1

How could research on testing of communicating systems become more industrially relevant ?

R. Lai
School of Computer Science and Computer Engineering,
La Trobe University, Victoria 3083, Australia.
E-mail : lai@latcs1.oz.au

Abstract

Academic research has made significant advances in the generation of test sequences from formal specifications and in the development of computer-aided test tools with the aims of improving the effectiveness of testing communicating systems. However, this state-of-the-art research is not necessarily the state-of-the-practice; these methods and tools are seldom used in the communications industry. As academia tends to work on problems that are not too industrially related and does not quite address the problems facing testers in the industry, testers generally regard academic testing techniques impractical and irrelevant to solving real problems. There is a big gap between testing practice and research results published in journals and reported at conferences. This paper argues that academic research on testing of communicating systems needs to become more industrially relevant, describes the means by which this change can be facilitated, and suggests some research topics that are relevant to the industry. It aims to help effect a change in the direction of academic research on testing of communicating systems.

Keywords

Formal Method, Formal Description Techniques, Conformance Testing, Test Sequence Generation

1. INTRODUCTION

In the last decade, some of the research work on testing of communicating systems include conformance testing, interoperability testing, test case generation, test execution, analysis of results, theoretical framework, and theories of testing. In particular, there have been much research work conducted on testing using formal methods. Three Formal Description Techniques (FDTs) - SDL (CCITT, 1987), Estelle (ISO, 1989a) and LOTOS (ISO, 1989b) - have been standardised. One of the main aims in developing FDTs is to have a formal specification acting as a sound basis for such a testing. Successes have been achieved by academia on automatic test case generation from formal specifications

B. Baumgarten et al. (eds.), *Testing of Communicating Systems*

to optimize the number of test cases in a test sequence so that minimal effort is required to obtain an acceptable level of confidence in the communication software (Sidhu and Leung, 1989).

The communications industry, with its own method of testing, is shipping to the customer communication software that still has bugs, even after it has been tested in the factory. With the merits of the academic testing methods, one would expect that they would be accepted widely in the industry. However, this state-of-the-art research is not necessarily the state-of-the-practice; they seem to have had very little impact on testing conducted in the industry (Lai and Leung, 1995). There is not much progress in the use of test sequence generation techniques for practical testing of communication networks. Test design is still largely performed by testers by interpreting the specifications written in a natural language.

Academia tends to work on problems that are not too industrially related and does not quite address the problems facing testers in the industry. The academic community likes to work on theories, with very little attention paid to real application and problems. There are very few researchers who are working on real projects and trying out new methods. Testers have to face real problems like deadlines, resource and economic constraints. For a researcher, he deals with a communicating system usually described in a finite state machine and tries to come up with a technique for testing the system on paper; for a practitioner, he is presented with an implementation of a communicating systems and needs to figure out ways to uncover most of the bugs within some economic, resource and time constraints. This is the fundamental difference between the work of a researcher and a practitioner. Testers therefore generally regard academic methods being impractical and irrelevant to solving their real problems. On the other hand, the industrial community only wants applications and have not much interests in the underlying theories. Academia and practitioners basically operate in two different cultures. There is a big gap between testing practice and research results published in journals and reported at conferences. This big gap between academic and industrial testing practices and that academia has not been addressing the real-life testing issues and problems account for the fact that academic testing methods are seldom used in the industry (Lai and Leung, 1995). There is a tremendous need to find a bridge between the theories and practices of testing.

This paper argues that academic research on testing of communicating systems needs to become more industrially relevant, describes the means by which this change can be facilitated, and suggests some research topics that are relevant to the industry. It aims to help effect a change in the direction of academic research on testing of communicating systems.

2. WHY CHANGE ?

How much longer can academia afford doing research that are not very industrially related ? The answer is "Not very long". The circumstances under which universities function in societies have changed. Gone were the days that academia could hide themselves in ivory towers to do research and afford not interacting with outsiders. Due to these changes, conducting researchs that are more industrially relevant is absolutely important for academia. The reasons are listed below.

Students : our direct customers In Australia and many countries, the level of funding that a university receives is proportional to the number of students that it has. The reputation of a university is built over a long period through its graduates, courses, and research output. More and more, universities are required to be accountable to the communities for their qualities of teaching. The contents of courses academics teach are usually related to their research interests. Slowly, communities get to know which universities' courses are more practical and which ones produce graduates that are more employable. Students tend to shy away from universities which offer courses that are less practical. As a result, such universities will receive a lower level of funding from the government; the job securities of some academics will be jeopardised. Students are the customers of academics; providing a good service to students is their duty and a means for their survivals in today's society.

Industry : our indirect customer Graduates of university courses which are too theoretical tend to have a harder time in getting their preferred jobs. Managers and practitioners want to hire graduates who are grounded in theories and also able to tackle real problems in the industry. Students having a good exposure to industrial issues during their university courses tend to perform better at interviews and hence be employed. If a company is happy with the graduates of a particular university, it usually comes back to hire some more. Producing graduates that are more employable and effective at their work is a key to building a good reputation for universities. If a university has a good perception from the industry about its graduates, more students will be attracted to its courses and subsequently a higher level funding will be received from the government.

A barometer for academic research quality What are the means to judge the quality of academic research ? Some barometers that have been suggested are : how many times a research paper is cited by other researchers and the reputation of the conference proceedings/journals in which a paper is published. In this fast moving communications field, a paper is usually cited by others once or twice. After a couple of years, the work would probably become obsolete. Is there a way to judge which journal/conference is better ? The answer is "Yes" because by and large, papers published in journal and presented at conferences are judged by academic standards; in terms of how useful the papers are in practice, the answer is probably "No". Would it be better that a different criterion be used for assessing the quality of research ? It needs to be judged by how suitable the technique is for use in the industry and how extensive it is used. This would be a more practical and objective way of measuring the usefulness and effectiveness of research work.

Expertised knowledge not just located at universities Gone were the days that universities were the only places where intellectural knowledge are located. A large percentage of the community's intellectural strength is not currently located in universities. Academics are not the only people that have the expertised knowledge. In fact, much of the leading research and development are conducted in the defence and aerospace industries; unfortunately, very seldom these reports are available to public use. The notion that the links with the wider community end when students graduate is really inappropriate. Academia can benefit by staying in touch with professionals in the industry. By working closely together, academia and practitioners can generate new intellectural products for the next century.

Time for a change For more than a decade, academic researchers have been doing work that are largely academic and have very little relevance to the real problems encountered in the industry. Is academia finding the work very interesting ? Maybe, the real problems are far more interesting than the problems discussed in the literature. Solving real problems could be more interesting and challenging for academic. There is already a huge backlog of research results not being put into practice. Therefore, it is doubtful if there is much value in continuing to stack more research results on top of existing ones which are not being used. It is high time that researchers do more work on how existing techniques can be applied to solving real problems and on deriving new methods to address industrial issues.

Government research funding In industry, the focus is on short-term results. In the academic community, conventional funding structures are rapidly disappearing, with a similar focus on short-term relevance, as opposed to basic research. In this modern down-sizing and leaner world, more and more government funds for research are required to be accountable to taxpayers. These changes make it harder for "pure researchers" to obtain government research grants. Millions of dollars have been spent on research on protocol engineering and alike in the last decade around the world. Can any one say that we have got the value for money and we have generated a worthwhile return ? Unless we address the real issues, we will not be able to get more research grants and will not deserve further funding.

Demonstrate or die The old saying for the survival of academia is "Publish or Perish". However, the system has changed to some extent for some academics. A new saying, "Demonstrate or Die", has sprung up. The ability to attract research grant from commercial firms is another way of judging the standing of a academic researcher. For instance, in the Boston area where the headquarters of many high technology computer and communications companies are located, researchers at institutions, like Massachusetts Institute of Technology, need to demonstrate to companies that their ideas are going to work on real-life systems before companies would give them research grants. Having a good idea is not good enough for attracting research funding from commercial firms. To gain promotion, academics are now very much pressurised to obtain research grants which are getting more and more competitive. Without the demonstration of the applicability of theories, there is very little hope that researchers are able to attract research funding from commerical companies.

Contribution to a nation's economy Academia in the western countries might be publishing more papers than their counterparts in countries like Taiwan. But what are really the contributions of publication to a nation's economy if the theories are not put into practice ? Sadly, there is not really a lot. This is evidenced by the trade deficits that some countries like Australia have. This is partly due to the fact that theories are seldom put into practice. On the other hand, Taiwan has one of the biggest trade surplus in the world. Taiwanese computer firms capture a substantial portion of the personal computer market of the world. One of the factors that contributes to the technological and economic successes of Taiwan is the government-initiated research infrastructure for helping the computer, electronics and communications industries. In Taiwan, research results tend to get more practised in the industry, hence making a contribution to the nation's economy.

3. CATALYSTS FOR THE CHANGE

With the huge gap between research and practice and the facts that academia are more interested in publishing papers and that practitioners are too busy to meet deadlines, making academic research on testing of communicating systems to become more industrially relevant would not happen naturally. Some catalysts are required to help bring about this change. These catalysts are discussed below.

3.1. Paper reviews

It is largely academics who review papers and control conferences, workshops, and publications. Academic reviewers do not like simple approaches and tend to reject papers which use simple techniques, even if they are useful. They usually do not like papers that only deal with real problems and large systems as they generally consider such papers not being original enough, even though practitioners have to solve problems in a real world environment. Vast majority of reviewers are in academic positions. So the standards by which papers are judged are by and large academic. Reviewing a paper from an academic point of view is not too helpful towards judging the true value of a paper. Their attitudes towards industrial papers and reviewing them somehow need to be changed.

Recently, it is encouraging to see that some journals emphasis the applications of theories to real problems. An example is the Journal of Systems and Software published by Elsevier Science; the editor-in-chief spells out that papers should contain an evaluation of the theories used and those that do not meet this criterion are returned to the authors without being sent off for review. We need more editors and reviewers who take a different view of reviewing and judging the values and the qualities of those papers that discuss real problems and systems.

3.2. Bridging agents

Academia are too busy writing papers and industrial people are under much pressure to finish their projects by the deadlines. Neither academia nor practitioners would have the time for each other. Agents are required for practitioners and researchers to interact, assuming that they have the genuine desire to interact. Who should act as these agents ? Professional bodies are the most promising agents to accomplish this task as many conferences are sponsored by organisations like IEEE, ACM and IFIP. Professionals in the industry need to take advantages of the availability of such conferences and journal publications to examine the work of academia more closely. An important mission of these societies should be to narrow the ever-growing gap between research and practice and to reconcile these two worlds. These professional societies need to take a better lead to improve the dialogue between academia and developers in the industry.

In general, the majority of participants at conferences are university researchers, and professionals are the very minorities. Part of the problem is due to the conference/symposium program committee in the way they select papers and organise the program. IWTCS must serve as an agent for practititioners and researcher to interact. Its purpose should be to influence research directions, stimulate new approaches, promote collaboration, and generate interactive exchanges among researchers and practitioners. IWTCS should aim at being a meeting place for researchers who want to work with real problems and developers will have a hope that researchers will have something relevant to say.

3.3. Fitting new technologies into companies

The main goal of companies is to make profits. Productivity is an important factor that helps achieve this goal. Productivity is largely determined by the methods, tools and skills used in software development. A company, with the goal of making profit, always searches for new and better methods and tools to improve its productivity and competitiveness This is also a survival strategy. Only if substantial improvements to productivity can be guaranteed, a company will seriously consider new methods and will invest money and time to introduce them into practice. It will invest in technology, provided that it pays. Researchers need to demonstrate to companies how their new techniques can hence productivity. Academics need to show the benefits and demonstrate its success to the industry Successful uses of new techniques on large systems needs to be reported at conferences and to be published in journals.

To ensure success of business, the choice of technology is an important decision. From a manager's point of view, trying a new idea poses risk. Managers are only interested in new ideas if they do not pose significant risks, do not cost too much to develop, and can be introduced gradually. The old approach may not be good, but it is predictable. This is good from a business point of view. A new technique might be good but not necessary predictable. New techniques need not be perfect, but in order to be used industrially they must fit in with the organisation and the way it has been working.

3.4. Consumable mathematics

Unfortunately, the methods offered by academia so far have been mathematical, and there has been excessive amount of effort focussing on proofs of theories. There is nothing wrong with mathematics which is simply a tool that supports analysis. Engineering personnels like civil and electrical engineers use mathematics for modelling and solving problems; likewise, mathematical methods are a key to improving effectiveness and productivity in testing of communicating systems. However, it s worth noting that engineering is a more matured discipline than computer science. Moreover, testers are not mathematicians, and they do not have the same type of knowledge as academics have. It may take a tester much longer to learn the method than to do the tests, so that in practice it may be just as easy for a tester to make mistakes in the use of mathematics as in testing. This slows down the process of testing without increasing its quality. Researchers need to make mathematical parts more consumable for testers in order for the new techniques to be in wider use. There are different levels of mathematics required by different personnels. Mathematical details need to be hidden from testers when they use the new techniques.

3.5. Education

The gap between theory and practice is partly due to the fact that testers have not been educated in formal methods for testing. Ultimately, future design and testing of communicating systems using formal methods will be initiated by engineers and testers who have been educated in formal methods. Education will be a key factor influencing the adoption of formal methods by the industry. This educational gap could be filled by university education.

Tertiary courses that focussed on formal methods of development of communicating systems became available in the late 80's (Lai, 1992). Apart from teaching theories, such courses need to aim at helping students apply formal methods in real life complex situations. Ideally, students are required to pass a practical component on development of communicating systems. Academics needs to actively solicit involvement from companies in the set-up of a curriculum. Universities should set up industrial advisory task forces from which input to courses can be sought. At times, practitioners and developers could be invited to deliver a couple of guest lectures. Unfortunately, course on formal method and with input from the industry on the development and testing of communicating systems is not common enough and needs to be more widespread.

3.6. Engaging in real projects

One of the practical ways forward is to have researchers working closely with practitioners on real projects. They can be housed in the same building. Then researchers and practitioners can interact regularly, and researchers will then be more aware of the industrial issues. A typical research project can be carried out to experiment the new idea in parallel to the actual project. Researchers can then present results back to the practitioners who are working on real projects. This will take the pressure off the practitioners as most of the risks involved in using new technique are removed.

Both sides will benefit from this type of collaboration. University researchers are not usually rewarded for their ideas that are put into practice on real projects. They are rewarded for publications of their research results and for getting research grants. If an idea works well, the technique can be adopted. If it does not, researchers still can write a paper explaining what went wrong. However, if the idea works, researchers can write several papers. In the end, university researchers still have papers published and do not lose any thing. In fact, they will have more to gain : they will have the credentials to attract more research funding as their ideas have been demonstrated to work on real projects. As for the practitioners, they will appreciate how new techniques can be of practical use to their problems and how their productivities could be increased. The gap between theory and practice will then become narrower.

3.7. Better tools

Tools are not very frequently used in the industry, but they play an important role in the industry as they will be an effective means of reducing human effort and will cover more test cases. A number of test tools have been developed by academia (Grassl et al., 1990; Shiratori et al, 1989) but such tools are seldom used commercially. They are usually prototypes and tend to be small and unreliable as normally there is not enough man power involved in developing them. Often such a type of work is performed by postgraduate students as part or all of their theses. Could we trust small and cheap tools ? The general perception is that it can not be good if it is small and cheap. Low cost tools are generally useful for a small or one-person project. Most academic tools are not of industry strength and therefore can not be commercialised. They tend not to measure up to expectation, and cannot be extensively used in the industry. For a large scale project, an industry strength tool that can support a large problem is required. Industrial developers tend to favour large and expensive tools. For instance, a jumbo 747 aeroplane cannot be designed with a small tool. The size and feature of a tool need to be matched to the complexity of the project. Academia should seek the human resources and financial sponsorship of companies in developing industry strength tools.

4. SOME INDUSTRIALLY RELEVANT TOPICS

Some industrially relevant research topics are suggested in this section. These topics are by no means less challenging and of lower standard than those that have already been conducted by academia. As not much research results for these topics can be found in the present literature, university researchers can easilly published some papers on each of these areas if publication is still a concern for them. Should academia work on these topics, the gap between the theories and practices will become even narrower.

Resource constraint As the industry has to operate in a profitable manner, constraints on resources often hamper the effectiveness of testing (Genuchten, 1991). One can use coverage percentage as an indicator of when to move the testing resources onto another area.

Conflict resolution Internal politics are often involved : designers see testing as a threat, while test managers want to test the product thoroughly. Developers can get upset if lots of discrepancies are found. They perceive that the the test team is slowing them down. However, they should be delighted at each discrepancy found, as this allow them to improve the quality of the product. Conflict thus arises. Conflict resolution and negotiation techniques are critical to software success. There is not too much research that recognises these realities.

Completion criteria Test completion criteria are hard to define. Communication software is often tested until the probability of failure is believed to be small, or until the deadline for the product release is reached (Sherer, 1991). There is nothing wrong with financial and economic criteria as the industry has time and financial constraints. Perfection can not be a completion criterion as software can never be perfect and will always have bugs. The usual practice in the industry is that after releasing a product to the market, suppliers will fix the bugs reported during the warranty period. Knowing how much of the source codes has been covered by the test suite can help estimate the risk of releasing a piece of software to users and thus could form a basis for completion criterion.

Point of diminishing return Inexperienced testers tend to execute down the same path of the software, and no new information is thus yielded. Points of diminishing return need to be identified. There is not much research work on determining where best to concentrate the testing efforts and pointing out where additional testing would be fruitful.

Prioritising Since it is impossible to test every possible case and there is time constraint, priority has to be given to testing the most important and vulnerable parts of the communicating system. Prioritising is necessary, as it is not too fruitful to set identical test coverage goals for all components of an implementation and some components are more critical than others. There is a need to identify the high risk areas so that more testing effort is devoted to these areas and a higher test coverage goals is set for them.

Test case value Since the number of test cases is enormous, there needs to be an optimal way of designing a test suite. By assigning a value to a test case, the relative importance of test cases can be ranked and an optimal test suite can then be designed. The value of a test case needs to be measured in economic terms which could be based on the probability that a particular case will occur and the its consequence of disaster if it fails. Some failures could bring a much larger catastrophic effect than others. Again, there is not much research work on this area.

Code review and measurement Bad codes depend heavily on testing to uncover errors and good codes can only come from good methods. There are not much research on code review and measurement for testing purpose. Statistics need to be collected about the codes in order to get some hints as to what codes need to be improved and tested more rigorously. The gathering of statistics is the start for getting measurement in practical use (McCabe, 1976); the aim is to predict error-proneness by measurements. Conducting a measurement of the codes gives an objective method of locating the error prone codes. Designers and testers usually have a fairly good idea of where the problems might lie, but measurement results affirm their judgements and support their arguments.

Functional aspect Most research focus on the control part of a communicating system. More efforts need to be spent on its functional and robust aspects such as multiplexing, boundary cases, timing, invalid parameter, inopportune events and capacity tests. These types of tests are even more important, without which communication software will not be too reliable.

Empirical comparative Studies There are so many techniques published by academia. Which one should testers adopt ? Empirical comparison of different techniques are seldom conducted and will be useful. A forum must be provided for researchers and practitioners to report both original and replicated studies, varying from controlled experiments to field studies. This experience can be used to evaluate and reveal the relationship between testing theories and practice. Replicated studies, both successful and unsuccessful, highlighting what can be learned from them for improving future work, need to be encouraged for publication.

Test coverage tool The design of a test suite is a slow, labour intensive and costly exercise. The higher the coverage is, the higher is the cost; coverage and cost need to be compromised. The use of tools for test coverage measurement is essential for advancing test coverage methods. Test selection is something that cannot be completely automated. Test sequences generated are not usually related to test purposes and are therefore not cost effective. Human insight continues to be necessary for cost-effective testing. Human guidance and intuition are important to its success. Good test tools should provide an interface that enables a tester to guide the test sequence generation, display test coverage choices and compute the test coverage attained by a particular test suite. Such tools should include input like PICS (Protocol Implementation Conformance Statement), protocol specification, and test suite.

Cost Estimation From experience, cost of testing is very high. There are many contributors to this high cost. Some of them are number of lines of codes, number of modules, number of interfaces, number of requirements and number of function points. There have been some models established for estimating the cost for software development (Boehm, 1981; Putnam, 1978), but none has been established for testing a communicating system yet. It would be nice if there is a simple formula into which we could plug the values for the cost contributors and which would, as a result, provide an estimate of the testing cost.

Others The failures of the relevant application of much of the theoretical work for testing on communicating systems are due to the failure to acknowledge other factors that are critical to projects. Some examples are risk control, functionality, user confidence, predictability, and early deliverable of visible. These areas have been hardly researchered into yet.

5. CONCLUSIONS

This paper has presented the argument that academic research on testing of communicating systems needs to become more industrially relevant and describes the means by which this change can be facilitated. The primary reason for this need is that academia can no longer afford not to work on real problems due to the changes in societies and the ways research grants are awarded to academia. This paper has also suggested some research topics that are useful to the testing community. To help industrialisation of academic techniques, empirical study and a new orientation towards real issues need to be given higher priorities by researchers, and real-life test coverage and field experience need to be more frequently reported . An effective academic testing method applicable to the industry must deal with the practical problems encountered in industrial testing.

It is the wish of the author that there will be a change in direction of academic research on testing of communicating systems after this IWTCS. To help effect such a change, academia must interact with practitioners in the industry. There are massive benefits as well as the costs of having such an interaction. It requires an enormous amount of efforts by both sides, as well as the resources to support the efforts. On the other hand, developers must recognise that they have to change their work habits if they wish to exploit research results. This process is difficult and expensive and it takes up time. Unless both sides are willing to work together despite the cost, the chance of succeeding is low; and we cannot be blind to the difficulties and costs. In conclusion, we still have to do it as the costs of not doing it is even higher than the cost of doing it.

6. REFERENCES

Boehm, B., (1981) *Software Engineering Economics*, Prentice Hall, Englewood Cliffs, N.J.

CCITT (1987) Recommendation Z.100 : Specification and Description Language SDL, Com X-R15-E.

Genuchten, M. van, (1991) Why is software Late? An Empirical Study of Reasons For Delay in software Development, *IEEE Transaction on Software Engineering*, vol.17, no.6.

Grassl, W. and Kossmann, H. (1990) Automated Implementation and Testing of SDL Protocol Specification, *Proceedings of the 2nd International Workshop on Protocol Test Systems*, North-Holland.

ISO (1989a) Information Processing Systems - Open Systems Interconnection - ESTELLE - A Formal Description Technique Based on an Extended State Transition Model, ISO 9074.

ISO (1989b) Information Processing Systems - Open Systems Interconnection - LOTOS - A Formal Description Technique Based on the Temporal Ordering of Observational Behaviour, ISO 8807.

Lai, R. (1992) Teaching Protocol Engineering in Honours Year, *Lecture Notes in Computer Science*, Vol 640, Springer-Verlag.

Lai, R. and Leung W. (1995) Academic and Industrial Protocol Testing - The Gap and the Means of Convergence, *Computer Network and ISDN Systems*, 27(1995) 537-547.

McCabe, T.J. (1976) A Complexity Measure, *IEEE Transaction on Software Engineering*, 2(4):308-320.

Putnam, L., (1978) A general empirical solution to the Macro software sizing and Estimating Problem, *IEEE Transactions on Software Engineering*, Vol. SE-4, no. 4, 345-61.

Sherer, S.A. (1991) A Cost-Effective Approach to Testing, *IEEE Software*, March, 1991.

Shiratori, N., Takahashi K. and Noguchi, S. (1989) An Intelligent Use-Friendly Support System for Protocol and Communication Software Development, *Protocol Specification, Testing and Verification VIII*, North Holland.

Sidhu, D.P. and Leung, T.K. (1989) Formal Methods for Protocol Testing: A Detailed Study, *IEEE Transactions on Software Engineering*, vol.15, no.4.