NATIONAL SCIENCE FOUNDATION CNS 10-16193

User Guide for iTrust over HTTP Installation and Use

University of California, Santa Barbara

By: Yung-Ting Chuang

This documentation provides explanations for installing iTrust, customizing iTrust, and using iTrust for distributing metadata, distributing requests (queries), and retrieving information.

Table of Contents

1. Overview

The iTrust over HTTP System is a decentralized and distributed publication, search and retrieval system, which is designed to make it difficult to censor or filter information accessed over the Internet.

This documentation is organized as follows. In Section 2, we provide an explanation of how to install iTrust. In Section 3, we provide an overview of the iTrust user interface. In Section 4, we explain how to customize the iTrust over HTTP system, such as setting control parameters for the detecting/defending and membership algorithms. In Section 5, we explain how to run the iTrust over HTTP system. Finally, we provide an Appendix that describes the control parameters and system variables of the iTrust over HTTP system.

2. Installing the iTrust over HTTP System

In this section, we first describe the requirements for setting up the iTrust over HTTP System. We then provide guidelines to enable you to install, configure, and set up the required software and extensions for iTrust. Next, we teach you how to set up a virtual Web host so that you can have multiple iTrust nodes running on the same server (*optional*). Then, we show you how to install the iTrust HTTP code on the Webroot of the virtual Web host. Finally, we teach you how to reset the database (through the Web browser) and make it ready to use.

A. *Software Required by iTrust*

- Apache Web server
- PHP
- SQlite extensions
- PECL HTTP extensions
- cURL extensions
- Subversion (SVN)
- A static HTTP connection to the Internet (such as a static IP address and port number for Web access).

B. *Walk-through Guide to Set Up an iTrust Node*

1) First, you must have installed a working Apache Web server (version 2 or greater), PHP 5 (or greater), SVN, SQLite extensions, PECL HTTP extensions, and cURL extensions. There is nothing particularly notable about these open source packages, and the default settings for these packages are acceptable. In a Debian-based Linux (Debian, Ubuntu, Mint, etc.) distribution, the easiest way to install this software is by typing the 'apt' command, as follows:

```
$ sudo apt-get update
$ sudo apt-get upgrade --show-upgraded
$ sudo apt-get install apache2 apache2-doc apache2-utils
$ sudo apt-get install libapache2-mod-php5 php5 php-pear php5-xcache
$ sudo apt-get install php5-sqlite
$ sudo apt-get install php5-curl
$ sudo pecl install pecl_http
$ sudo apt-get install php-http
$ sudo apt-get install subversion
```

The above commands will automatically install the following software for you: Apache 2 (or higher), PHP 5, Subversion, cURL extensions, SQLite extensions, and PEAR/PECL HTTP extensions.

Other operating systems will have their own install procedures. RPM-based distributions, such as Red Hat EL or CentOS, have procedures similar to the apt-get procedures. MacOS users may use whatever method Apple or the Mac open source community recommends. Windows users may use whatever method is easiest for them. We recommend that you setup iTrust using a Linux-based distribution, because it is the easiest way to install and set up

iTrust.

2) Next, you need to create a file named ‘http.ini’ and place it under the PHP5 configuration folder (e.g., in a Debian-based Linux distribution, you will create the ‘http.ini’ file and locate it under the /etc/php5/conf.d/ folder). Inside the http.ini file, you will create the following line:

    extension=http.so

3) You will need to set up a virtual Web host for your iTrust node after you have completed the previous two steps. The default Apache virtual Web host file suffices to use as a template (e.g., in a Debian-based Linux distribution, it is located in /etc/apache2/sites-available/default). The defaults are fine, and you only need to ensure that the virtual Web host is named whatever you want (e.g., if the virtual Web host is named ‘test’, then ServerName should be set to ‘test’). In addition, you need to make sure that the Webroot is pointed to a file path on the server where you can store the iTrust code (e.g., the DocumentRoot needs to be set to /users/webroot/filepath). You can now access the Webroot through: http://test.yourdomain.com

4) Next, you need to download all of the iTrust code to the Webroot. The easiest way to do this is by checking out the code repository directly to the Webroot. You first navigate your Webroot file path and then check out the code with Subversion. Using the command line with any UNIX-based operating system, you can check out the code by typing the following command:

    $ cd /users/webroot/filepath

    $ svn checkout svn://itrustlab.org/itrust/trunk

Please be aware that it can take a while to check out the entire code, because the spell checking and other auxiliary JAR files are rather large.

5) The final installation step is to create the iTrust database. You can complete this step by first clicking the ‘Tools’ tab at the top of the menu, and then click the 'Remove database' link (please refer to Section 5D). Alternatively, you can go to the URL to create the iTrust database: http://example.yourdomain.com/scripts/remove_database.php

6) You should now have a fully working iTrust node. However, if the server does not have some type of static connection to the Internet (such as a public IP address and port number for Web access), it will not be able to form a membership with other nodes on the Internet. Essentially, you need a way for other nodes or people to 'see' your machine on port 80 (or whichever port number you wish to use). If you have a static IP address on your machine, then you already have a static route. However, if you do not have a static IP address, then you can consider using a dynamic DNS or NAT server to obtain an IP address. Both of those setups are dependent on your router hardware, which is outside the scope of this guide.

Please note that, if you want to use private IP addresses on a local network, you do not need to worry about routing. The iTrust membership can still be formed with private IP addresses. But then you can distribute metadata and make requests only to nodes in the local network.

3. Overview of the iTrust User Interface

Figure 1. Default iTrust Web page.

In this section, we present an overview of the iTrust user interface. Figure 1 shows the default iTrust Web page. Please note that some of the parameters and variables are not fully explained in this section, and you can refer to the Appendix for more details. The iTrust over HTTP system on a node consists of six parts, which are circled in red in Figure 1 and are described below.

A. *Search Interface*

Figure 2. Search Web page.

Figure 2 shows the search Web page, which allows you to make queries, view query results, and retrieve resource information.

B. *User Settings Interface*

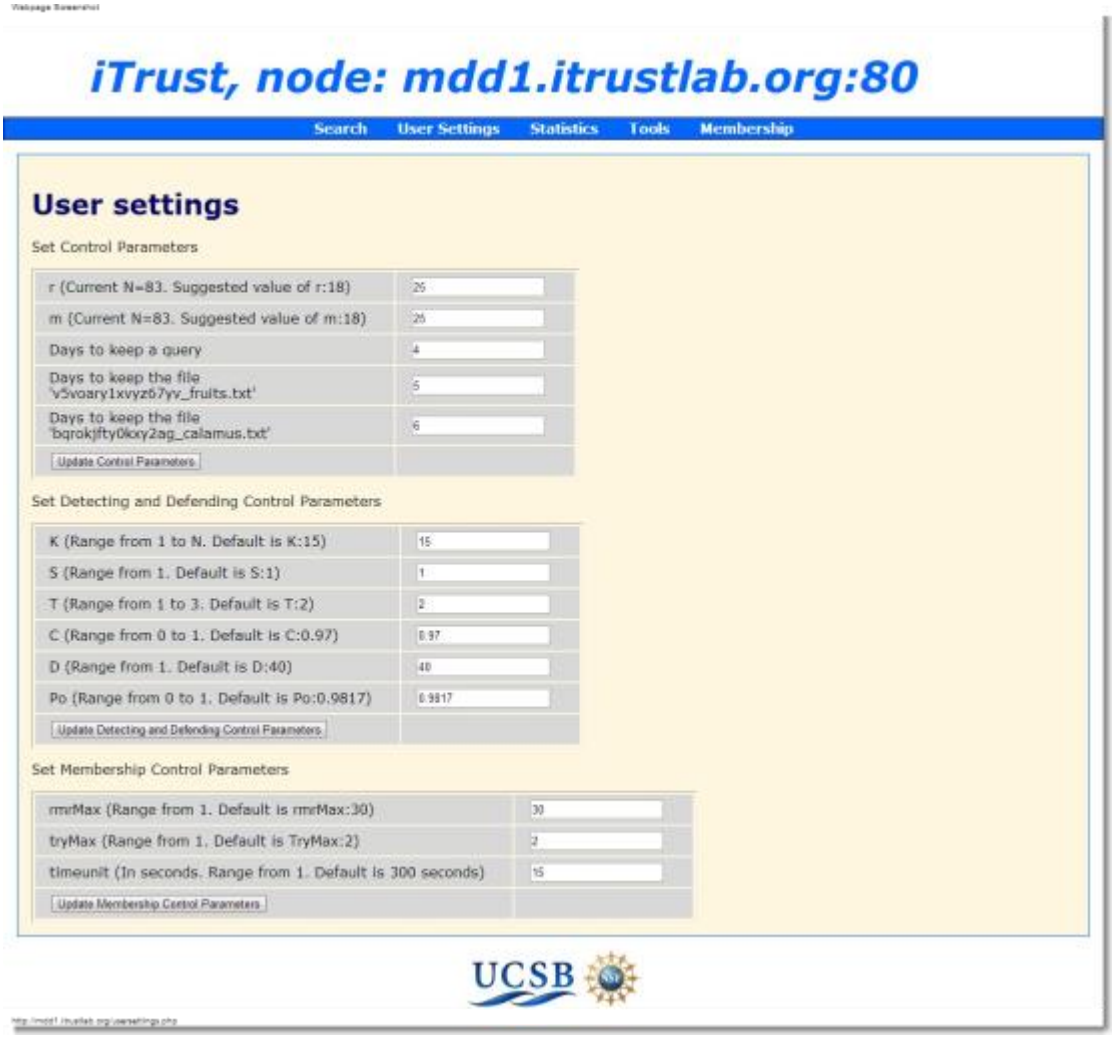

Figure 3. User settings Web Page.

Figure 3 shows the user settings Web page. There are three tables in the user settings Web page. The top table allows you to set *m, r*, and the number of days to keep a query/resource in the system. In iTrust, we provide the current size of the membership, as well as the suggested values of *m* and *r* for you to set these values appropriately. Moreover, you can set the number of days that the system will save a query, and the query will be removed after this duration. Similarly, you can reset the number of days that the system will keep the resource(s). Any blank value is set to the default value, where the default value

is discussed in the Appendix.

The middle table allows you to set rmrMax, tryMax, and timeunit. These variables are used by the membership algorithm to estimate how long a node should wait before it initiates its next distribution of a request/metadata.

The bottom table gives you the ability to set the *K, S, T, C, D*, and *Po* variables, which are used by the detecting and defending algorithms to calculate x' and r' in the iTrust over HTTP system.

C. *Statistics Interface*

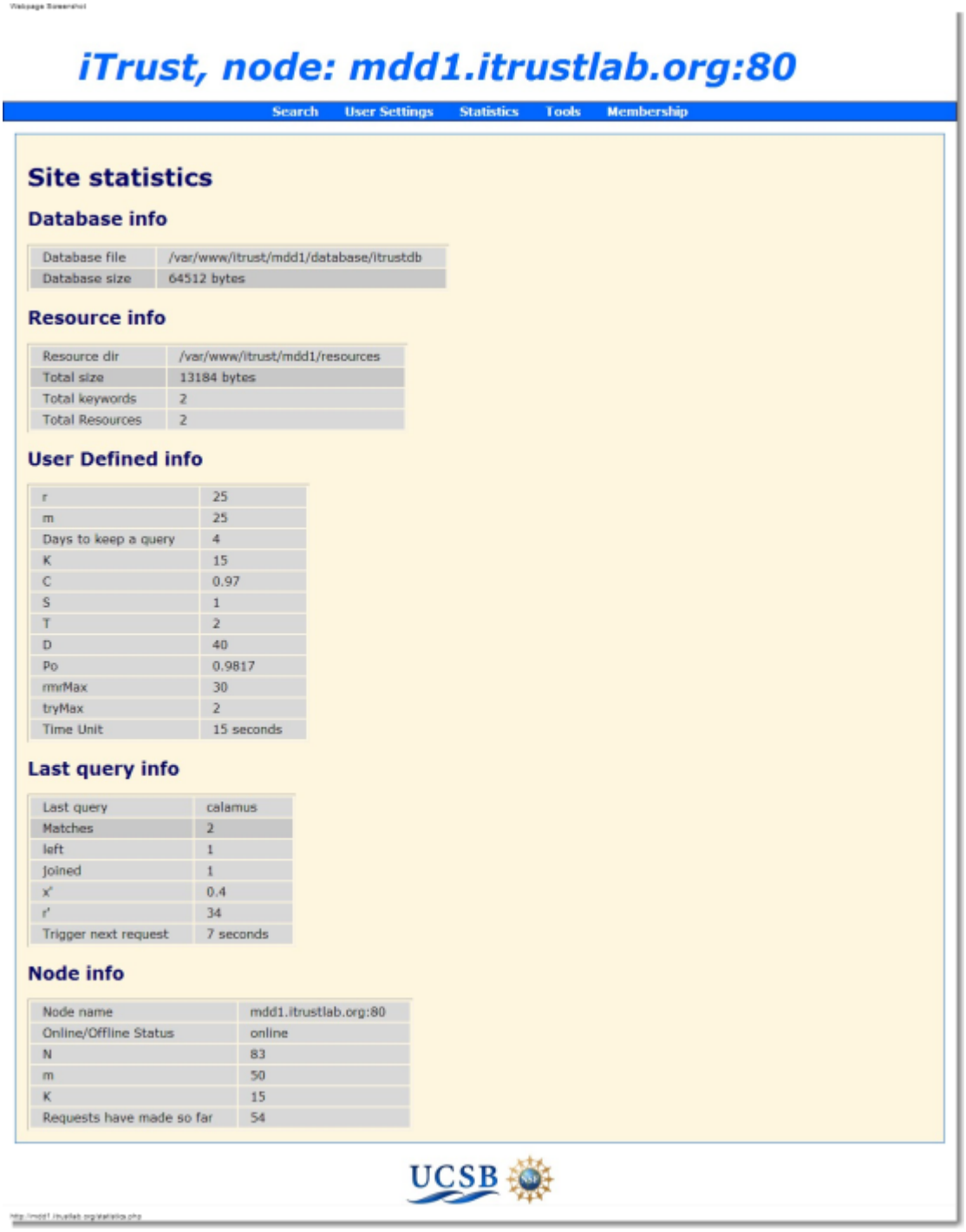

Figure 4. Statistics Web page.

Figure 4 shows the statistics Web page. This Web page displays the following information:

I. Database-related information, including: database filename and total size.

II. Resource-related information, including: total size, resource directory, total keywords, and total resources.

III. User-defined variables, including: number of nodes to which the metadata/requests are distributed, number of days to keep metadata/requests, control parameters for the detecting and defending algorithms, and control parameters for the membership algorithm.

IV. Latest request information, including: query keywords, number of nodes reporting matches, nodes detected to have left the system, nodes detected to have joined the system, estimated proportion of operational nodes in the current node's view, estimated number of nodes to which the requests are distributed to compensate for non-operational nodes, and the time a node should wait before it initiates distribution of its next request/metadata.

V. Current node information, including: current node's name, online/offline status, number of members in its current view of the membership, number of nodes that have the metadata, upper bound *K* on the number of buckets for the detecting algorithm, and number of requests that occurred on this node (e.g., 15 indicates that this node had made 15 requests since it first joined the membership).

D. *Tools Interface*

Figure 5. Tools Web Page.

Figure 5 shows the tools Web page. Using the tools Web page, you can customize the following features:

I. Database tools: Allow you to create a database, add new keywords, or add new members to the database file.

II. Resource tools: Allow you to upload a resource, manually add more keywords to an existing resource, or delete all resources from the current node.

III. Metadata tools: Allow you to create metadata and distribute the metadata to other nodes, or extract the metadata from the node's Inbox.

E. *Membership Interface*

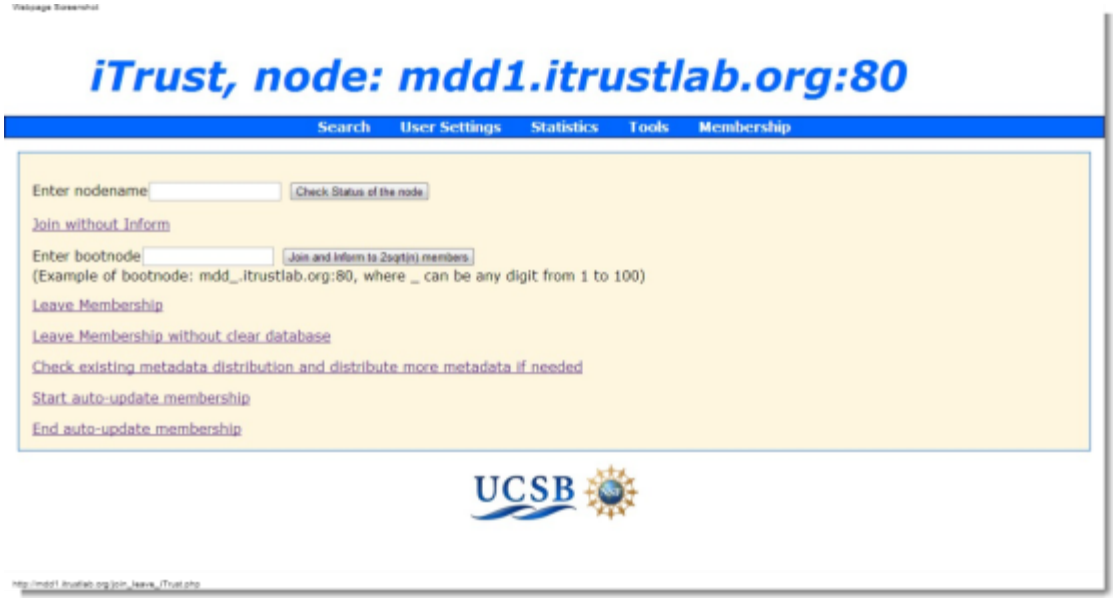

Figure 6. Membership Web Page.

Figure 6 shows the membership Web page. Using the membership Web page, you can perform membership-related functions, such as:

I. Check the status of a node (whether it is online or offline).

II. Join the membership by first obtaining some members from a bootstrapping node, or just simply join the membership

III. Leave the membership and clear the existing resources and database, or just simply leave the membership

IV. Check the number of nodes that have the metadata and re-distribute the metadata to more nodes. Some nodes that have the metadata might have left the system voluntarily. In addition, you may reset *m* at any time. This feature helps a node to determine whether to distribute the metadata to more nodes.

V. Enable/Disable the auto-update membership feature. The auto-update membership feature allows the current node to keep updating its local

view of the membership in the background, to maintain its local view of the membership close to the actual membership. You can disable this feature at any time.

4. Customizing the iTrust over HTTP System

In this section, we describe how you can customize the iTrust over HTTP system.

A. *Setting r, m, and Days to Keep a Query/Resource:* You can define *r*, *m*, and days to keep a query/resource by clicking the "User settings" tab at the top of the menu, and entering the desired values in these text fields (shown in Figure 7). A blank value is set to the default value, which is discussed in the Appendix.

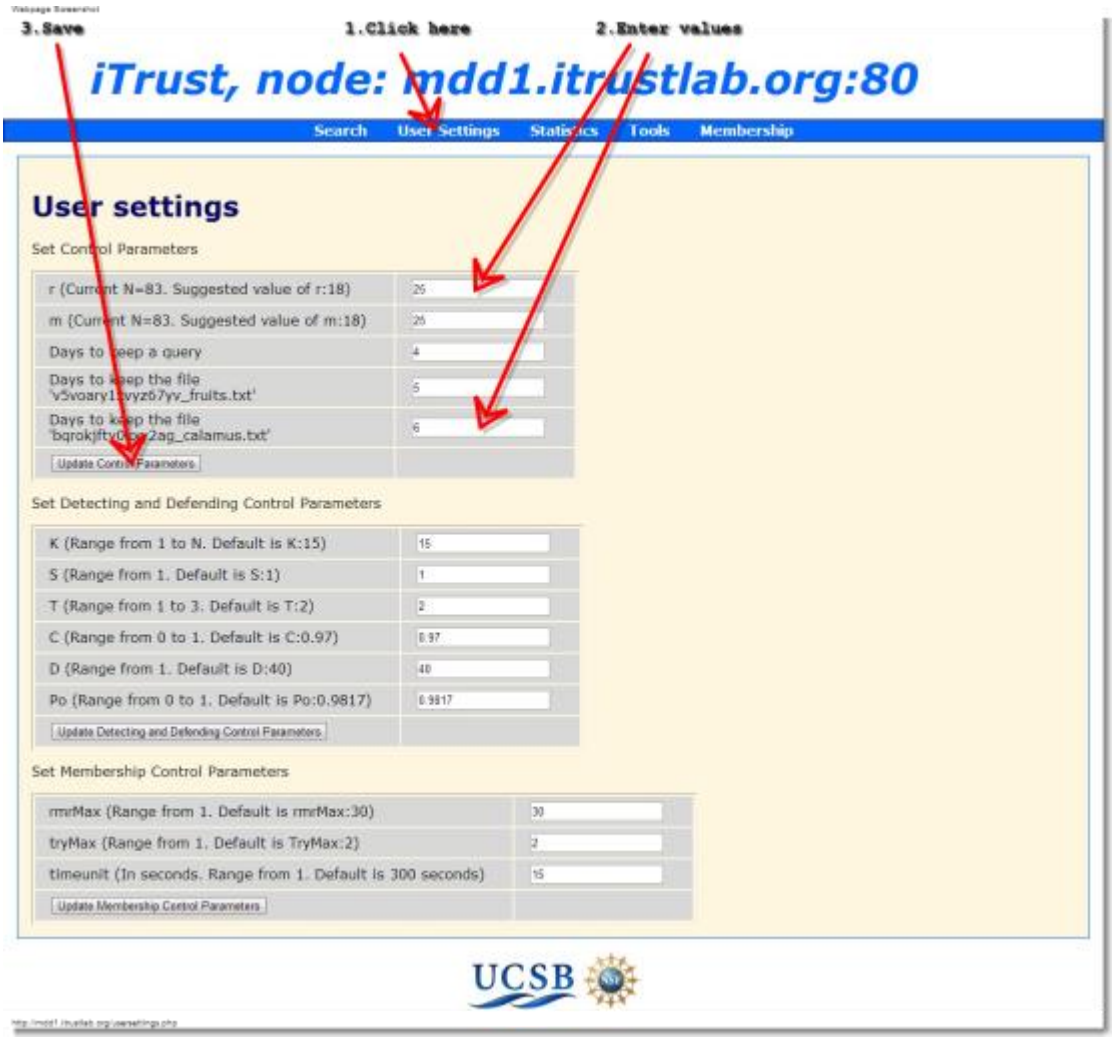

Figure 7. Steps in setting r, m, and days to keep a query/resource.

B. *Setting Detecting and Defending Control Parameters:* You can also set the parameters that are used for the detecting and defending algorithms (such as K, S, T, C, D, Po) by clicking the "User Settings" tab at the top of the menu. Once you have finished inputting these values, you need to click the "Update Detecting and Defending Control Parameters" button (shown in Figure 8). A blank value is set to the default value, which is discussed in the Appendix.

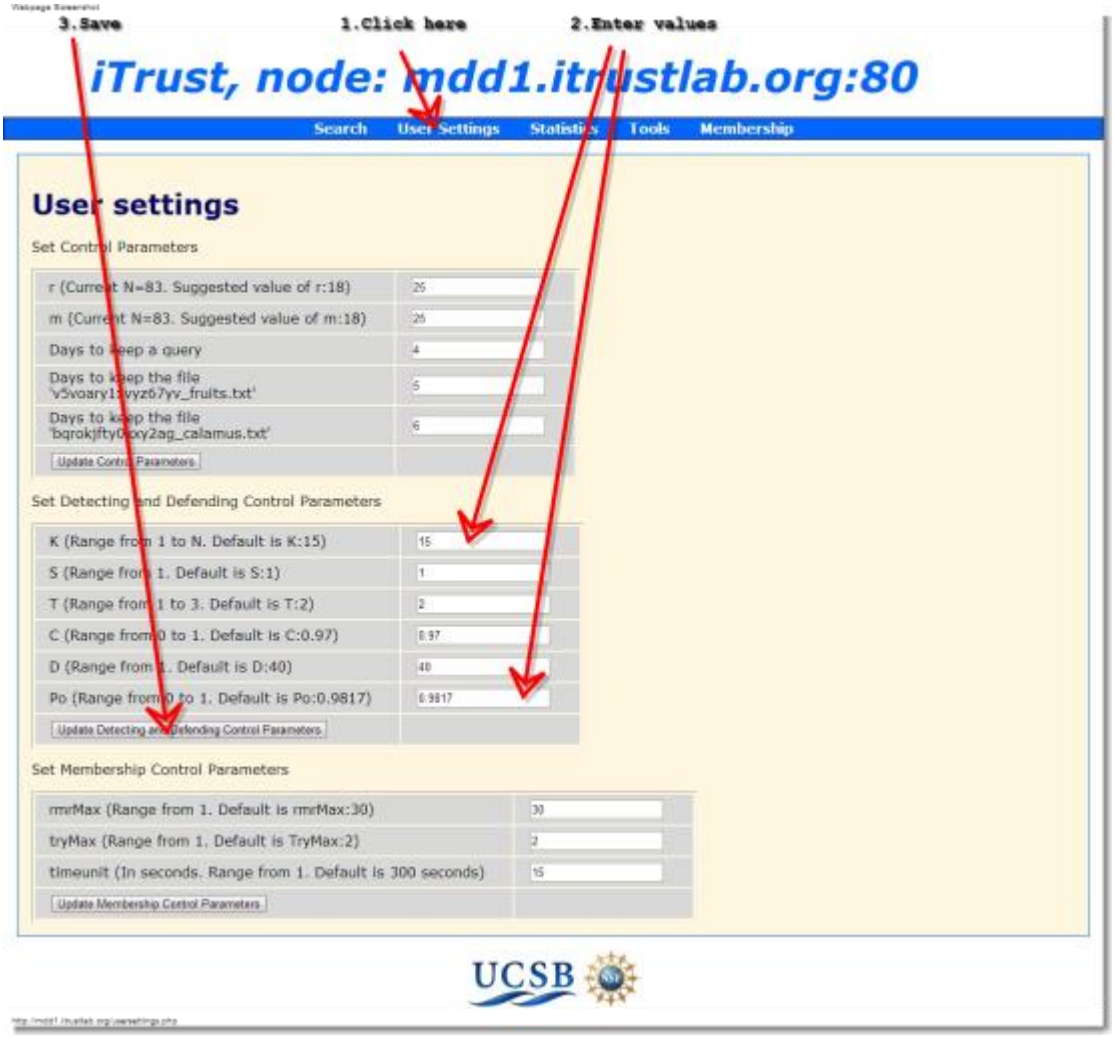

Figure 8. Steps in setting the detecting and defending control parameters.

C. *Setting the Membership Control Parameters*: You can define the membership control parameters (such as rmrMax, tryMax, and timeunit) by clicking the "User Settings" tab at the top of the menu. Once you have finished inputting these values, you need to click the "Update Membership Control Parameters" button (shown in Figure 9). A blank value is set to the default value, which is discussed in the Appendix.

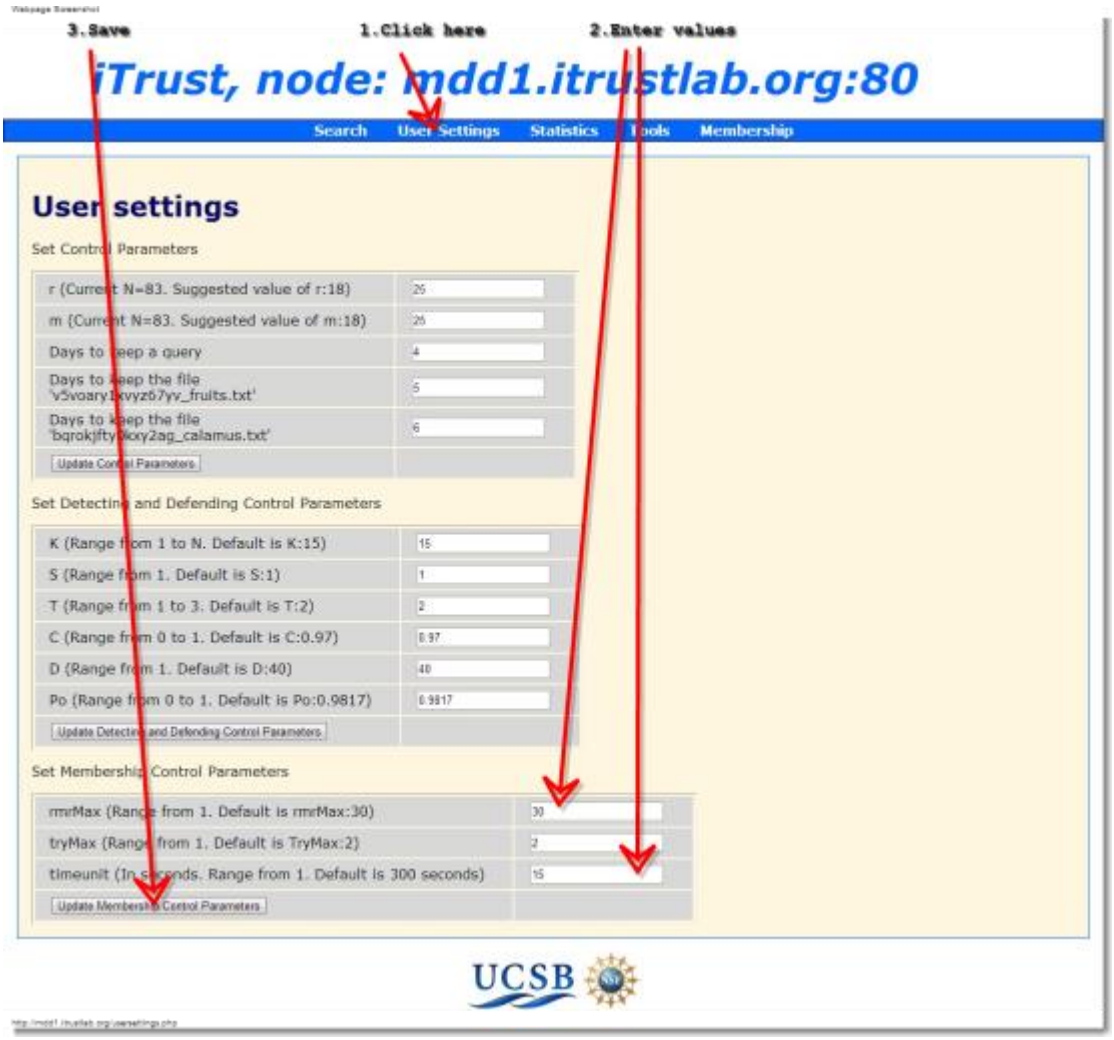

Figure 9. Steps in setting the membership control parameters.

5. Using the iTrust over HTTP System

In the iTrust over HTTP System, you can perform many functions, such as: deleing resources, creating the database, forming a membership, uploading resources, creating and distributing metadata, reading metadata from the Inbox, and distributing requests (queries). In this section, we provide a walk-through guide that describes these functions in the iTrust over HTTP system. We then discuss other optional features for you to customize and execute the iTrust over HTTP system.

There are two types of the users of the iTrust over HTTP system:

1) Users (nodes) that *intend* to upload their resources, create metadata, and publish metadata to other nodes in the iTrust network. We call such users *source nodes*.
2) Users (nodes) that *do not intend* to upload resources or to publish metadata to other participating nodes in the iTrust network. We call such users *requesting nodes*.

Please note that some features are designed only for source nodes, and you should skip these features if you do not wish to upload resources or to distribute metadata to other nodes.

A. Delete All Resources

Before you start executing the iTrust over HTTP system, you should delete all resources from the current node by first clicking the “Tools” tab at the top of the menu, followed by the “Delete all resources” link (shown in Figure 10). This step deletes all uploaded resources from the current node.

Figure 10. Steps in deleting all resources from the current node.

B. Create the Database

After you delete all of the resources in the previous step, the next step is to create your node's database. You can do so by first clicking the "Tools" tab at the top of the menu, followed by the "Create database (and delete old database)" link (shown in Figure 11).

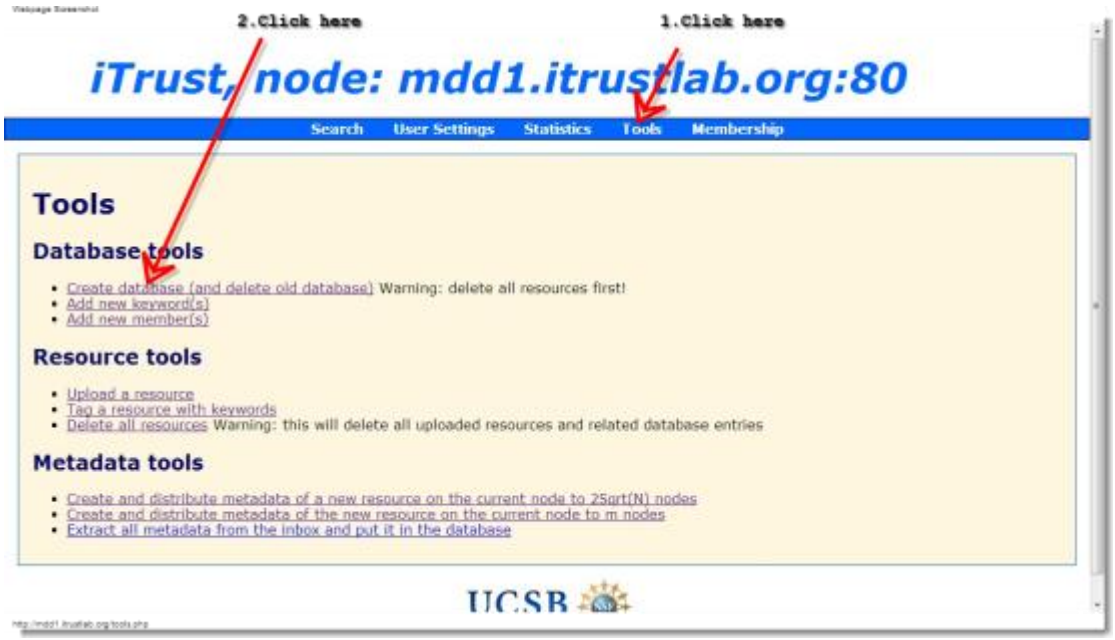

Figure 11. Steps in creating the database.

C. *Forming an iTrust Membership*

Next, you need to join an iTrust membership before you can start distributing metadata or retrieving information. You can join the membership by first clicking the "Membership" tab at the top of the menu, and then choosing from the following options:

- Join the membership by first obtaining *N* members from a bootstrapping node, and then distributing a join message to $2\sqrt{N}$ members. This option is the easiest way for a newly joining node to discover other nodes and to let other nodes known of its existence. In this option, you would first enter a bootstrapping node's name, and then click the "Join and Inform 2sqrt(n) members" button (shown in Figure 12). If you know a bootstrapping node, you can simply enter its name. Otherwise, you can use the provided bootstrapping node (which is mdd_.itrustlab.org:80, where _ is any number between 1 and 100).

Figure 12. Joining the membership by first obtaining members from a bootstrapping node.

- Alternatively, you can choose to form a membership yourself. In this option,

you would simply click the “Join without Inform” link to join the membership without obtaining any members from the bootstrapping node (shown in Figure 13). This option is less preferable, and is intended for a user who wants to create the membership manually. The steps in manually adding members to the current membership are described in Section H.a.

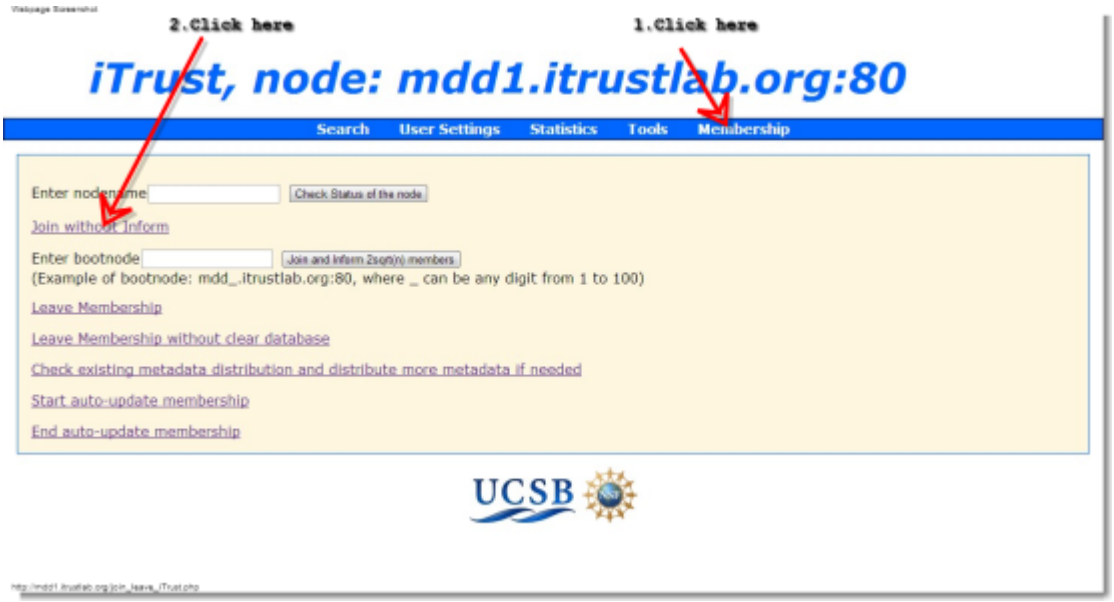

Figure 13. Steps in joining the membership without contacting a bootstrapping node.

D. Upload Resource (For source nodes only)

You can upload a resource by clicking the "Tools" tab at the top of the menu, followed by the "Upload a resource" link. You are then directed to another Web page (shown in Figure 14), where you can:

- Upload a file (either from your hard drive, or remotely from another Web site)
- Tag the file with metadata (either manually generated by you, or automatically generated by the system)
- Set the duration of the file (either set by default or manually set by you)

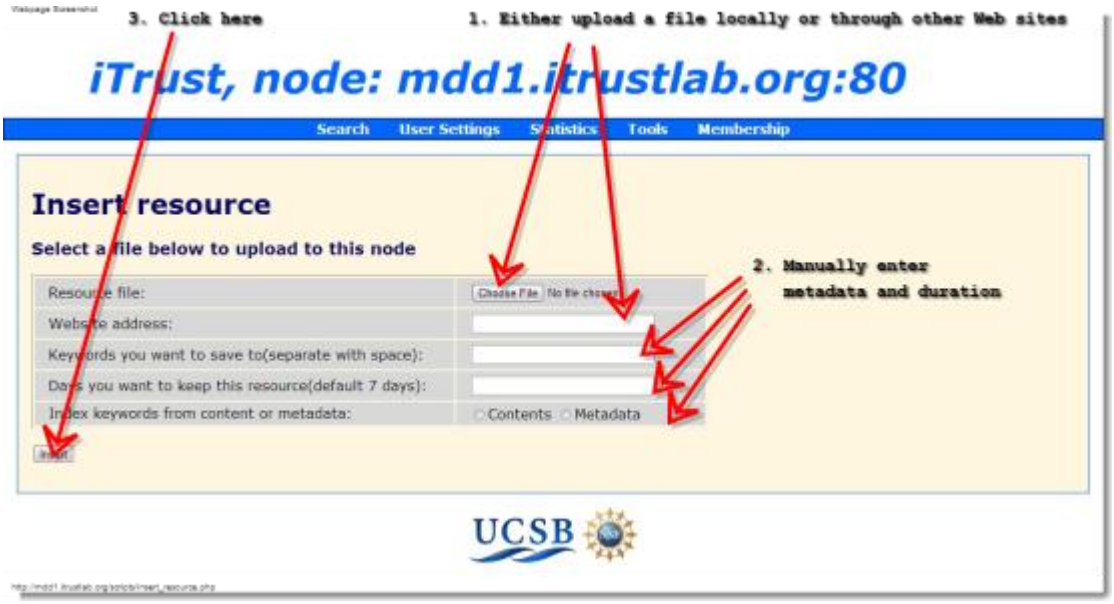

Figure 14. Steps in uploading a resource.

E. Create and Distribute Metadata (For source nodes only)

You can now create and distribute metadata to other nodes by clicking the "Tools" tab at the top of the menu, and then choosing one of the following options:

- You can click the "Create and distribute metadata of a new resource on the current node to 2Sqrt(N) nodes" link, where the metadata are distributed to $2\sqrt{N}$ nodes (shown in Figure 15).
- Alternatively, you can click "Create and distribute metadata of a new resource on the current node to m nodes" link (shown in Figure 15). In this option, you distribute metadata to *m* nodes rather than to $2\sqrt{N}$ nodes.

  We strongly suggest that you perform this step every time right after you have uploaded a new file, so that the system picks different sets of nodes to which to distribute the metadata. Otherwise, the probability of obtaining one or more matches will be affected (slightly) if you choose to perform this step after you have uploaded multiple files.

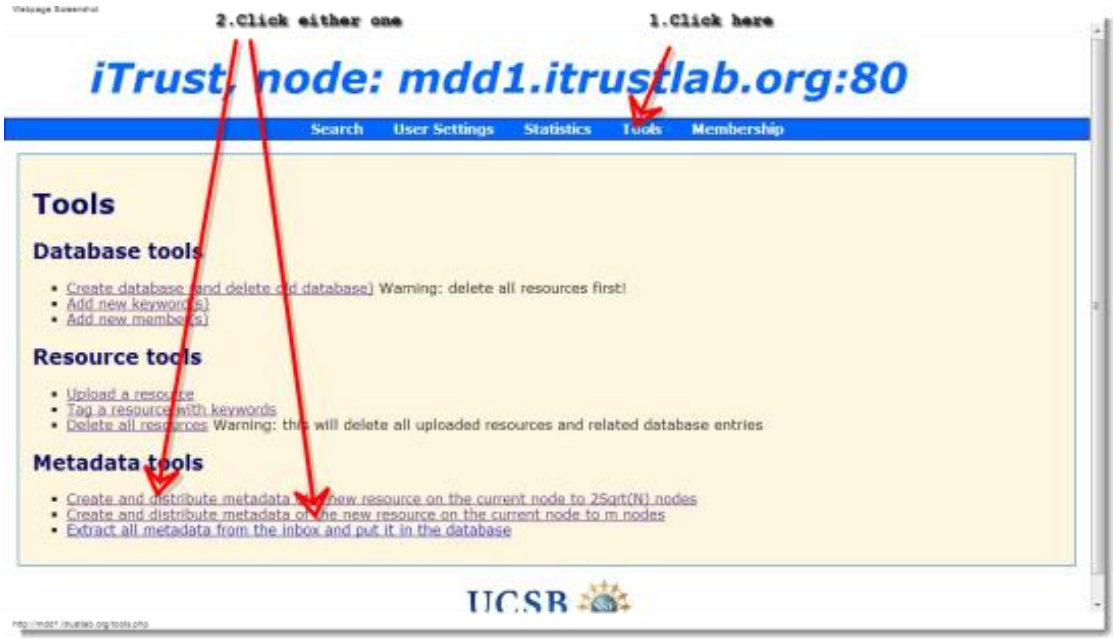

Figure 15. Steps in creating and distributing metadata.

F. Read the Metadata in the Inbox

In the iTrust over HTTP system, every node has its own Inbox which allows it to receive messages from other nodes (you can think of this Inbox as a typical email inbox that receives messages for the user). This feature allows you to check your Inbox, read all of the metadata in your Inbox, and save all of the metadata to the database all at once. You can perform this function by first clicking the “Tools” tab at the top of the menu, followed by the “Extract all metadata from the Inbox and put it in the database” link (shown in Figure 16).

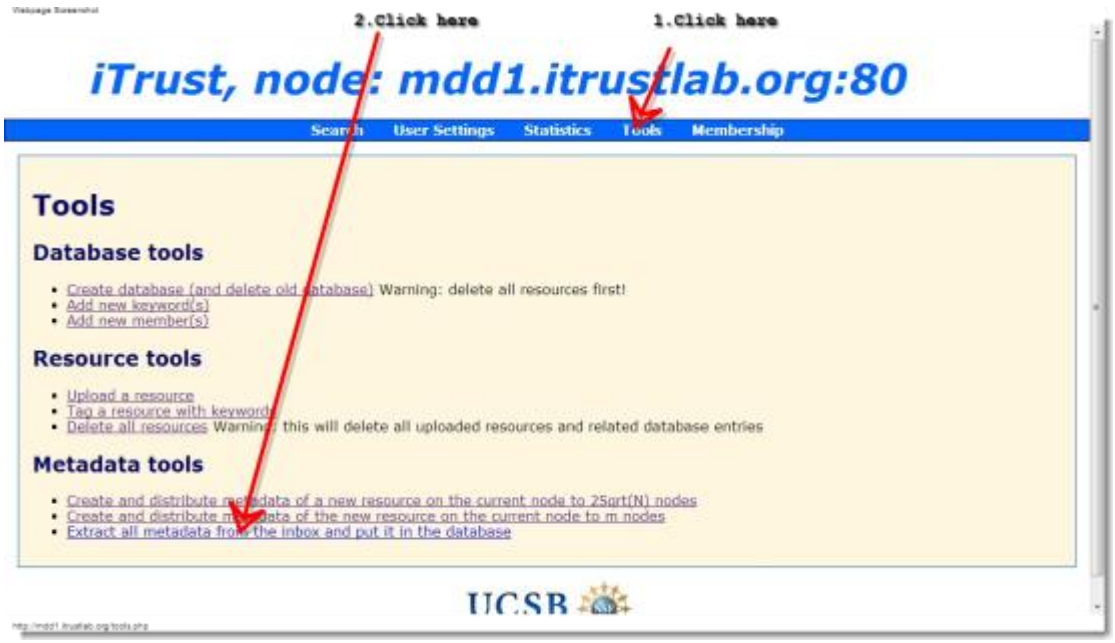

Figure 16. Steps in extracting the metadata from the Inbox.

G. Make Requests

In the iTrust over HTTP System, any node can make a request (query) at any time. You can make a request by clicking the “Search” tab at the top of the menu, and then entering the keywords in the search box. The iTrust over HTTP system displays all of the possible URLs of the matching nodes, from which you can retrieve the documents (shown in Figure 17).

Figure 17. Steps in making requests.

H. Optional Features

In the iTrust over HTTP System, there are also some optional features for you to customize the iTrust over HTTP system. There is no particular order for these functions, and you can execute any function at any time.

a. *Add Members*

You can add new members to the current membership at any time by clicking the "Tools" tab at the top of the menu, followed by the "Add new member(s)" link (shown in Figure 18). The Web page will be directed to another Web page, where you can manually add new members to the membership. The iTrust code checks for duplicate nodes, and won't add a node that is already a member.

Figure 18. Steps in adding members.

b. *Add Keywords* (For source nodes only)

You can manually add keywords associated with a particular resource to the database at any time by first clicking the “Tools” tab at the top of the menu, followed by the “Add new keyword(s)” link. You will be directed to another Web page where you can first view the existing keywords associated with the resource, and then you can add more keywords (shown in Figure 19).

Figure 19. Steps in adding keywords.

c. *Tag a Resource with Keywords* (For source nodes only)

You can manually tag a resource with more keywords at any time by first clicking the “Tools” tab at the top of the menu, followed by the “Tag a resource with keywords” link. The Web page first asks you to select a resource. After you select a resource, you will be able to tag or un-tag the resource with keywords (shown in both Figure 20 and Figure 21). After you finish tagging the resource with keywords, you will click “Finish”.

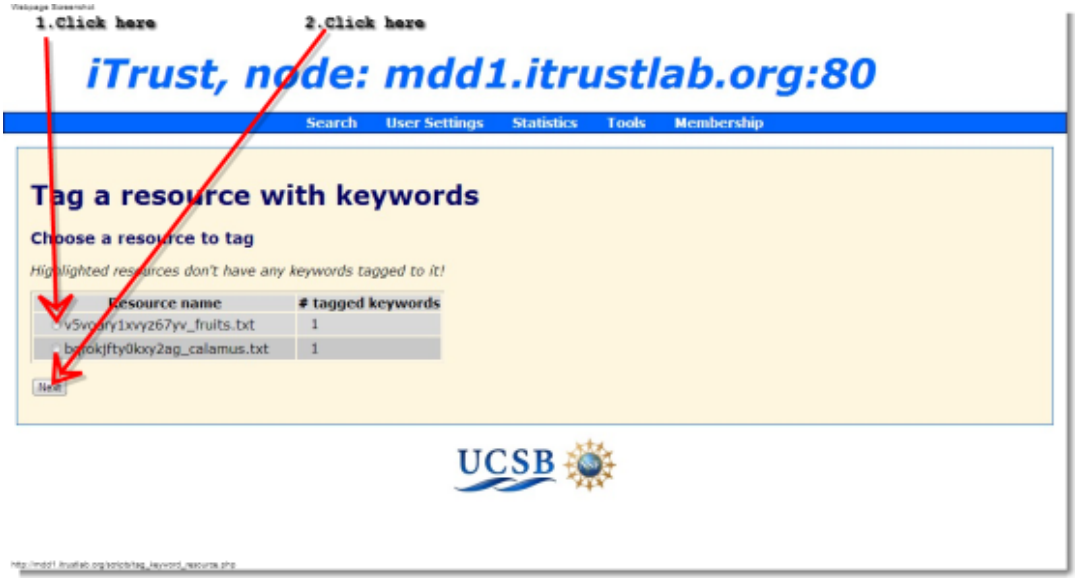

Figure 20. Steps in tagging a resource with keywords (Part 1).

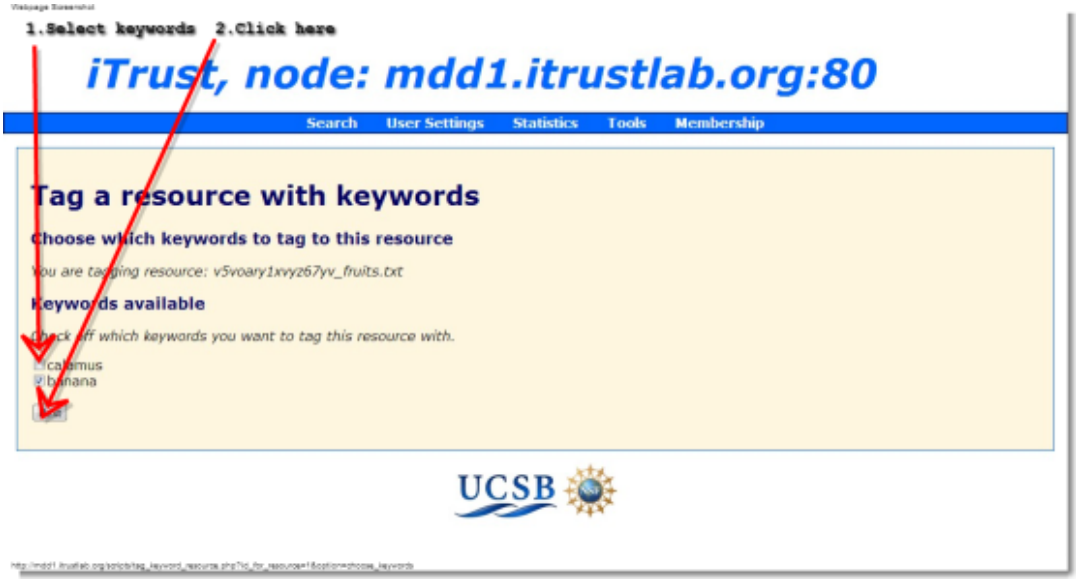

Figure 21. Steps in tagging a resource with keywords (Part 2).

d. *Re-distribute Metadata* (For source nodes only)

You can check the nodes to which metadata have been distributed to determine whether they are still alive, and distribute more metadata to them at any time. For example, you might discover that a node that had metadata has left the membership. Moreover, you can increase the value of *m* at any time. Thus, you can check and re-distribute metadata to more nodes by clicking the “Membership” tab at the top of the menu, followed by “Check existing metadata distribution and distribute more metadata” link (shown in Figure 22).

Figure 22. Steps in re-distributing metadata.

e. *Leave the Membership*

A node can leave the membership at any time. There are two options for leaving the membership. You can choose to clear the existing resources and database before leaving the membership by clicking the “Membership” tab at the top of the menu, followed by the “Leave Membership” link. Alternatively, you can choose to leave the membership without clearing the resources and database by clicking the “Membership” tab at the top of the menu, followed by the “Leave Membership without clearing the database” link (shown in Figure 23).

Figure 23. Steps in leaving the membership.

f. *Auto-Update Membership*

The auto-update membership feature allows the current node to continue updating the membership in the background to keep its local view of the membership close to the actual membership. You can enable this feature by first clicking the “Membership” tab at the top of the menu, followed by the “Start auto-update membership” link. If later you change your mind and want to disable this feature, you can click the “End auto-update membership” link (shown in Figure 24).

Figure 24. Steps in enabling (or disabling) the auto-update membership feature.

g. *View Node Information*

You can check and view the current node’s information at any time by clicking the “Statistics” tab at the top of the menu (as discussed in Section 3C).

Appendix A. Parameters and Variables

In this section, we explain the control parameters (which are set by the user) and the system variables of the iTrust over HTTP system. The control parameters and system variables are:

- Control parameters:
  - m: Number of nodes to which the metadata are distributed
  - r: Number of nodes to which the requests are distributed
- System variables:
  - N: Number of nodes in the current node's view of the membership
  - Matches: Number of nodes that report matches to the current node.

The values of the control parameters and system variables determine the performance of the iTrust over HTTP System.

A. Detecting and Defending Algorithms

In the iTrust over HTTP System, nodes might behave maliciously, where they do not match requests related to sensitive topics. Therefore, we employ detecting and defending algorithms to detect and defend against such malicious nodes in iTrust. The control parameters and system variables for the detecting and defending algorithms are defined as follows:

- Control parameters:
    - Po: The target probability of one or more matches, which is used to calculate r’ in the defending algorithm. It is set to 0.9817 by default.
    - c: The weighting factor of the exponential weighted moving average method used by the detecting algorithm, where its range is between 0 and 1. A value close to 1 means that the detecting algorithm has a higher accuracy in estimating x’ but a longer response time, whereas a value close to 0 has a lower accuracy in estimating x’ but a shorter response time. It is set to 0.97 by default.
    - d: The number of requests in the initial transient for the exponential weighted moving average method used by the detecting algorithm. The detecting algorithm does not start estimating x’ until it has *d* requests. It is set to 40 by default.
    - s: The number of requests that are accumulated before the detecting algorithm estimates x’. A larger value of s means that the detecting algorithm has a longer response time but that it makes fewer mistakes. Conversely, a smaller value of s means that the detecting algorithm has a shorter response time but that it makes more mistakes. It is set to 1 by default.

- t: The number of successive estimates by the detecting algorithm that indicate the same change in the value of x' before the algorithm accepts that change and the defending algorithm takes action to change the value of r'. For example, t=2 means that the defending algorithm does not accept the change until the detecting algorithm encounters the same value of x' twice. A smaller value of *t* results in a "hair trigger" algorithm that responds more quickly to changes but that makes more mistakes. A larger value of *t* results in a more conservative algorithm that responds more slowly to change.
- K: The upper bound on the number k of responses to a request. It is set to 15 by default.

- System variables:
  - x': The estimated proportion of non-malicious nodes in the current node's view, i.e., 1-x is the estimated proportion of malicious nodes in its view.
  - r': The estimated number of nodes to which the requests are distributed to compensate for non-operational (including malicious) nodes to achieve the same probability of a match as when all the nodes are operational. This value is calculated based on *Po*, *x'*, *n*, and *m*.

B. Membership Algorithm

In the iTrust over HTTP System, the membership changes dynamically as nodes join and leave the membership from time to time. Nodes might leave the membership voluntarily or involuntarily because they have failed or have become disconnected. The control parameters and system variables of the iTrust membership algorithm are defined as follows:

- Control parameters:
    - rmrMax: The maximum rate at which a node is allowed to distribute requests (or metadata). A higher value of *rmrMax* means that the current node generates requests (or metadata) and updates its view of the membership faster but with higher message costs. A lower value of *rmrMax* means that the node generates requests (or metadata) more slowly but with lower message costs. This value is set to 30 by default.
    - timeunit: The number of seconds in a time unit. A higher value of *timeunit* means that the distribution rate is high, whereas a lower value of timeunit means that the distribution rate is low. It is set to 300 seconds by default.
    - tryMax: The maximum number of times that a node sends a message in an attempt to receive responses from *r* (or *m*) nodes. For example, tryMax=2 means a node sends a message in a second attempt, if some nodes failed to respond in the first attempt. It is set to 2 by default.
- System variables:
    - left: The number of nodes that the current node detected have left the membership since its last request.
    - joined: The number of nodes that the current node discovered have joined the membership since its last request.