Chapter 5

Serendipity Reloaded:
Fair Loading in Event-Based Messaging

Daniel Cutting
University of Sydney, Australia

Aaron Quigley
University College Dublin, Ireland

ABSTRACT

Client/server approaches to event-based message can scale to millions of users, but at great administrative and financial cost. By contrast, distributed peer-to-peer (P2P) systems offer the promise of smooth scalability from small to large numbers of participants without dedicated infrastructure. Some forms of event-based messaging, such as publish/subscribe, require events to be delivered to groups of consumers based upon their characteristics or interests. Such groups are undefined until the moment of publication and may be very large, posing significant delivery and load distribution problems in P2P environments. This chapter presents Ice, a structured P2P overlay design with scale-free properties that can be used to construct fairly loaded and efficient event-based messaging architectures.

INTRODUCTION

A community is an enduring group of people defined by common characteristics. Historically, a community has described a group of people living together in a particular geographic area, or a social group within a larger society. Formal associations are another form of community but typically members of an association share a common purpose. More recently, Communities of Practice (CoP) (Lave & Wenger, 1991) have described groups of people participating in shared social or business settings. CoP has been extended to several specific types of communities including Communities of Interest (people who share a common interest or goal) (Fischer, 2001) and Communities of Circumstance (people with illnesses or minority groups, for example).

These notions of community have progressed into the online domain. First used by Rheingold in relation to the WELL web site in the early nineties, the term “virtual community” (Rheingold, 2000) described physically distant members bound intellectually, socially and technically.
Today the Internet has grown to be a dominant medium for social interaction with more than 1.4 billion people already online (Miniwatts Marketing Group, 2008). More broadly speaking the term “community” now refers to practically any online shared social activity such as frequent visitors to web sites, chat rooms, and online games (Zhang & Weiss, 2003). Online communities typically require their members to explicitly visit (e.g., websites, portals and member sites), join (e.g., Facebook (Facebook, 2008) and LinkedIn (LinkedIn, 2008)) or opt-in (e.g., mailing lists and RSS feeds). Such approaches often rely on client/server models and sites such as Google (Google Inc., 2008b) have shown they can scale to handle hundreds of millions of users daily. However, this comes at a great administrative and financial cost, and repeated incidents of outage or the scaling problems which plague services such as Twitter (Twitter, 2008) are common.

In parallel to the growth of such explicit communities has been the rise of countless content-oriented tools such as YouTube, Blogger (Google Inc., 2008a) and Flickr (Yahoo! Inc., 2008) which rely on large server farms to deliver their services. These tools have significantly lowered the technical difficulty of producing and publishing content. Combined with increasingly ubiquitous broadband connectivity (some two-thirds of all users in 2006, based on forecasts from 2004 data) (Organisation for Economic Co-operation and Development (OECD), 2008), these tools have driven the proliferation of specialised weblogs (blogs), films, technical documents and podcasts (Pew Internet & American Life Project, 2004, Lindahl & Blount, 2003, Kumar, Novak, Raghavan, & Tomkins, 2003). By 2008, the web site Technorati (Technorati Inc., 2008) was tracking more than 133 million interconnected blog records in the “blogosphere”. Four of the top ten entertainment sites accessed are blogs according to a comScore Media Metrix study in July of 2008 (comScore, 2008). “Open source journalism” is taking root (OhmyNews, 2008) and the concurrent evolution of connected mobile devices such as the Apple iPhone (iPhone, 2008) enables people to publish and receive photographs and other content “on-the-go” as participants in mobile virtual communities (Rheingold, 2002).

However, the Internet is now a mass medium facilitating tremendous social interaction on a scale difficult for an individual to fully grasp or appreciate. While there are hundreds of blogs, thousands of websites and millions of pictures which might be of interest to an individual over the course of their lives, how would one ever find them? As a result, people flock to well known websites or limit themselves to content others have placed into well-defined communities that they opt into over time. The question becomes, how can one serendipitously be exposed to information when online interaction is inherently designed around getting specifically what is requested?

One approach to serendipity relies on the fact that publishers create content with a specific audience in mind. When publishing, publishers could specify the consumer demographic for each article of content. This demographic defines an implicit group or a set of consumers that have some inherent features in common. Implicit groups are defined by the characteristics of its members, rather than their explicit names. Implicit groups are wider in scope than traditional communities. Although they can describe communities of people with shared interests or circumstances, they are, more specifically, subsets of a population of objects selected by the requirements for membership. They may be enduring or ephemeral, generic or highly specific, social or technical. Implicit groups may specify groups not considered to be communities in the traditional sense, such as a set of computers with specific functionality and available resources. Sending messages to these sorts of groups is called implicit group messaging (IGM) (Cutting, 2007, Cutting, Quigley, & Landfeldt, 2007a, 2007b). Some examples include: the PEACH project (Busetta, Merzi, Rossi, & Zancanaro, 2003) which uses the concept of “implicit organization” in order