Introduction

- Modern society and commerce depend on access to information over the Internet.
- Information is accessed over the Internet using centralized search engines and search indexes, for efficiency and scalability.
- We cannot assume that centralized search engines will always deliver the information we seek, uncensored and unbiased.
- iTrust is a system for publishing, searching for, and retrieving information over the Internet that provides trustworthy access to information.

Overview of iTrust

Main objectives are to:
- Provide users with information they need.
- Publish information for other people to access.
- Search for, and retrieve, published information.
- Detect that the system is under attack.
- Increase the probability of distribution and retrieval when the system is under attack.

Basic elements of iTrust:
- Participating nodes: Form the membership of iTrust.
- Source node: A participating node that produces metadata describing the information, which it makes available to other participating nodes chosen at random.
- Requesting node: A participating node that generates requests and distributes the requests to a subset of participating nodes chosen at random.

How iTrust works:
- When a participating node receives a request, first it compares the metadata it holds with the metadata in the request. If there is a match, it returns the URL of the associated information to the requesting node.
- The requesting node then uses the URL to retrieve the information from the source node.

Implementation of iTrust

The iTrust implementation is based on HTTP.

(a) Web server foundation: Uses Apache Web server with PHP extensions: cURL for inter-node communication and SQLite for node / keyword / resource tracking.

(b) Application infrastructure: XML component generates resource lists; helper functions manage node and resource information; uses Tika / Lucene / WordNet JARs to add tagging and synonyms.

(c) Public interface. Machine interface (green) used for node searching, administration, and uploading resources.

Probabilistic Analysis

Variables:
- Membership contains n participating nodes.
- Metadata are distributed to m nodes.
- Requests are distributed to r nodes.
- Proportion x of the participating nodes are operational.

Formulas:

- Probability q of no match on r trials:
  \[ q = \frac{n!}{(n-r)!} \frac{1}{n!} \]

- Probability p of a match on r trials:
  \[ p = 1 - \frac{n!}{(n-r)!} \frac{1}{n!} \]

Conclusions and Future Work

- iTrust is a trustworthy information distribution and retrieval network with no centralized mechanisms and no centralized control.
- iTrust involves distribution of metadata and requests, matching of requests and metadata, and retrieval of information corresponding to the metadata.
- With iTrust, the probability of matching a query is high for 100%, 80%, and 60% operational nodes, given a reasonable number of participating nodes.
- We plan to do experimental evaluations of the prototype implementation using PlanetLab.
- We are investigating other implementations of iTrust based on SMS, Twitter, etc.
- We plan to make the iTrust source code, tools, documentation, etc. freely available for all to use.

Acknowledgments

This research is supported in part by NSF CNS 10-16193.