. In addition, a BSR (Baseline-Statistic-Refinement) method has also been proposed for establishing and refining the software process performance baseline, which has been applied in many software organizations and seems to be effective and efficient in the quantitative process management [68]. Productivity is a critical performance metric of process resources and an improved productivity time series prediction method named ARIMAmse is developed, which can facilitate the software organizations to predict the trend of successive history productivity time series in an improved precision [71]. Software development task is the atomic unit of a software project. A novel software development tasks benchmarking approach based on data envelopment analysis (DEA) is developed, which can support software developers to identify the high performance tasks under multivariate and variable return to scale requirements and to provide quantitative personal software process improvement [72]. Effective capability assessment of individual software processes is a key issue in the validation of the past and the adjustment of the future development processes. A novel combined method for capability assessment of individual software processes and practices is presented in Ref. [73]. Evaluation of project quality can lead to a better control of the schedule, cost and resources allocation and smooth the way for process improvement efforts. A DEA-based approach is proposed to evaluate the project quality [74].

- Requirements engineering

A Tridimensional Requirements Model (TRISO-RM) has been proposed, which describes the win conditions of stakeholders on different aspects of software development. TRISO-Elements, each of which is formed by interconnected actors, an artifact, and an activity, are used as the medium to build and maintain the relationships among stakeholder win conditions. The production mechanism of model clashes is discussed and the process to find and avoid them is presented based on TRISO-RM [75]. A novel user-driven requirements elicitation method has been proposed. Based on users' individualities and context, support for users to present requirements is provided, including the recommendation of domain assets and advice about multi-user cooperation [76]. A risk-driven requirements prioritizing method, which combines adaptive planning and risk-driven methodologies, has been proposed. Requirements are prioritized adaptively, with risks as the foundation of priorities decisions [77]. A risk-driven release planning method for extreme programming has also been proposed. The method uses risks analysis to expose the strengths and the weaknesses of a release plan and to improve developers' understanding of system goals and awareness of their abilities [78].

- Tool support

Based on the methods and techniques mentioned above, an integrated software process services technology and management system has been developed. The system integrates quality assurance, development assistance and service support. It is quite extensible and easy to deploy. So far, the system has been applied successfully and widely in many software organizations including enterprises, universities and research institutes. This work won a second-class prize of the National Science and Technology Progress Award in 2005.

---

## 8 Distributed computing and middleware

The Distributed Computing and Middleware group focuses on developing a variety of middleware for distributed environments and tools supporting distributed system development. We work on the principles, technologies and systems for distributed computing. Our major research interests include:

- Distributed algorithms
- Key technologies for mobile computing environment
- Network-based content management technologies
- Models and technologies for adaptive middleware
- Service-oriented computing
- Component-based software methodology
- Model-driven architecture
- Testing and verification of distributed software

For the transactional composite service (TCS), we proposed a relaxed transaction model [79], including system mode, relaxed atomicity criterion, static checking algorithm and dynamic enforcement algorithm. With this model, the users are able to define different relaxed atomicity constraint for different TCS according to application-specific requirements, including acceptable configurations and the preference order. The checking algorithm determines whether the constraint can be guaranteed to be satisfied. The enforcement algorithm monitors the execution and performs transaction management work according to the constraint. Our approach can handle complex application requirements, avoid unnecessary failure recoveries and perform the transaction management work automatically.

We have also designed and implemented some software tools. For example, OnceAS/Q [80] is a web application server (WAS) with Quality of Service (QoS) guarantee. It provides a set of QoS related services and a framework to support QoS guarantee.

---

## 9 Computer graphics

The major research areas of computer graphics group are virtual reality, realistic image synthesis and scientific

visualization. Since 1990's, the group has been doing research in these areas. Some results are summarized in the following:

- In the early nineties, when the global illumination of radiosity technology was developed in its infancy stage, for the first time, a novel approach on the solution for bump texture generation by radiosity calculation was proposed. The work solved one of the crucial problems in the development of radiosity technology, and was accepted by ACM/SIGGRAPH90 [81]. At the same time, the radiosity solution to the generation of other related geometries, such as furry surfaces and fractal surfaces, was also proposed [82,83]. By application of novel techniques in global illumination and image based modeling to natural objects and environment, excellent contribution has been made to render special effect, such as knitwear, snowy scenery, etc [84,85]. Study on visualization of scalar data also started in the late nineties [86–89].

- The advance of graphics processing technology and hardware in recent years has been strongly pushing the progress of real time processing of realistic image synthesis. The computer graphics group has focused the investigation on this aspect, and has contributed mainly to two subjects: physically based modeling and animation, and graphics processing to large environments.

On the physically based modeling and rendering, work has been conducted on the simulation of fluid dynamics based on contemporary graphics processing unit (GPU). In particular, design and implementation has been made on real time simulation of fluid dynamics for medium size environment with arbitrary boundary conditions, for the first time, fulfilled on GPU [90–92]. More realistic interaction has been simulated between fluid & solid substrate and among fluid mixtures between different features of fluids [93–95]. These works have been presented at important journals or conferences. We have also been invited as keynote speakers by prestigious international conferences, like CASA2004 [90] and CyberWorlds2006 [95].

- On the subject of geometric processing to large out-of-core models, we proposed an approach, which is recognized as one of the pioneering works for non-uniform simplification to out-of-core models [96].

- On the fundamental research for geometric processing, a geometric subdivision method was proposed for inclusion test of points within a geometry expressed as polygons(2d) or Polyhedrons(3d). The



work has been accepted by IEEE Transaction and European publications [97,98].

## 10 Human-computer interaction techniques

- Ink processing based on pen-paper metaphor

Pen and paper are natural and effective interaction patterns used by people for a long time. The Human-Computer interaction group focuses on digital ink description based on temporal data model and ink understanding based on cognitive psychology. We proposed an ink structure understanding algorithm for handwriting digital ink, which can separate textual ink from graphical ink, recognize paragraphs, segment lines and extract characters. Based on the ink understanding algorithm, we proposed a method to correct handwriting recognition errors by speech. Users can repeat the text by speech to correct several character recognition errors in a line of Chinese handwriting. Our results were presented at ACM IUI [99–101].

- Visualization of dynamic cursor

We explored the visualization method of pen tilt information in pen-based user interface and its influence on interaction process and proposed a method to generate cursor shape on the computer screen which visualizes pen tilt information. Empirical evaluation shows that the method can enhance the stimulus-response compatibility of touchpad in pen-based user interface. This work was presented at ACM CHI 2007 [102].

- Large hierarchical data visualization

We proposed a circle packing technique to visualize large hierarchical data. Compared with other methods, it can show both data hierarchy and single data nodes. It provides a new way to visualize large volumes of hierarchical data. This work was presented at ACM CHI 2006 [103].

## 11 Concluding remarks

Computer science is a young field and is evolving rapidly. In addition to the major areas outlined in the previous sections, we are also interested in other research topics, such as algorithms and complexity [104,105], software testing [106,107], data mining [108], semantic web, quantum computing, and so on.

We hope that our results will have a significant impact on the basic research in computer science and will also be helpful to Chinese software industry.

## References

1. Lin H M. Complete inference systems for weak bisimulation equivalences in the  $\pi$ -calculus. *Information and Computation*, 2003, 180(1): 1–29
2. Lin H M, Wang Y. Axiomatizing timed automata. *Acta Informatica*, 2002, 38(4): 277–305
3. Jiao L, Huang H J, Cheung T Y. Property-preserving composition by place merging. *Journal of Circuits, Systems and Computers*, 2005, 14(4): 793–812
4. Huang H J, Jiao L, Cheung T Y. Property-preserving subnet reductions for designing manufacturing systems with shared resources. *Theoretical Computer Science*, 2005, 332(1–3): 461–485
5. Jiao L, Cheung T Y, Lu W M. Handling synchronization problem in Petri-net-based system design by property-preserving transition-reduction. *The Computer Journal*, 2005, 48(6): 692–701
6. Jiao L, Cheung T Y. Compositional verification for workflow nets. *Journal of Circuits, Systems and Computers*, 2006, 15(4): 551–570
7. Zhang W H. Combining static analysis and case-based search space partitioning for reducing peak memory in model checking. *Journal of Computer Science and Technology*, 2003, 18(6): 762–770
8. Lv Y, Lin H M, Pan H. Computing invariants for parameter abstraction. In: Schneider K, Hoe J, eds. *Proceedings of MEMOCODE*. Nice: IEEE Press, 2007, 29–38
9. Wu P, Lin H M. Model-based testing of concurrent programs with predicate sequencing constrains. *International Journal of Software Engineering and Knowledge Engineering*, 2006, 16(5): 727–746
10. Lin H M. A predicate mu-calculus for mobile ambients. *Journal of Computer Science and Technology*, 2004, 20(1): 95–104
11. Lin H M. A predicate spatial logic and model checking for mobile processes. In: Liu Z M, Araki K, eds. *Proceedings of ICTAC*. Berlin: Springer, 2004, 36
12. Zhan N J, Majster-Cederbaum M. Deriving non-determinism from conjunction and disjunction. In: Wang F, ed. *Proceedings of FORTE*. Berlin: Springer, 2005, 351–365
13. Zhan N J, Wu J. Compositionality of fixpoint logic with chop. In: Hung D V, Wirsing M, eds. *Proceedings of ICTAC*. Berlin: Springer, 2005, 136–150
14. Zhang W H. Model checking with SAT-based characterization of ACTL formulas. In: *Proceedings of ICFEM*. Berlin: Springer, 2007, 191–211
15. Zhou C C, Hoare C A R, Ravn A. A calculus of durations. *Information Processing Letters*, 1991, 40(5): 269–276
16. Damm W, Hungar H, Olderog E-R. Verification of cooperating traffic agents. *International Journal of Control*, 2006, 79(5): 395–421
17. Zhou C C, Hansen M R. Duration calculus: a formal approach to real-time systems. *EATCS Series of Monographs in Theoretical Computer Science*. Berlin: Springer, 2004
18. Xu Q W, Zhan N J. Formalizing scheduling theories with duration calculus. *Nordic Journal of Computing*, to appear
19. Liu C L, Layland J W. Scheduling algorithms for multiprogramming in a hard real-time environment. *Journal of ACM*, 1973, 20(1): 46–61
20. Li G Y, Tang Z S. Modeling real-time systems with continuous-time temporal logic. In: George C, Miao H, eds. *Proceedings of ICFEM*. Berlin: Springer, 2002, 231–236
21. Tang Z S. A program development support environment based on temporal logic. In: *Proceedings of PLSD*. North Holland, 1983

22. Li G Y, Tang Z S. Translating a continuous-time temporal logic into timed automata. In: Ohori A, ed. Proceedings of APLAS. Berlin: Springer, 2003, 322–338
23. Yan R J, Li G Y, Tang Z S. Symbolic model checking of finite precision timed automata. In: Hung D V, Wirsing M, eds. Proceedings of ICTAC. Berlin: Springer, 2005, 272–287
24. Fu Y, Wang H. Distributed utilization control for real-time clusters with load balancing. In: Proceedings of IEEE RTSS. Rio, 2006, 137–146
25. Zhou H, Wang Y J, Wang Q. Measuring internet bottlenecks: location, capacity, and available bandwidth. In: Lu X C, Zhao W, eds. Proceedings of CNMC. Berlin: Springer, 2005, 1052–1062
26. Wang X X, Wang Y J, Zhou J H, et al. Congestion control algorithm based on improved model in large-delay networks. Acta Electronica Sinica, 2005, 33(5): 842–846
27. Wang X L, Wang Y J, Zhou H, et al. PSO-PID: a novel controller for AQM routers. In: Proceedings of IEEE/IFIP WOCN. Bangalore, 2006, 1–5
28. Wang X L, Wang Y J, Zeng H T, et al. Particle swarm optimization with escape velocity. In: Proceedings of CIS. Guangzhou: IEEE Press, 2006, 457–460
29. Shen Y D, You J H, Yuan L Y, et al. A dynamic approach to characterizing termination of general logic programs. ACM Transactions on Computational Logic, 2003, 4(4): 417–430
30. Shen Y D, You J H, Yuan L Y. Enhancing global SLS-resolution with loop cutting and tabling mechanisms. Theoretical Computer Science, 2004, 328(3): 271–287
31. Dowek G, Jiang Y. Eigenvariables, bracketing and the decidability of positive minimal predicate logic. Theoretical Computer Science, 2006, 360(1–3): 193–208
32. Zhang W H, Huang Z, Zhang J. Parallel execution of stochastic search procedures on reduced SAT instances. In: Proceedings of PRICAI. Berlin: Springer, 2002, 108–117
33. Jia X X, Zhang J. Predicate-oriented isomorphism elimination in model finding. In: Proceedings of IJCAI. St Louis: Morgan Kaufmann Publishers, 2005, 1525–1526
34. Jia X X, Zhang J. A powerful technique to eliminate isomorphism in finite model search. In: Proceedings of IJCAR. Berlin: Springer, 2006, 318–331
35. Dong Y M. MLIRF method for specification acquisition and reuse. In: Proceedings of the 9th National Conference of China Computer Federation. Chongqing, 1996, 21–27 (in Chinese)
36. Dong Y M. Recursive functions of context free languages (I) --- the definitions of CFPRF and CFRF. Science in China Series F, 2002, 45(1): 25–39
37. Dong Y M. Recursive functions of context free languages (II) - validity of CFPRF and CFRF definitions. Science in China Series F, 2002, 45(2): 1–21
38. Dong Y M. An interactive learning algorithm for acquisition of concepts represented as CFL. Journal of Computer Science and Technology, 1998, 13(1): 1–8
39. Dong Y M, Li K, Chen H, et al. Design and implementation of the formal specification acquisition system SAQ. In: Proceedings of Conference Software: Theory and Practice, IFIP 16th World Computer Congress 2000. Beijing, 2000, 201–211
40. Sun J C. Multivariate fourier series over a class of non tensor-product partition domains. Journal of Computational Mathematics, 2003, 21(1): 53–62
41. Sun J C, Yao J F. Fast generalized discrete fourier transforms on hexagon domains. Chinese Journal of Numerical Mathematics and Application, 2004, 26(3): 351–366
42. Sun J C, Li H Y. Generalized fourier transform on an arbitrary triangular domain. Advances in Computational Mathematics, 2005, 22: 223–248
43. Yao J F, Sun J C. HFFT on parallel dodecahedron domains and its parallel implementation. Numerical computation and computer applications, 2004, 4: 304–313 (in Chinese)
44. Sun J C. Multivariate fourier transform methods over simplex and super-simplex domains. Journal of Computational Mathematics, 2006, 24(3): 305–322
45. Zhang Y Q. Performance optimizations on parallel numerical software package and study on memory complexity. Dissertation for the Doctoral Degree. Beijing: Institute of Software, Chinese Academy of Sciences, 2000
46. Zhang Y Q. DRAM(*h*): a parallel computation model for high performance numerical computing. Chinese Journal of Computers, 2003, 26(12): 1660–1670
47. Zhang Y Q, Chen G L, Sun G Z, et al. Models of parallel computation: a survey and classification. Frontiers of Computer Science in China, 2007, 1(2): 156–165
48. Zhang Y Q, Sun J C, Tang Z M, et al. Memory complexity in high performance computing. In: Proceedings of HPC'ASIA. Singapore, 1998: 142–151
49. Alpern B, Carter L, Feig E, et al. The Uniform Memory Hierarchy Model of Computation. Algorithmica, 1994, 12(2–3): 72–109
50. Culler D, Karp R, Patterson D, et al. LogP: towards a realistic model of parallel computation. In: Proceedings of PPOPP. San Diego, 1993, 1–12
51. Li M S. Expanding the horizons of software development processes: a 3-D integrated methodology. In: Proceedings of SPW. Beijing, 2005, 54–67
52. Li M S. Assessing 3-D integrated software development processes: a new benchmark. In: Proceedings of SPW/ProSim. Shanghai, 2006, 15–38
53. Wang Q, Li M S. Software process management: practices in China. In: Proceedings of SPW. Beijing, 2005, 317–331
54. Li M S. TRISO-model: a new approach to integrated software process assessment and improvement. Software Process: Improvement and Practice, 2007, 12(5): 387–398
55. Wang Q, Xiao J, Li M S, et al. A process-agent construction method for software process modeling in SoftPM. In: Proceedings of SPW/ProSim. Shanghai, 2006, 204–213
56. Li N, Li M S, Wang Q, et al. A negotiation model for the process agent in an agent-based process-centered software engineering environment. In: Proceedings of SEKE. San Francisco, 2006, 664–669
57. Zhao X, Chan K, Li M S. Applying agent technology to software process modeling and process-centered software engineering environment. In: Proceedings of SAC. Santa Fe, 2005, 1529–1533
58. Xiao J, Osterweil L J, Zhang L, et al. Applying little-JIL to describe process-agent knowledge in SoftPM. In: Proceedings of SPW/ProSim. Shanghai, 2006, 214–221
59. Zhang L, Wang Q, Xiao J, et al. A tool to create process-agents for OEC-SPM from historical project data. In: Proceedings of ICSP. Berlin: Springer, 2007, 84–95
60. Xiao J, Osterweil L J, Zhang L, et al. Applying little-JIL to describe process-agent knowledge and support project planning. Software Process: Improvement and Practice, 2007, 12(5): 437–448
61. Li M S, Yang Q, Zhai J, et al. On mobility of software processes. In: Proceedings of SPW/ProSim. Shanghai, 2006, 105–114
62. Yang Q, Li M S, Wang Q, et al. An algebraic approach for managing inconsistencies in software processes. In: Proceedings of ICSP. Berlin: Springer, 2007, 121–133
63. Yuan F, Li M S, Wan Z. SPEM2XPDL-towards SPEM model enactment. In: Proceedings of SERP. Las Vegas, 2006, 240–245
64. Li J, Li M S, Wu Z, et al. A SPEM-based software process metamodel for CMM. Journal of Software, 2005, 16(8): 1366–1377
65. Yang D, Wan Y, Tang Z, et al. COCOMO-U: an extension of COCOMO II for cost estimation with uncertainty. In: Proceeding of SPW/ProSim. Shanghai, 2006, 132–141

66. Yang D, Boehm B, Yang Y, et al. Coping with the cone of uncertainty: an empirical study of the SAIV process model. In: Proceedings of ICSP. Berlin: Springer, 2007, 37–48
67. He M, Yang Y, Wang Q, et al. Cost estimation and analysis for government contract pricing in China. In: Proceedings of ICSP. Berlin: Springer, 2007, 134–146
68. Wang Q, Jiang N, Gou L, et al. BSR: a statistic-based approach for establishing and refining software process performance baseline. In: Proceedings of ICSE. Shanghai, 2006, 585–594
69. Wang Q, Li M S. Measuring and improving software process in China. In: Proceedings of ISESE. Noosa Heads, 2005, 183–192
70. Wang Q, Li M S, Liu X. An active measurement model for software process control and improvement. Journal of Software, 2005, 16(3): 407–418
71. Ruan L, Wang Y, Wang Q, et al. ARIMAmse: an improved ARIMA-based software productivity prediction method. In: Proceedings of COMPSAC. Chicago, 2006, 17–21
72. Ruan L, Wang Y, Wang Q, et al. Empirical study on benchmarking software development tasks. In: the International Conference on Software Process. Minneapolis, 2007, 221–232
73. Zhang S, Wang Y, Yuan F, et al. Mining software repositories to understand the performance of individual developers. In: Proceedings of COMPSAC 2007. Beijing, 2007, 625–626
74. Zhang S, Wang Y, Tong J, et al. Evaluation of project quality: a DEA-based approach. In: Proceedings of SPW/ProSim. Shanghai, 2005, 88–96
75. Wang J, Li M S. A tridimensional requirements model and its support for stakeholder coordination. Journal of Software, 2007, 18(10): 2380–2392
76. Shu F, Zhao Y, Wang J, et al. User-driven requirements elicitation method with the support of personalized domain knowledge. Computer Research and Development, 2007, 44(6): 1044–1052
77. Huang M, Shu F, Li M S. A risk-driven method for prioritizing requirements in iteration development. Journal of Software, 2006, 17(12): 2450–2460
78. Li M S, Huang M, Shu F, et al. A risk-driven method for extreme programming release planning. In: Proceedings of ICSE 2006. Shanghai, 2006, 423–430
79. Huang T, Ding X N, Wei J. An application-semantics-based relaxed transaction model for internetware. Science in China Series F, 2006, 49(6): 774–791
80. Huang T, Chen N J, Wei J, et al. OnceAS/Q: a QoS-enabled web application server. Journal of Software, 2004, 15(12): 1787–1799
81. Chen H, Wu E H. An efficient radiosity solution for bump texture generation. Computer Graphics, 1990, 24(4): 125–134
82. Chen H, Wu E H. Radiosity for furry surfaces. In: Post F H, Barth W, eds. Proceedings of EUROGRAPHICS. North-Holland: Elsevier Science Publishers, 1991, 447–457
83. Wu E H. A radiosity solution for illumination of random fractal surfaces. The Journal of Visualization and Computer Animation, 1995, 6(4): 219–229
84. Chen Y Y, Sun H Q, Wu E H. Modeling and rendering snowy natural scenery using multi-mapping techniques. The Journal of Visualization and Computer Animation, 2003, 14(1): 21–30
85. Xu Y Q, Chen Y Y, Lin S, et al. Photo-realistic rendering of knitwear using the lumislice. In: Proceedings of ACM SIGGRAPH. New York: ACM Press, 2001, 391–398
86. Liu X H, Wu E H. Hierarchical structure with focus criterion for rendering height field. Journal of Computer Science and Technology, 1998, 13(12): 1–8
87. Wang W C, Wu E H. Adaptable splatting for irregular volume rendering. Computer Graphics Forum, 1999, 18(4): 213–222
88. Wang W C, Wu E H, Max N. A selective rendering method for data visualization. Journal of Visualization and Computer Animation, 1999, 10(3): 123–131
89. Wang W C, Zhou D H, Wu E H. Accelerating techniques in volume rendering of irregular data. Computers and Graphics, 1997, 21(3): 289–295
90. Wu E H, Liu Y Q, Liu X H. An improved study of real time fluid simulation on GPU (invited paper). Computer Animation and Virtual World, 2004, 15(3–4): 139–146
91. Liu Y Q, Liu X H, Wu E H. Real-time 3D fluid simulation on GPU with complex obstacles. In: Proceedings of Pacific Graphics. Seoul: IEEE Computer Society, 2004, 247–256
92. Liu Y Q, Liu X H, Wu E H. Fluid simulations on GPU with complex boundary conditions. In: Proceedings of Poster Workshop on GPGPU, at ACM SIGGRAPH, 2004
93. Liu Y Q, Zhu H B, Liu X H, et al. Real time simulation of physically based on-surface flow. The Visual Computer, 2005, 21(8–10): 727–734
94. Zhu H B, Liu X H, Liu Y Q, et al. Simulation of miscible binary mixtures based on lattice boltzmann method. Computer Animation and Virtual Worlds, 2006, 17(3–4): 403–411
95. Wu E H, Zhu H B, Liu X H, et al. Physically based fluid dynamics and interactions. In: Proceedings of CyberWorlds. Lausanne: IEEE Computer Society, 2006, 3–13
96. Fei G Z, Cai K Y, Guo B N, et al. An adaptive sampling scheme for out-of-core simplification. Computer Graphics Forum, 2002, 21(2): 111–119
97. Wang W C, Li J, Wu E H. 2D point-in-polygon test by classifying edges into layers. Computers and Graphics, 2005, 29(3): 427–439
98. Wang W C, Li J, Sun H Q, et al. A layer-based representation of polyhedrons for point containment tests. IEEE Transactions on Visualization and Computer Graphics, 2007, to appear
99. Ao X, Wang X G, Tian F, et al. Cross-modal error correction of continuous handwriting recognition by Speech. In: Proceedings of ACM International Conference on Intelligent User Interfaces. Honolulu, 2007, 243–250
100. Ao X, Li J F, Wang X G, et al. Structuralizing digital ink for efficient selection. In: Proceedings of ACM International Conference on Intelligent User Interfaces. Sydney, 2006, 148–153
101. Li J F, Zhang X W, Ao X, et al. Sketch recognition with continuous feedback based on incremental intention extraction. In: Proceedings of ACM International Conference on Intelligent User Interfaces. San Diego, 2005, 145–150
102. Tian F, Ao X, Wang H A, et al. The tilt cursor: enhancing stimulus-response compatibility by providing 3D orientation cue of pen. In: ACM Conference on Human Factors in Computing Systems. San Jose, 2007, 303–306
103. Wang W X, Wang H, Dai G Z, et al. Visualization of large hierarchical data by circle packing. In: ACM Conference on Human Factors in Computing Systems. New York: ACM Press, 2006, 517–520
104. Xia M. Maximum edge-disjoint paths problem in planar graphs. In: Cai J Y, Cooper S B, Zhu H, eds. Proceedings of TAMC, 566–572
105. Zhang P. An approximation algorithm to the  $k$ -steiner forest problem. In: Cai J Y, Cooper S B, Zhu H, eds. Proceedings of TAMC, 2007, 728–737
106. Yan J, Zhang J. Backtracking algorithms and search heuristics to generate test suites for combinatorial testing. In: Proceedings of COMPSAC. Chicago: IEEE Computer Society, 2006, 385–394
107. Xu Z X, Zhang J. A test data generation tool for unit testing of C programs. In: Proceedings of QSIC, 2006, 107–116
108. Shen Y D, Zhang Z, Yang Q. Objective-oriented utility-based association mining. In: Proceedings of ICDM, 2002, 426–433