STOCHASTIC AND EPIDEMIOLOGICAL MODELS FOR PERFORMANCE EVALUATION OF PEER-TO-PEER NETWORKS

By

Krishna Kumar Ramachandran

An Abstract of a Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: ECSE

The original of the complete thesis is on file
in the Rensselaer Polytechnic Institute Library

Examining Committee:

Prof. Biplab Sikdar, Thesis Adviser
Prof. Koushik Kar, Member
Prof. Alhussein Abouzeid, Member
Prof. Ananth Krishnamurthy, Member

Rensselaer Polytechnic Institute
Troy, New York

July 2007
(For Graduation August 2007)

ABSTRACT

Peer-to-peer (P2P) networks provide a paradigm shift from the traditional client server model of most networking applications by allowing all users to act as both clients and servers. The primary use of such networks so far has been to swap media files within a local network or over the Internet as a whole. These networks have grown in their popularity in the recent past and the fraction of network traffic originating from these networks has consistently increased. Currently, there exist different flavors of the concept such as Napster, KaZaa, Gnutella and BitTorrent networks, to name a few. P2P networks can be broadly categorized into three architectures: centralized (e.g. Napster), distributed (e.g. Gnutella) and hybrid (e.g. BitTorrent) systems. In this thesis, we confine our focus to the latter two architectures since they are more prevalent.

Peer-to-peer networks typically are complex structures with millions of users interacting with each other. The peers in a P2P network behave selfishly, in that the prime objective of a node is to optimize the information retrieval time contingent on the constraints imposed by the network. The bulk of the current thesis work is devoted to investigating relevant problems native to Gnutella type decentralized systems, and seek to quantify the system dynamics in their wake. To this end, we first develop an analytic framework to quantify the impact of the underlying IP router topology on the quality of relevant peer overlay metrics such as file download time and search efficiency. We then address the issue of malware spread in decentralized networks, a serious security threat since the popularity of Gnutella provides malicious peers with a potent means for rapid proliferation, thereby adversely affecting the network performance. We further extend this model by applying it to a wired-cum-wireless scneario, wherein the presence/absence of malware in the Gnutella network governs the malware intensity in the wireless cell phone network. In our final study, we highlight how inefficiencies arise in content replication in hybrid P2P systems like BitTorrent and suggest a possible framework to iron out the kinks.

The current work is an attempt towards understanding the factors affecting

the performance of P2P networks and to quantify their impact so as to facilitate the judicious development of such networks and ensure the proper utilization of the networking infrastructure. We address this issue by developing an analytic framework for modeling and evaluating the performance of peer-to-peer networks while accounting for architectural, topological and user related factors.