Disaster Tolerant Computer and Communication Systems

Stephen A. SZYGENDA
Department of Engineering Management, Information and Systems, SMU
Department of Computer Science and Engineering, SMU
Dallas, Texas, 75275, U.S.A.

and

Mitchell A. THORNTON
Department of Computer Science and Engineering, SMU
Department of Electrical Engineering, SMU
Dallas, Texas, 75275, U.S.A.

ABSTRACT

Disaster Tolerance is the characteristic attributed to a system that can withstand a catastrophic failure and still function with some degree of normality. Disaster tolerance of computer and communication systems is described and methods for modeling this form of system robustness are described. Definitions and descriptions of disaster tolerant computing and communications systems are provided and related to more familiar forms of system robustness such as fault tolerance. This paper concludes with a description of future areas of investigation for this new area of systems engineering.

Keywords: Disaster Tolerance, Catastrophic System Failure, Reliable Computing and Communications Systems

1. INTRODUCTION

Modern computing and communications systems are deeply integrated into the daily function of government, commercial, and private entities. Our society has heavy reliance on these systems for critical life support functions as well as business and other commercial reasons. This heavy dependence motivates us to investigate a new area that we term “disaster tolerant computing and communications systems”. This area involves both the design of new disaster tolerant systems and the modification of existing systems for disaster tolerance. We note that the term ‘disaster tolerance’ has been used in the past, particularly with respect to business data protection and availability. Our intention is to study disaster tolerance at the systems level, thus, data storage robustness is included as well as the disaster tolerance of underlying hardware, software, and interconnecting communications channels that form the overall system.

Disaster tolerance in computing and communications systems refers to the ability of such systems to maintain a degree of functionality after a disaster has occurred.

Definition: *A **disaster** is an event that can cause a system-wide malfunction as a result of one or more failures within a system. Disasters may occur due to a single-point failure or by a plurality of single-point failures that occur either simultaneously, or nearly simultaneously in a temporal sense and may be caused from either a man-made or natural event.*

Definition: *A **catastrophe** can occur as the result of the occurrence of a disaster. Catastrophes may be avoided by using disaster avoidance mechanisms.*

We differentiate between the terms “fault tolerance” and “disaster tolerance”. Fault tolerant system design has been studied for the last few decades and is usually based upon a single point of failure in a system. Typical strategies to provide fault tolerance at the physical level include the incorporation of redundant system components and a voting mechanism [1]. Other methods include the use of error-checking and correcting methods, hot-swap support, and special “watchdog” types of software. Disaster tolerance is a superset of fault tolerance in that a disaster may be caused by multiple points of failure in a system that occur very close together in time as well as a single point of failure that escalates into a wide catastrophic system failure.

In this work, we are most interested in very large systems such as a large distributed computing network that would preclude the use of physical component redundancy due to prohibitive cost. As an example, a single network router may have a degree of fault tolerance by the incorporation of redundant output line drivers; however, a data network that is distributed over a large geographical area may be impossible to replicate fully and thus redundancy is not a valid method for incorporation of disaster tolerance. Even if such a geographically wide-spread system were replicated, communications channels for voting would also be required and would also require redundancy in order to achieve any type of disaster tolerance.

We differentiate between disaster tolerant systems and disaster recovery systems. Disaster recovery is the ability to resume normal operations after a disaster has occurred while disaster tolerance is the ability to continue operations in an uninterrupted manner despite the presence of a disaster. This implies that the main difference between disaster recovery versus tolerance is one measured in terms of the delay that occurs after a disaster before normal operations resume. Whether a typical implementation is a disaster recovery or tolerance method depends upon the application.

For example, the ability of a datacenter to provide information within 5 minutes of a disaster may be a tolerant feature if this delay does not affect normal operations, alternatively, if the data is needed in real-time, this approach would be classified as a disaster recovery mechanism.

2. BASIS FOR SYSTEM MODELS

We believe that results from the areas of non-linear system analysis such as chaos and catastrophe theory may form the basis for the development of a model for large computing and communications systems that are subject to failure from disasters. Non-linear models are often used for large systems with variables that affect overall behavior in varying non-disproportionate ways. Such systems can typically be modeled with feedback loops. Critical points of failure can be modeled as singularities that occur in an otherwise monotonically smooth function.

The basis of catastrophe theory is to model a system using a normally smooth transfer function and to observe abrupt changes that arise as a sudden response due to an anomaly as compared to an otherwise normal and smooth change. When such abrupt changes occur due to a change in external conditions, a “disaster” is said to have occurred. From a mathematical point of view, the theory of singularities of smooth mappings allows for the development of a rigorous theory of catastrophes [2,3,4]. While this is an abstract point of view, we believe that investigation into this area may lead researchers to develop more sophisticated models of disaster in computing and communications systems. We believe that the occurrence of such singularities in computing and communications systems obey our definition of a disaster and that further system failures resulting from such a disaster can be modeled as a “catastrophe” as defined previously.

We recognize that the use of catastrophe theory in mathematics may not produce workable models for disaster tolerance. Nevertheless, we propose that this be a starting point for investigation for the modeling of disasters. Once a disaster is successfully modeled, the means for the incorporation of robust design techniques into new and existing system specification tasks will be devised and implemented.

The approach of using catastrophe theory (and in general, chaos theory) for systems’ models can likely be used in conjunction with approaches that have been used in the past for providing fault tolerance in computing and communications systems. As an example, redundant systems such as quadded [5], triple modular redundancy [6], duplicate and match and automatic module replacement systems [7] that use various codes [8], as well as the use of various networking redundancy techniques such as those developed at AT&T including rollback and recovery techniques [9] and reconfigurable and self repair systems [10, 11] are planned to be investigated. While these techniques are based on *single fault* assumptions and long *mean time to failure* assumptions, modifications and extensions may be possible for *Disaster Tolerance* applications

3. FUTURE INVESIGATION

There are many areas of future research for *Disaster Tolerant Computing and Communications Systems* and it is likely that these areas will benefit from a combination of the use of past fault tolerance, reliability, and system synthesis and analysis methods.

In terms of the implementation of disaster tolerant systems, techniques and metrics for quantifying aspects of dependability for various systems must be developed. Tools, concepts, and techniques for tradeoff optimization among availability, performance, correctness, and security need to be better refined. Novel uses of other methods to enhance disaster tolerance should also be investigated such as was done in the use of erasure codes for geoplex communications in the Myriad project [12].

In terms of disaster recovery, better techniques for automated failure management need to be developed. These types of methods would allow for systems to adapt to real-time transient faults that may occur such as software bugs that occur only when a particular input data set is present. In concert with automatic failure management, techniques for better diagnosis and detection of failures is also required. Finally, from an analysis point of view, better forensic tools for system administrators and other software and systems professionals should be developed for use after a disaster has occurred.

4. CONCLUSIONS

This paper has outlined a new area of research: disaster tolerant computing and communications. Whether disasters are man-made or a consequence of nature is not discussed here. We believe that computing and communications artifacts will always be plagued by the phenomena that we define as a disaster. Here, we have outlined a framework for future investigation to cope with these occurrences.

REFERENCES

[1] D.K. Pradan, Ed., **Fault-Tolerant Computing – Theory and Techniques**, Prentice-Hall, 1986.

[2] V.I. Arnold, **Catastrophe Theory**, Springer-Verlag, Third Edition, Second Printing, 2004, (translated from Russian by G.S. Wasserman), ISBN 3-540-54811-4.

[3] R. Gilmore, **Catastrophe Theory for Scientists and Engineers**, Dover Publishers, 1981, ISBN 0-486-67539-4.

[4] T. Poston and I. Stewart, **Catastrophe Theory and its Applications**, Dover Publishers, 1978, ISBN 0-486-69271-X.

[5] J.C. Tryon, Quadded Logic, in Redundancy **Techniques for Computing Systems**, W.C. Mann and R.C. Wilcox, Ed., Washington DC: Spartan, 1962, pp. 205-228.

[6] R.E. Lyions and W. Vanderkulk, The use of triple modular redundancy to improve computer reliability, IBM J. Res. Develop., vol. 7, pp. 200-209, 1962.

[7] T. Bloom, Dynamic Module Replacement in a Distributed Programming System, Ph.D. dissertation, Massachusetts Institute of Technology, 1983.

[8] E.R. Berlekamp, **Algebraic Coding Theory**, McGraw-Hill, New York, 1968.

[9] R.E. Ahmed, R.C. Frazier, et al., Cache-aided Rollback Error Recovery (CARER) Algorithms for Shared-Memory Multiprocessor Systems, in Proc. of Int. Symp. on Fault Tol. Comp. Sys., 1990, pp. 82-88.

[10] S.A. Szygenda and M.J. Flynn, Failure Analysis of a Memory Organization for Utilization in a Self-Repair Memory System, IEEE Trans. on Reliability, R-20(2), May 1971, pp. 64-70.

[11] S.A. Szygenda and M.J. Flynn, Coding Techniques for Failure Recovery in a Distributive Modular Memory Organization, in Proc. of American Federation of Information Processing Societies, 1971, pp. 459-566.

[12] F. Chang, M. Ji, S.-T. Leung, J. MacCormick, S. Perl, and L. Zhang, *Myriad: Cost-effective Disaster Tolerance*, Proceedings of the USENIX Conference on File and Storage Technologies, January 2002.