

**TIMESTAMP-BASED CORRELATION MEASURES FOR
FINDING HIDDEN GROUPS IN CHAT ROOMS**

By

Christopher P. Willmore

An Abstract of a Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

Major Subject: **COMPUTER SCIENCE**

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

Approved:

Mark K. Goldberg, Thesis Adviser

Rensselaer Polytechnic Institute
Troy, New York

April 2008
(For Graduation May 2008)

Abstract

This thesis describes a new two-step algorithm for finding hidden groups from chat transcripts, that is, transcripts of communication where the recipient of a message is not known. The algorithm is presented in two steps: calculating a correlation value between every pair of users in the chat transcript, and finding clusters in the weighted undirected graph that results. The inter-user correlation can be calculated in a number of different ways, some of which are accomplished by projecting individual user transcripts into an inner product space. The clustering step uses the existing iterative-scan algorithm, with some new modifications. This approach is found to work under limited conditions.