CENIIT research project 00-11
Fault Tolerance in Real-time Distributed Systems

Simin Nadjm-Tehrani
Dept. of Computer and Information Systems

The above project proposal was originally formulated around development of an incremental proof management system to support an incremental design methodology for analysis of system fault tolerance. The work was expected to be valuable in development of safety-critical systems and in particular aerospace systems. As a result of early feedback by other industrial members of the CENIIT board, the project was later reformulated to cover availability and thereby analysis of fault tolerance in networked systems. This complete redirection of the topic should perhaps be seen as the single most benefit of the exposure to the CENIIT context, from which a whole group of new research activities were spawned in the years to come.

1 Results during 00-05

Today's systems are to a large extent built as networks in which critical services are to be guaranteed. In recent years I have been facinated by the many threats to service availability in networked systems. Many of my papers treat threats to availability in terms of overloads, failures or attacks. The developed methods are evaluated using realistic scenarios from telecommunications and IP networks, or tested in specially developed simulated environments with realistic models of traffic patterns. I have also managed to pursue my original interest in safety-critical systems thereby obtaining broad expertise in dependable systems. I will present my research-related publications during 00-05 that directly relate to the above topic areas at the end of this report and comment below.

Server crashes and network partitions

Papers 1-6 study support for treatment of server failures using automatic failover mechanisms in middleware. The problem is studied both empirically and theoretically, using queueing theory. In the empirical studies a service in a telecom operations and management network was used on top of a standard CORBA platform that we extended with support for fault tolerance. In theoretical studies, optimised checkpointing interval for primary servers was found in order to minimise average response time or maximise average availability. Paper 4 combines the benefit of a middleware-based approach for automatic treatment of failover, with the use of application writers knowledge to minimise the overhead (wait times for operations that have to be logged and potentially replayed upon failures).

Paper 7 studies recovery from network partitions, in particular, middleware support for higher availability in distributed object systems, and the resulting trade-off against data consistency and performance. As the network partitions, the primaries in various partitions continue to serve the arriving requests from clients. Upon reunification of the network, automatic reconciliation of state for primary objects is carried out in the middleware. This activity has been carried out in the context of a European project in which air traffic control is the main application.

Overloads

Overloads can be optimally treated by using admission control and service differentiation. The users of a service then experience a trade-off in terms of delivered quality of service (QoS). Papers 8-13 provide optimised methods for differentiated resource allocation in future generation cellular, wireless adhoc, and hybrid networks. The main concept to enable adaptation of the system in presence of overloads is a resource-utility function. Heuristics based on convex hull approximations are compared to optimal (linear programming) solutions to the problem with surprisingsly good results. Paper 12 combines routing and resource allocation in ad hoc networks using shadow prices. Paper 13 compares the convergence and performance properties of our algorithm in comparison to another prominent group's algorithm and shows superior behaviour in presence dynamic load changes and mobility.

Attacks and intrusions

Papers 14-18 study survivability of critical infrastructures and propose an agent architecture for safeguarding them. In a survivable network, parts of the system suffer crashes, successful intrusions and overloads, but the use of timely detection, response and recovery mechanisms allows that critical services are provided at all times. Paper 14 shows how the Swarm simulation environment can be used to characterise the trade-off between integrity, availability and response time in models of survivable networks. Papers 16 and 17 propose adaptive real-time anomaly detection algorithms based on clustering techniques. This work has been successfully evaluated on a test network developed at the Swiss telecom operator Swisscom, and shown to have good performance and scalability properties. The published results on accuracy of the detection results, and the relatively low false alarm rates have resulted in recent cooperation with university of Cincinnati on a project in a completely different domain: recognition of anomalies due to water contamination, commissioned by the Homeland security.

Fault tolerance and timeliness in component-based systems

In 2003 a project granted by SSF enabled going back to the original formulation of the CENIIT project: supporting upgrades in safety-critical systems built from components (papers 21-26).

In safety-critical systems fault tolerance is often analysed at an application level (unlike telecom and IP-based systems in which fault-related functions can be embedded in agent platforms and middleware). Tools like Fault Tree Analysis (FTA) and Failure Modes and Event Analysis (FMEA)

have their roots in mechanical engineering, and do not support managing the combinatorial complexity of faults affecting digital systems. Our work supports automatic and formal analysis of digital systems, in presence of faults, and the integration of this analysis in the system development process in a model-based paradigm. This problem was interesting for collaborators at Saab Aerospace and was cast in a context of model-based development of software-like hardware (reconfigurable components). Design languages that assist both automatic code generation and provide tool support for formal verification were selected to illustrate our ideas on a "safety analysis pattern". Analysis of realistic fault modes for the hydraulic leakage detection system of JAS 39 Gripen aircraft was used as a proof of concept and presented in papers 21 and 22.

Paper 25 further develops the analysis pattern in the context of *component-based systems*. A component model is presented where specific interfaces with formal semantics are used to capture sensitivity of the component to single or multiple faults. A method for compositional reasoning about system level safety property of assemblies, based on component interfaces is proposed. Cooperation with Ed Clarke's group at Carnegie Mellon university has just started based on this work. Paper 26 constructs a similar formal interface for components in which *timing properties* are of prime concern. A model based on timed automata is used to formalise reconfigurable components and their modification using *aspects* together with a weaving operation.

Papers 27 and 28 take the analysis of fault tolerance for safety-critical systems further by considering special issues arising in distributed control systems. These are illustrated by studying a distributed flight control system (FCS) as a future development of the current centralised FCS in JAS 39 Gripen.

2 Summary of Graduate Degrees awarded

The students in my group with a partial support by CENIIT have all started in 2000 or later years. Among these, Diana Szentivanyi and Calin Curescu obtained Licentiate degrees in year 2002 and 2003 respectively, and PhD degrees in year 2005. The date of defence for Kalle Burbeck's Licentiate thesis has been set in February 2006. Aleksandra Tesanovic whom I was formally an advisor for until 2005 completed her Licentiate thesis in 2003.

Jonas Elmqvist and Mikael Asplund are planned to obtain Licentiate theses in 2006 and 2007 respectively.

3 Summary of Masters thesis

During the years 2000-2005, 28 students completed their Masters thesis or final year projects under my supervision. These were: Christofer Hallen, Åsa Karlsson, Markus Nilsson, Bengt Bergman, Li Cai, Henrik Leion, Per Andersson, Anders Grahn, Andreas Eriksson, Jonas Elmqvist, Johan Ydren, Daniel Garpe, Tobias Chyssler, Kenth Fransson, Fredrik Ruben, Jerker Hammarberg, Robert Jonasson, Sara Garcia Andres, Tomas Lingvall, Pia Johansson, Wang Wei, Marison Escalada Malibran, Elvira Svensson, Jan

Bäckström, Håkan Oswaldsson, Tomas Berntsson, Mikael Hult, Johan Liljegren.

4 Researchers supported by the grant

Over the years the following researchers (and students) have been partially supported by the grant: Simin Nadjm-Tehrani (advisor), Diana Szentivanyi, Kalle Burbeck, Jonas Elmqvist, Jerker Hammarberg. Several Masters students have obtained supervision and contributed to research papers under the CENIIT grant.

5 Industrial Contacts

Contacts at Ericsson Radio Systems have been initiated and strengthened within the project (Torbjörn Örtengren, Johan Moe, Pär Emanuelsson, Mikael Patel, Fredrik Gunnarsson) as well contacts within Ericsson Finland (Andras Vajda). During this period, however, the potential for research at Ericsson radically changed due to economic conditions.

Lars Holmlund, Jan-Erik Ericsson, Hans Sjöblom, and Rikard Johansson, among others, at Saab Aerospace, have made valuable contributions to our research as well as Kristina Forsberg (Saab Tech). Contacts at European companies have contributed to the development of our work: Stefan Buschka, Thomas Dagonnier, Michael Semling at Swisscom, and several discussion partners at AIA (Spain) and Frequentis (Austria).

6 Contacts with other CINIIT projects

Some contacts with the group of Pär Värbrand were made during the CENIIT years, but did not result in concrete cooperation. Pär was on the thesis committee of one thesis in my group in 2005. We established contacts with John Noble in the Mathematics department and with Teresa Dahlberg (University of North Carolina) as well as Rachid Guerraoui (EPFL) and Marius Minea (university of Timisoara) leading to common papers. Several other cooperations existed due to participation in EU projects.

7 Formation of groups

The CENIIT grant is of course not an enabler for formation of groups, but has had a definite role in evening out funding from other sources during the years 00-05. During 2005 I was the advisor for 6 PhD students.

References

[1] Diana Szentivanyi and Simin Nadjm-Tehrani. Building and Evaluating a Fault-tolerant CORBA Infrastructure. In *Proceedings of the Workshop on Dependable Middleware-Based Systems at the International Conference on Dependable Systems and Networks (DSN02)*, pages G31 – G38. IEEE, June 2002.

[2] Diana Szentivanyi, Isabelle Ravot, Simin Nadjm-Tehrani, and Rachid Guerraoui. Dependable Distributed Middleware: Pay Now or Pay Later!. In *Poster Session at the ACM/IFIP/USENIX International Middleware Conference, Middleware 2003 Companion*, ACM/IFIP/USENIX, June 2003.

[3] Diana Szentivanyi and Simin Nadjm-Tehrani. *Middleware Support for Fault Tolerance*. Book chapter in M. Quasay (Eds.) *Middleware for Communications*, John Wiley and Sons. June 2004.

[4] Diana Szentivanyi and Simin Nadjm-Tehrani. Aspects for Improvement of Performance in Fault-tolerant Software. In *Proceedings of the 10th Pacific Rim Dependable Computing Conference (PRDC04)*, pages 283–291. IEEE, March 2004.

[5] Diana Szentivanyi, Simin Nadjm-Tehrani and John Noble. *Configuring Fault-tolerant Servers for Best Performance*. In *Proceedings of the First International Workshop on High Availability of Distributed Systems (HADIS05)*. IEEE Computer Society, August 2005.

[6] Diana Szentivanyi, Simin Nadjm-Tehrani and John Noble. *Optimal Choice of Checkpointing Interval for High Availability.*. In *Proceedings of 11th Pacific Rim Dependable Computing Conference (PRDC05)*. IEEE Computer Society, December 2005.

[7] Mikael Asplund and Simin Nadjm-Tehrani. *Post-Partition Reconciliation Protocols for Maintaining Consistency.*. In *Proceedings of the 21st ACM/SIGAPP symposium on Applied computing (SAC06), Dependable Distributed Systems Track*. ACM, April 2006.

[8] S. Nadjm-Tehrani, K. Najarian, C. Curescu, T. Lingvall, and T. A. Dahlberg. Adaptive Load Control Algorithms for 3rd Generation Mobile Networks. In *Proceedings of the 5th ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWIM02)*, pages 104–111, ACM, September 2002.

[9] C. Curescu and S. Nadjm-Tehrani. Time-aware Utility-based QoS Optimisation. In *Proceedings of the 15th Euromicro Conference on Real-time Systems (ECRTS03)*, pages 83–92, IEEE, July 2003.

[10] C. Curescu and S. Nadjm-Tehrani. Time-aware Utility-based Resource Allocation in Wireless Networks. *IEEE Transactions on Parallel and Distributed Systems*, July 2005.

[11] C. Curescu, S. Nadjm-Tehrani, B. Cao and T. A. Dahlberg. Utility-based Adaptive Resource Allocation in Hybrid Wireless Networks. In *Proceedings of the Second International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (Qshine05)*, IEEE, August 2005.

[12] C. Curescu and S. Nadjm-Tehrani. Price/Utility-based Optimized Resource Allocation in Wireless Ad hoc Networks. In *Proceedings of the 2nd Conference on Sensor and Ad Hoc Communications and Networks (SECON05)*, IEEE, September 2005.

[13] M. Leuthi, S. Nadjm-Tehrani and C. Curescu. Comparative Study of Price-based Resource Allocation Algorithms for Ad Hoc Networks. In *Proceedings of the 11th International Parallel and Distributed Systems Symposium (IPDPS)*, IEEE, April 2006.

[14] K. Burbeck, S. Garcia Andres, S. Nadjm-Tehrani, M. Semling, and T. Dagonnier. Time as a Metric for Defence in Survivable Networks. In *Proceedings of the Work-in-Progress session of 24th IEEE Real-Time Systems Symposium (RTSS03)*, December 2003.

[15] T. Chyssler, S. Nadjm-Tehrani, S. Burschka, and K. Burbeck. Alarm Reduction and Correlation in Defence of IP Networks. In *Proceedings of the 13th International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE04)*, pages 229–234. IEEE Computer Society, June 2004.

[16] K. Burbeck and S. Nadjm-Tehrani. Adwice: Anomaly Detection with Real-time Incremental Clustering. In *Proceedings of 7th International Conference on Information Security and Cryptology (ICISC04)*. Springer Verlag, December 2004.

[17] K. Burbeck and S. Nadjm-Tehrani. Adaptive Real-time Anomaly Detection with Improved Index and Ability to Forget. In *Proceedings of the Workshop on Security in Distributed Computing Systems (SDCS) at the 25th International Conference on Distributed Computing Systems (ICDCS05)*, IEEE Computer Society, June 2005.

[18] D. Gamez, S. Nadjm-Tehrani, J. Bigham, C. Balducelli, T. Chyssler, and K. Burbeck. *Safeguarding Critical Infrastructures*. Book chapter in H. B. Diab, A.Y. Zomaya (Eds.) *Dependable Computing Systems: Paradigms, Performance Issues and Applications*. John Wiley and Sons, Inc., November 2005.

[19] S. Nadjm-Tehrani. Formal Methods for Analysis of Heterogeneous Models of Embedded Systems. In *proceedings of International Symposium on Computer-Aided Control Systems Design (CACSD00)*, IEEE, September 2000.

[20] S. Tudoret, S. Nadjm-Tehrani, A. Benveniste, and J.-E. Strömberg. Co-simulation of Hybrid Systems: Signal-Simulink,. In *proceedings of the 6th International Conference on Formal Techniques in Real-Time and Fault-Tolerant Systems, (FTRTFT00), LNCS 1926*, pages 134–151. Springer Verlag, September 2000.

[21] J. Hammarberg and S. Nadjm-Tehrani. Development of Safety-critical Reconfigurable Hardware with Esterel. In *Proceedings of the 8th International Workshop on Formal Methods for Industrial Critical Systems (FMICS03)*, Norway, 2003.

[22] J. Hammarberg and S. Nadjm-Tehrani. Formal Verification of Fault Tolerance in Safety-critical Configurable Modules. *International Journal of Software Tools for Technology Transfer, Springer Verlag*, December 2004.

[23] J. Elmqvist and S. Nadjm-Tehrani. Intents, Upgrades and Assurance in Model-based Development. In *2nd Real-Time and Embedded Technology and Applications Symposium (RTAS) Workshop on Model-Driven Embedded Systems*, May 2004.

[24] J. Elmqvist and S. Nadjm-Tehrani. Intents and Upgrades in Component-based High-assurance Systems. In S. Beyeda, M. Book, and V. Gruhn (Eds.) *Model-driven Software Development, Volume II of Research and Practice in Software Engineering.* Springer Verlag, August 2005.

[25] J. Elmqvist, S. Nadjm-Tehrani, and M. Minea. Safety Interfaces for Component-based Systems. In *Proceedings of the 24th International Conference on Computer Safety, Reliability and Security (SAFECOMP05)*, LNCS, Springer Verlag, September 2005.

[26] A. Tesanovic, S. Nadjm-Tehrani, and J. Hansson. Modular Verification of Reconfigurable Components. Book chapter in C. Atkinson, C. Bunse, H-G. Gross, and C. Peper (Eds.), *Component-Based Software Development for Embedded Systems - An Overview on Current Research Trends*. Springer Verlag, LNCS 3778, November 2005.

[27] K. Forsberg, S. Nadjm-Tehrani, and J. Torin. Fault Analysis of a Distributed Flight Control System. In *Fault-Tolerant and Dependable Distributed Systems Minitrack of the Software Technology Track 38th IEEE Hawaii International Conference on System Sciences(HICSS05)*, ACM, January 2005.

[28] K. Forsberg, S. Nadjm-Tehrani, J. Torin, and R. Johansson. Maintaining Consistency among Distributed Control Nodes. In *Proceedings of the 23rd Digital Avionics Systems Conference (DASC04)*. IEEE, October 2004.

[29] J. Bäckström and S. Nadjm-Tehrani. Design of a Contact Service in a Jini-based Spontaneous Network. In *proceedings of ITCOM, Java-Jini technologies track*. SPIE, 2001.

[30] K. Burbeck, D. Garpe, and S. Nadjm-Tehrani. Scale-up and performance studies of three agent platforms. In *Proceedings of International Performance (IPCCC04), Communication and Computing Conference, Middleware Performance workshop*, pages 857–863. IEEE Computer Society, April 2004.