CCIT REPORT #732 May 2009

Correctness of Gossip-Based Membership under Message Loss

Maxim Gurevich* Idit Keidar

Dept. of Electrical Engineering
Technion, Haifa 32000, Israel
gmax@tx.technion.ac.il, idish@ee.technion.ac.il

Abstract

Due to their simplicity and effectiveness, gossip-based membership protocols have become the method of choice for maintaining partial membership in large P2P systems. A variety of gossip-based membership protocols were proposed. Some were shown to be effective empirically, lacking analytic understanding of their properties. Others were analyzed under simplifying assumptions, such as lossless and delay-less network. It is not clear whether the analysis results hold in dynamic networks where both nodes and network links can fail.

In this paper we try to bridge this gap. We first enumerate the desirable properties of a gossip-based membership protocol, such as view uniformity, independence, and load balance. We then propose a simple *Send & Forget* protocol, and show that even in the presence of message loss, it achieves the desirable properties.

1 Introduction

Large-scale dynamic systems are nowadays being deployed in many places, including peer-to-peer networks over the Internet, in data centers, and computation grids. Such systems are subject to *churn*, i.e., their membership constantly changes, as nodes dynamically join and leave. Moreover, such systems are often comprised of unreliable components, where node failures and message losses are frequent.

In order to allow nodes to communicate with each other, each node must know the ids (for example, IP addresses and ports), of some other nodes. Such ids are stored at each node in a *local view* (sometimes called membership), or *view* for short. In large systems, it is uncommon to store full views including all nodes in the system, not only because of the amount of memory this would require, but also because of the high maintenance overhead that churn would induce. Instead, one typically stores small views, e.g., logarithmic in system size [8, 2]. Local views are maintained by a distributed *group membership protocol*.

The views of all nodes induce a *membership graph* (overlay network), over which communication takes place. Two nodes are *neighbors* if one of their views includes the id of the other. The properties

*Supported by the Eshkol Fellowship of the Israeli Ministry of Science.