

# ***Public Review: CAPRI: A common architecture for autonomous, distributed diagnosis of Internet faults using probabilistic relational models***

**George J. Lee**

*Public Reviewer: Ming Zhang*  
*Microsoft Research*

Traditionally, Internet research has focused on usability, such as reliability and performance, while overlooking network manageability. However, as the Internet has undergone an exponential growth in recent years, so has its complexity. For example, a typical end-to-end communication has to rely on the correct functioning of many network components, such as firewalls, proxies, DNS systems, routers and links, to succeed. The malfunctioning of any individual component or the improper interaction between these components may easily lead to network disruption. Such increasing complexity poses a great challenge to the existing way of network management which requires significant amount of human effort. This paper provides a timely answer to this challenge by proposing a new architecture, called CAPRI, which tries to automate the process of fault diagnosis.

There have been several recent works on Internet fault diagnosis [2, 3, 1] and numerous other works on applying Bayesian inference to fault diagnosis. The first category of papers mostly relies on techniques that are specific to the problems in their target domains, e.g. routing systems. While these works provide lots of valuable empirical results, it is unclear how to generalize their techniques to problems outside their target domains. The second category of papers focuses on constructing various theoretical models of existing systems and proposing algorithms to diagnose faults within these models. However, it is unclear whether these models and algorithms can be directly applied to real systems. CAPRI is one of the few papers that try to bridge the gap between these two categories of work.

Although this paper is at a preliminary stage, it does cover quite a few interesting issues. As the author mentions, two key challenges in Internet fault diagnosis are: 1) how to express diagnostic knowledge, and 2) how to do fault inference. The author resorts to probabilistic relational models and Bayesian networks to deal with these two issues. The main contribution of this paper is the author makes case for the applicability of these two techniques and provides some initial validation by diagnosing HTTP connection failures

in the CoDeeN content distribution networks.

Internet fault diagnosis is definitely a very important networking research topic and the solutions proposed by the author have shown great promise. However, there are still many open issues that have not been fully addressed in this paper. First, while there has been much work on measuring Internet failures, we have not had a good understanding of what are the most common types of failures. The answer to this question will provide us insight of the types of data that need to be collected to diagnose the majority of Internet faults. Second, since the Internet is highly decentralized, any Internet-scale fault diagnosis system will require cooperation between multiple administrative domains. How can we provide incentive for ISPs to cooperate with each other without revealing any proprietary information. Finally, what is the limitation of fault diagnosis imposed by the existing Internet architecture? If we are going to have a clean slate design, what are the built-in capabilities that can facilitate fault diagnosis?

## **References**

- [1] M. Caesar, L. Subramanian, and R. H. Katz. Towards localizing root causes of BGP dynamics. Tech. Report UCB/CSD-04-1302, U.C. Berkeley, November 2003.
- [2] A. Feldmann, O. Maennel, M. Mao, A. Berger, and B. Maggs. Locating internet routing instabilities. In *Proc. of SIGCOMM*, September 2004.
- [3] J. Wu, Z. M. Mao, J. Rexford, and J. Wang. Finding a needle in a haystack: Pinpointing significant BGP routing changes in an IP network. In *Proc. Networked Systems Design and Implementation*, May 2005.