Bridget Gutting
CSCI 6500: Distributed Computing over the Internet
Project Proposal

Reputation Biased Random Walks for Finding Useful Files in Peer-to-Peer Networks

Research Question and Significance

Unfortunately, in P2P networks, it is easy to select a file which seems useful, but is actually malicious (spyware, adware, virus, etc.). The intent of this project is to expand upon the work done in [1] using a biased *k*-random walk to find useful files in an unstructured P2P network. I will expand upon this by implementing a reputation system in which a peers reputation is affected by the types of files it shares (malicious vs. not). This system will modify the biased *k*-random walk such that the popularity of a peer and the reputation are considered when determining the next neighbor to select.

I also will run the simulations using different evaluations of usefulness. In [1], the usefulness was determined with two generic metrics A, B represented in a vector. Given two files with metrics <a,b> and <x,y> the file with <a,b> is more useful if when comparing elements of the vectors at least one of the elements is larger while the rest are at least equal (ex. $a > x$ and $b \geq y$).

Project Design

Reputation system simulation:

I will implement a reputation system such that the probability of selecting a file is influenced by the reputation of the peer it originates from. So files from peers with a positive reputation are more likely to be selected than files from peers with negative reputations. In order to implement a reputation system, about 5% of the peers will have approximately 50% of their files labeled “malicious”. When these malicious files are selected as the most useful file discovered by the search, the reputation of the peer will be decreased. Conversely, when a nonmalicious file is selected, the reputation of the peer will be increased.

At the start of a run, 1000 peers will be generated the same reputation and each 5000 files generated. These 5000 files will be assigned some usefulness evaluation, and will be in order of popularity (lower file numbers have higher popularity). Each peer has a 5% chance of being malicious; each file of a malicious peer has a 50% probability of being a malicious file. One run will perform a biased k-random walk searching for the i^{th} file with the highest usefulness above the threshold. Each run will perform 100 queries from randomly generated start points. After each query, the reputation will be updated for the peer whose file was selected.

I will implement the four different methods used in [1] and compare the percentage of malicious files selected in these, to the percentage of malicious files selected when using the reputation based system.

Usefulness Evaluations:
To expand on the usefulness evaluation, I will use two different evaluation functions: vector comparison and weighted sum. I will vary the vector dimensions from 2 to 5, where each dimension represents a different generic metric of usefulness. I will use the same scale as [1] , so a file will have metrics that vary from 1 to 5. For vector comparison, a useful file will be above 4 in all dimensions of the vector. For weighted sum, a sum greater than 4 will be considered a useful file. For the weight vector *w*, $\sum_{i=1}^{d} w_i = 1$, where d is the number of metrics being used. The weighted sum, $\sum_{i=1}^{d} w_i v_i$ where *w* is the weight vector and *v* is the vector of metrics for a file.

Reference
[1] Kitamura, H.; Fujita, S.; , "A Biased k-Random Walk to Find Useful Files in Unstructured Peer-to-Peer Networks," Parallel and Distributed Computing, Applications and Technologies, 2009 International Conference on , vol., no., pp.210-216, 8-11 Dec. 2009
doi: 10.1109/PDCAT.2009.49
Abstract:
“In this paper, we consider a problem of finding "useful" files matching a given query in unstructured P2Ps. The proposed scheme is a variant of k-random walk, which combines a synchronization mechanism proposed by Lv et al. with a mechanism to evaluate the usefulness of discovered files. In addition, we apply a variant of popularity-biased k-random walk to accelerate the file search in normal k-random walk under uniform distribution. The goodness of the scheme is evaluated by simulation. The result of simulations indicates that the proposed biased k-random walk scheme certainly finds useful files in short time, without significantly increasing the number of message transmissions.”
URL:
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5372801&isnumber=5372727