A Survey of Privacy Concerns With Dynamic Collaborator Discovery Capabilities

Robert L. Marchant
College of Information Sciences and
Technology
Pennsylvania State University
University Park, PA, USA

marchant@psu.edu

ABSTRACT
Dynamic Collaborator Discovery is concept that proposes using a person's patterns of information access to create models that can then be used to find others with similar interest. This concept may raise privacy concerns to end users. The poster will present the results of a survey conducted in May 2007 to determine if privacy concerns will exist for a dynamic collaborator discovery capability.

1. INTRODUCTION
We often look for someone who will work with us on a joint effort. Often our goal is someone to divide an effort into work shares that can be independently completed similar to how siblings might divide the task of raking leaves into front back and sides of the house. At other times we may hope to find someone with similar interests to share ideas, use as reviewers, combine data, or take advantage of multi-discipline experience (perspective) on a topic that is in common, but crosses discipline. In academia, collaborators may add expertise to a team; synergistically adding capability, or they may simply provide a sounding board; helping in the brainstorming process.

In "Dynamic Collaborator Discovery in Information Intensive Environments", Payton, Daily, and Martin Payton, Daily, Martin [1] describe a concept for dynamic collaborator discovery they call "Indirect Collaborator Discovery". This concept proposes using a person's patterns of information access to create models that can then be used to find others of similar interest. The paper uses accesses logged though a web proxy to create the models (profiles) for the potential collaborators.

An extension of the indirect collaborator discovery method could use database logs, security logs, transactions logs created by users and applications. This method could be most useful for the searchers who use a common database, a common environment or a common tool. For example, intelligence analysts who use a common database for research, analysis, and data aggregation.

Some organizations use a database approach to finding (and matching) potential collaborators. Skills database are created

using a predefined taxonomy from questionnaires, performance review results, supervisor inputs, education and professional seminar records. These databases can be supplemented with the result of analysis software that scans publications and resumes for additional information. McLean et al [2] show that using this type of structure can provide a significant improvement in expert finding.

A rational man who strongly supports his organization, institution, or profession would welcome discovery. But does this rational man care about his rights to privacy? The reasons for this resistance can range from an honest desire to retain control of personal collaborative workload to a desire to retain privacy. To sample this concern, the author conducted a survey based study focused on determining if any of these privacy concerns exist.

2. STUDY DESIGN
The survey is hosted on a commercial site (www.zoomerang.com) and distributed via e-mail to associates of the author. Recipients of the e-mail were requested to forward the URL to others in their social/professional networks. The survey contained the scenario presented below (the Bob scenario) and the following 10 questions that used 5 point Likert Scale answers. The survey was concluded with demographic questions.

THE BOB SCENARIO: Bob is an active computer user both at work and at home. Bob is an engineer working for a large technology company that has many international locations and over 75,000 employees. At home, Bob is active in several social clubs and is involved with a community theater. As Bob uses his computers, both at home and at work, his activities are recorded in the computers history logs. For example, every time Bob visits a website, a record is created of that visit. Bob knows that by using a desktop indexing tool, he can keep track of all the files on his computer and all of the websites he has visited. By using this tool, he can enter a query just as he would using an internet search tool, and locate files on his desktop as well as see the websites he has visited. Bob knows that this tool can keep track of the things that he is interested in. For example, if Bob visits 20 sites that have "recreational walking" as a subject, it would indicate that Bob is very interested in recreational walking. What if Bob could share this list of items he is interested in with others who have similar interests? Bob could use this ability to help find others in his community who are interested in public

speaking, or Community Theater. At work, he could use this ability to find others who share his work interests and problems.

The questions where arranged in the order below with the intent of answering four basic research questions: Are the participants concerned that such a utility would violate their privacy rights? Are there privacy concerns that may affect the use of a dynamic collaborator discovery tool? Is there any correlation between participants interested in finding collaborators and their privacy concerns? Is the concern (if any exists) the same at work as it is at home?

(Q1) I believe this capability could be constructed in such a way that it will not violate my privacy (e.g. it can be created in a way that will allow me to decide who can see my interests).

(Q2) Being able to find others who share my interests is more important to me than any potential loss of privacy (i.e. others finding out what my interests are).

(Q3) I often wonder if others are trying to solve the same problem I am currently working on (at work or at home).

(Q4) I often try to find others who share my interests or are working on a similar problem to mine.

(Q5) I find it difficult to find others who share my interests.

(Q6) I would be angry if I knew my company used this utility.

(Q7) If the capability described in the scenario was used at your place of work, how much do you agree with the following statement? I believe this capability would violate my right to privacy.

(Q8) Presume the capability described in the scenario was used at your place of work, but you were given the choice of using it as well as the choice of allowing others to see your list of personal interest, how much do you agree with the following statement? I believe this capability would violate my right to privacy.

(Q9) I would be angry if I knew that my favorite search engine provider used this capability.

(Q10) Presume the capability described in the scenario was available on your home computer, but you were given the choice of using it and the choice of allowing others to see your list of personal interest, how much do you agree with the following statement? I believe this capability would violate my right to privacy.

3. PRELIMINARY RESULTS
To date results (101 responses) indicate no overwhelming evidence of insurmountable privacy concern. When averaged, all results indicate a favorable perception of the usability of the capability. There is no indication of any measurable difference in attitude for use of the collaborator discovery tool at home or at work. General responses to each question are:

70 percent of respondents believe (agree completely or generally agree) this capability could be constructed without violating their privacy. (Q1)

65 percent feel that being able to find others who share their interests is NOT more important than any potential loss of privacy. (Q2)

58 wonder if others are trying to solve the same problem they are currently working on; 65 percent try to find others working on the same problem; but 75 percent   feel they don't need help finding collaborators. (Q3, Q4, Q5)

The majority of respondents do not feel that this capability if used at work or at home would violate their privacy. (Q6, Q7, Q8, Q10)

59 percent of the respondents would be angry if they knew their search engine provided used this type of utility. (Q9)

4. FUTURE WORK
Some indications are present that suggest that further study of a more qualitative nature (e.g. some semi-structured interviews) may be useful to further evaluate Q2.

Based on these results, the author is currently proposing a project to evaluate this potential by instantiating a discovery capability using volunteer employees at a large aerospace industry corporation. The goal of this initial project will be to provide a proof of concept prototype that will validate a capability to identify experts but for this prototype will keep the identity of the engineers anonymous to the researchers

5. REFERENCES
[1] Payton, D., M. Daily, and K. Martin, *Dynamic Collaborator Discovery in Information Intensive Environments*, ACM Computing Surveys (CSUR), 1999, 31:2, pp. 8 (article 8).

[2] Alistair McLean, Anne-Marie Vercoustre, MingFang Wu. *Combining Structured Corporate Data and Document Content to Improve Expertise Finding*, CSIRO – ICT Centre. Clayton, Victoria, Australia