Research Statement

Gheith A. Abandah
abandah@ju.edu.jo
June 25, 2015

I am interested in three areas of research. I am interested in improving application performance on parallel systems. I am also interested in providing solutions to problems related to the Arabic language. And I am interested in developing secure electronic voting systems that improve participation in the political life.

1. Past and Current Research

1.1 Parallel Processing Research

I have started research as a graduate student in the University of Michigan in 1994 by studying and modeling the computation performance of superscalar pipelined processors [1, 2]. This was within my research group, the Parallel Performance Project. In this group, we were trying to port and improve the performance of scientific and engineering applications on vector supercomputers and scalable multi-computers and multiprocessors. One of my first assignments was to port Ford Motor's car crash simulation code into the message passing multi-computer IBM SP2. In order to improve the performance of this application, I developed microbenchmarking techniques to evaluate and model the communication performance carried out through the exchange of explicit messages [1, 2, 3].

The group then got interested in porting the car crash code to scalable shared-memory multiprocessors. We worked on large parallel systems such as the cache-only memory architecture (COMA) Kendall Square Research's KSR-1 and KSR-2 machines and the scalable shared-memory multiprocessor Convex SPP-1000. Again to improve the performance on such machines, I developed microbenchmarking techniques to evaluate and model the implicit communication performance carried out through accessing shared memory locations [4, 5].

In 1996, I took a one-year research internship in Hewlett-Packard Labs in Palo Alto, CA with a large group studying design alternatives for building the next generation of scalable shared-memory systems. This group was interested in my evaluation and modeling work and invited me to join them. My main research focus with them was characterizing applications for scalable shared-memory multiprocessors. For this objective, I developed a set of tools to instrument such applications and analyze their characteristics that most affect their performance on shared-memory multiprocessors [6]. These tools analyze application memory traces on-the-fly and were successfully used, as a first step, to characterize the NAS parallel benchmarks [7]. These tools use a novel approach that allows performing configuration independent and configuration-dependent analyses [8, 9]. The configuration independent analysis allows characterizing the inherent characteristics of the application. Whereas, the configuration dependent analysis allows characterizing the application depending on the machine configuration it is executed on. The machine configuration aspects most important in this analysis are the memory and interconnection network configurations.

The tools and methodologies I developed in HP Labs proved to be very useful when I returned back to the University of Michigan to resume my PhD program. My PhD research focus was on reducing

communication cost in scalable shared memory systems [10]. In this work, I did an evaluation of three important cache-coherent nonuniform memory access systems: Stanford DASH, Convex SPP1000, and SGI Origin 2000 [11, 12]. I also studied the effects of architectural and technological advances on the memory and communication performance of two generations of the HP/Convex Exemplar's, the SPP1000 and SPP2000 [13]. Finally, I proposed design enhancements of example scalable shared-memory multiprocessor to better serve communication intensive applications [14].

After getting my PhD degree, I spent the Summer of 1998 in HP Labs. At that time, the group was interested in improving the performance of commercial applications on scalable shared-memory multiprocessors and we worked on characterizing transaction processing applications such as TPC-C and decision support applications such as TPC-D.

1.2 Arabic Language Technologies Research

When I joined the University of Jordan in Fall 1998, I found that it is difficult to continue research on parallel processing due to the lack of supercomputing facilities. Moreover, I got interested in solving problems related to the Arabic language. We did some work concerning the coding of Arabic letters and proposed new code system that matches the Arabic alphabet and does not have the problems found in other codes [15]. We also used genetic algorithm to find efficient Arabic keyboard layout [16]. However, the largest problem related to the Arabic language I work on is the optical character recognition of Arabic text. I had three funded research projects related to this problem. The first project was an initial study on printed and handwritten Arabic optical character recognition where we surveyed this area and collected a database of handwritten samples [17]. We also explored some techniques to recognize handwritten Arabic characters [18, 19].

I then got very interest in solving a hard research problem, the recognition of handwritten Arabic text. First, we concentrated on developing efficient feature extraction techniques suitable for this problem [20, 21]. We also used some feature extraction techniques to analyze handwritten Arabic letters in order to get insights that help in developing efficient recognition systems [22]. In order to select efficient subsets of features, we investigated the utility of several feature selection techniques [23, 24]. At the end of this phase, we got deep understanding of what feature sets give excellent recognition accuracy.

I then concentrated on developing a complete recognition system for handwritten words. To this end, we first developed efficient skeleton-based segmentation algorithm to segment words into their grapheme components [25]. Combining efficient feature extraction techniques with this segmentation algorithm and a recurrent neural network recognition engine, we got a system that achieves better accuracy than other leading systems on some benchmark test sets [26]. We used a design of experiments approach to efficiently tune the recurrent neural network used for best accuracy [27]. We also developed a word matching algorithm that uses multiple-sequence weighted edit distances [28].

I found that recurrent neural networks are very useful to solve other problems. Their sequence transcription capabilities are very effective in adding diacritics to Arabic text. The modern standard Arabic is usually written without diacritics; but diacritics are needed to remove many pronunciation ambiguities. The approach we are developing gives significant accuracy improvement over best published results [29]. We use deep bidirectional long short-term memory (LSTM) network to achieve state-of the art accuracy in diacritizing Arabic text.

We have also recently conducted a study to assess the status of Arabic language on the social networking forums and mobile phone communications [30]. We gathered many Arabic text samples from facebook, twitter, news site comments, blogging sites, and short messaging. We analyzed many language attributes such as language type, characters used, language level, and error rates and types.

1.3 Electronic Voting Research

I also have another funded project to develop secure national electronic voting system. We developed a system that satisfies the important requirements of an e-voting system through state-of-the-art technologies and secure processes. This system relies on homomorphic cryptography, zero-knowledge proofs, biometrics, smartcards, open-source software, and secure computers for securely and efficiently implementing the system processes over the various stages of the electoral process, without relying on online network connections [31]. I think that this system has commercial and practical feasibility and I am interested in its full implementation.

2. Research Plans

I have started to resume my graduate research work on parallel processing. I think that the tools and approaches I have developed for scalable multiprocessors can be modified for multicore processors. I plan to extend my previous work to enable analyzing and improving multicore applications and to explore cache, memory, and interconnect design ideas for multicore systems. I have recently modified the configuration-independent tool to work with the Pin instrumentation tool on Linux operating systems for SPLASH and PARSEC multicore application benchmarks.

I also have started to extend the handwritten Arabic word recognition system to handwritten Arabic document recognition. I plan to develop appropriate line and word segmentation algorithms. Moreover, the current approach needs to be extended from limited vocabulary recognition to unlimited vocabulary. I have recently done some work to segment handwritten Arabic documents of the MADCAT training set.

Finally, our cryptographic and verification algorithms used in the electronic voting system are computational intensive. The tallying process in a large election takes many hours. I am now developing parallel algorithms to reduce the tallying time to a reasonable time.

References

1. E. Boyd, **G. Abandah**, H. Lee, and E. Davidson, “Modeling Computation and Communication Performance of Parallel Scientific Applications: A Case Study of the IBM SP2,” Tech. Rep. CSE-TR-236-95, University of Michigan, May 1995.
2. **G. Abandah** and E. Davidson, “Modeling the Communication and Computation Performance of the IBM SP2,” Tech. Rep. CSE-TR-258-95, University of Michigan, May 1995.
3. **G. Abandah** and E. Davidson, “Modeling the Communication Performance of the IBM SP2,” in Proc. 10th Int'l Parallel Processing Symposium (IPPS'96), Apr 15-19, 1996, pp 249-257, Honolulu, Hawaii, USA.
4. **G. Abandah** and E. Davidson, “Characterizing Shared Memory and Communication Performance: A Case Study of the Convex SPP-1000,” Tech. Rep. CSE-TR-277-96, University of Michigan, Jan 1996.
5. **G. Abandah** and E. Davidson, “Characterizing Distributed Shared Memory Performance: A Case Study of the Convex SPP1000,” IEEE Trans. Parallel and Distributed Systems, Vol. 9, No. 2, Feb 1998, pp 206-216.
6. **G. Abandah**, “Tools for Characterizing Distributed Shared Memory Applications,” Tech. Rep. HPL-96-157, HP Labs, Dec 1996.
7. **G. Abandah**, “Characterizing Shared-Memory Applications: A Case Study of the NAS Parallel Benchmarks,” Tech. Rep. HPL-97-24, HP Labs, Jan 1997.

8. **G. Abandah** and E. Davidson, "Configuration Independent Analysis for Characterizing Shared-Memory Applications," Tech. Rep. CSE-TR-357-98, University of Michigan, Jan 1998.
9. **G. Abandah** and E. Davidson, "Configuration Independent Analysis for Characterizing Shared-Memory Applications," in Proc. 12th Int'l Parallel Processing Symposium & 9th Symp. On Parallel and Distributed Processing (1998 IPPS/SPDP), Mar 30-Apr 3, 1998, pp 485-491, Orlando, Florida, USA.
10. **G. Abandah**, "Reducing Communication Cost in Scalable Shared Memory Systems," PhD Dissertation, University of Michigan, 1998.
11. **G. Abandah** and E. Davidson, "A Comparative Study of Cache-Coherent Nonuniform Memory Access Systems," High Performance Computing Systems and Applications, Ed. J. Schaeffer, Kluwer Academic Publishers, pp 267-282, 1998.
12. **G. Abandah** and E. Davidson, "A Comparative Study of Cache-Coherent Nonuniform Memory Access Systems," in 12th Ann. Int'l Symp. High Performance Computing Systems and Applications (HPCS'98), May 20-22, pp 27, 1998, Edmonton, Canada.
13. **G. Abandah** and E. Davidson, "Effects of Architectural and Technological Advances on the HP/Convex Exemplar's Memory and Communication Performance," in Proc. 25th Ann. Int'l Symp. Computer Architecture (ISCA'98), Jun 27-Jul 1, 1998, pp 318-329, Barcelona, Spain.
14. **G. Abandah** and E. Davidson, "Origin 2000 Design Enhancements for Communication Intensive Applications," in Proc. 1998 Int'l Conf. on Parallel Architecture and Compilation Techniques (PACT'98), Oct 12-18, 1998, pp 30-39, Paris, France.
15. **G. Abandah** and F. Khundakjie, "Issues Concerning Code System for Arabic Letters," Dirasat Engineering Sciences Journal, Vol. 31, No. 1, Apr 2004, pp 165-177.
16. T. Malas, S. Taifour, and **G. Abandah**, "Toward optimal Arabic keyboard layout using genetic algorithm," in Proc. 9th Int'l Middle Eastern Simulation Multiconference (MESM 2008), Aug 26-28, 2008, pp. 50-54, Amman, Jordan.
17. **G. Abandah** and M. Khedher, "Printed and Handwritten Arabic Optical Character Recognition – Initial Study," A report on research supported by The Higher Council of Science and Technology, Amman, Jordan, Aug 2004.
18. M. Khedher and **G. Abandah**, "Arabic Character Recognition using Approximate Stroke Sequence," Third Int'l Conf. on Language Resources and Evaluation (LREC 2002), Arabic Language Resources and Evaluation - Status and Prospects Workshop, Jun 1, 2002, Las Palmas, Spain.
19. Mohammed Olama, Muhannad Aulama, Asem Natsheh, and **G. Abandah**, "Arabic handwritten optical character recognition using hidden Markov models," in Proc. 2011 SPIE Defense, Security, and Sensing Multi-conference, Optical Pattern Recognition XXII, Vol. 8055, Apr 28-29, 2011, Orlando, Florida.
20. **G. Abandah**, K. Younis, and M. Khedher, "Handwritten Arabic Character Recognition Using Multiple Classifiers Based on Letter Form," in Proc. 5th IASTED Int'l Conf. on Signal Processing, Pattern Recognition, and Applications (SPPRA 2008), Feb 13-15, 2008, pp 128-133, Innsbruck, Austria.
21. **G. Abandah** and N. Anssari, "Novel Moment Features Extraction for Recognizing Handwritten Arabic Letters," J. of Computer Science, Vol. 5, No. 3, Mar 2009, pp. 226-232.
22. **G. Abandah** and M. Khedher, "Analysis of Handwritten Arabic Letters Using Selected Feature Extraction Techniques," Int'l J. of Computer Processing of Languages (IJCPOL), Vol. 22, No. 1, Mar 2009, pp. 49-73.
23. M. Khedher, **G. Abandah**, and A. Al-Khawaldeh, "Optimizing Feature Selection for Recognizing Handwritten Arabic Characters," in Proc. 2nd World Enformatika Congress (WEC'05), Feb 25-27, 2005, pp 81-84, Istanbul, Turkey.
24. **G. Abandah** and T. Malas, "Feature Selection for Recognizing Handwritten Arabic Letters," Dirasat Engineering Sciences Journal, Vol. 37, No. 2, Oct 2010, pp. 242-256.
25. **G. Abandah** and F. Jamour, "Recognizing Handwritten Arabic Script through Efficient Skeleton-Based Grapheme Segmentation Algorithm," in 10th Int'l Conf. on Intelligent Systems Design and Applications (ISDA'10), Nov 29-Dec 1, 2010, Cairo, Egypt, pp. 977-982.
26. **G. Abandah**, F. Jamour, and E. Qaralleh, "Recognizing handwritten Arabic words using grapheme segmentation and recurrent neural networks," Int'l Journal on Document Analysis and Recognition, Springer, Vol. 17, No. 3, Sep 2014, pp. 275-291.
27. E. Qaralleh, **G. Abandah**, and F. Jamour, "Tuning recurrent neural networks for recognizing handwritten Arabic words," Journal of Software Engineering and Applications, Scientific Research, Vol. 6, No. 10, Oct 2013, pp. 533-542.

28. **G. Abandah** and F. Jamour, "A word matching algorithm in handwritten Arabic recognition using multiple-sequence weighted edit distances," International Journal of Computer Science Issues (IJCSI), SoftwareFirst, Vol. 11, No. 3, May 2014, pp. 18-26.
29. **G. Abandah**, A. Graves, B. Shagoor, A. Arabiyat, F. Jamour, and M. Al-Taee, "Automatic diacritization of Arabic text using recurrent neural networks," ," Int'l Journal on Document Analysis and Recognition (IJDAR), Springer, Vol. 18, No. 2, Jun 2015, pp. 183-197, DOI: 10.1007/s10032-015-0242-2.
30. M. Khedher, W. Anati, **G. Abandah**, A. Alzgoul, S. Ababneh, and M. Hattab, "Arabic Language in the Social Networking Forums and Mobile Phone Communications", the Jordan Academy of Arabic, Sep 2014.
31. **G. Abandah**, K. Darabkh, T. Ammari, and O. Qunsul, "Secure national electronic voting system," Journal of Information Science and Engineering, The Institute of Information Science, Vol. 30, No. 5, pp. 1339-1364, Sep 2014.