People: Yingang Li, Javed Mostafa, Xiaofeng Wang

{yinli, jm, xw7}@indiana.edu

Institution: School of Informatics, Indiana University, Bloomington

Project : *A Privacy Enhancing Infomediary for Retrieving Personalized Health Information from Web*

Motivation

Online health portals associated with well-known health care organizations are becoming popular for offering timely, credible health information (e.g. WebMD,MedLinePlus). However, there remain several critical concerns related to their practices in personalization and privacy.

The common practice of **personalization** is through locally stored user profile, meaning it is based on the interaction user have with each individual site, and is represented in terms the relevance of local contents to personal interest.
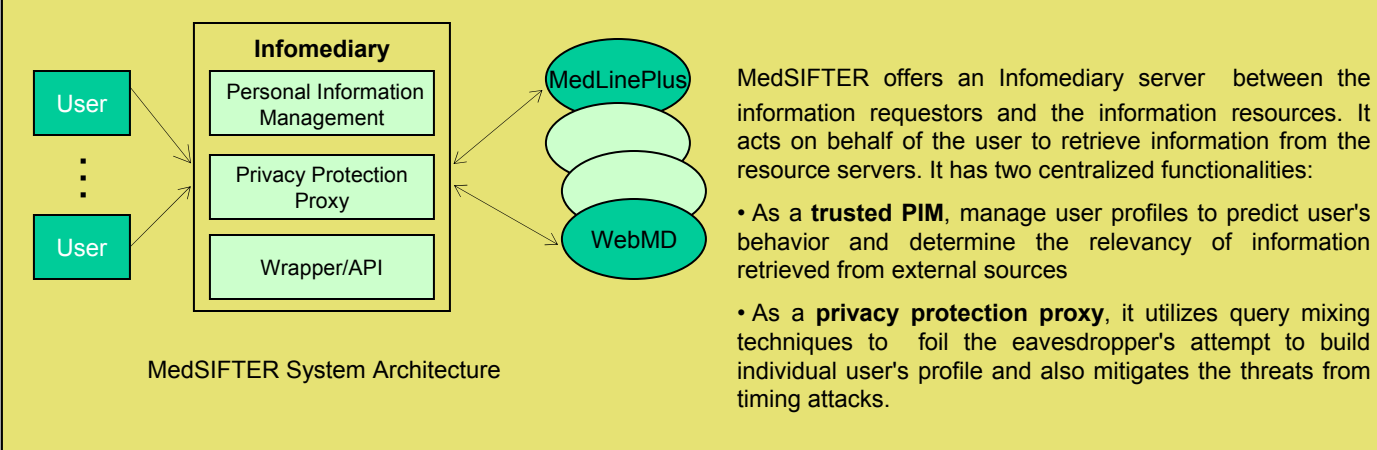
To ensure **privacy**, many web sites require "proof" of identity before allowing use service. Individuals who do not provide essential identifications might be denied services as a consequence.

In the online health-related information retrieval environment, users have to "hunt" for and aggregate information from multiple trust-based sites, and frequently expose profile and identifiable information in exchange of personalized information services.

Objective

In this work, our primary objective is to provide more tangible evidence of solid privacy practices in an personalized health information retrieval setting.

We present a multi-tier health information delivery environment, where a trusted **Infomediary** named *MedSIFTER* serves as a one-stop-shop access point to meet individual's need on quality health and medical information. The proposed Medsifter system centralizes the personal information management at the Infomediary as a unification of previously scattered and inconsistent personalization information. Privacy enhancing measures are also employed at Informediary to keep the user's information aggregation task not revealable to the external information sources, i.e. offering personalized information retrieval without sensitive client personal information being revealed to external eavesdropper.

MedSIFTER System Architecture

MedSIFTER offers an Infomediary server between the information requestors and the information resources. It acts on behalf of the user to retrieve information from the resource servers. It has two centralized functionalities:

• As a **trusted PIM**, manage user profiles to predict user's behavior and determine the relevancy of information retrieved from external sources

• As a **privacy protection proxy**, it utilizes query mixing techniques to foil the eavesdropper's attempt to build individual user's profile and also mitigates the threats from timing attacks.
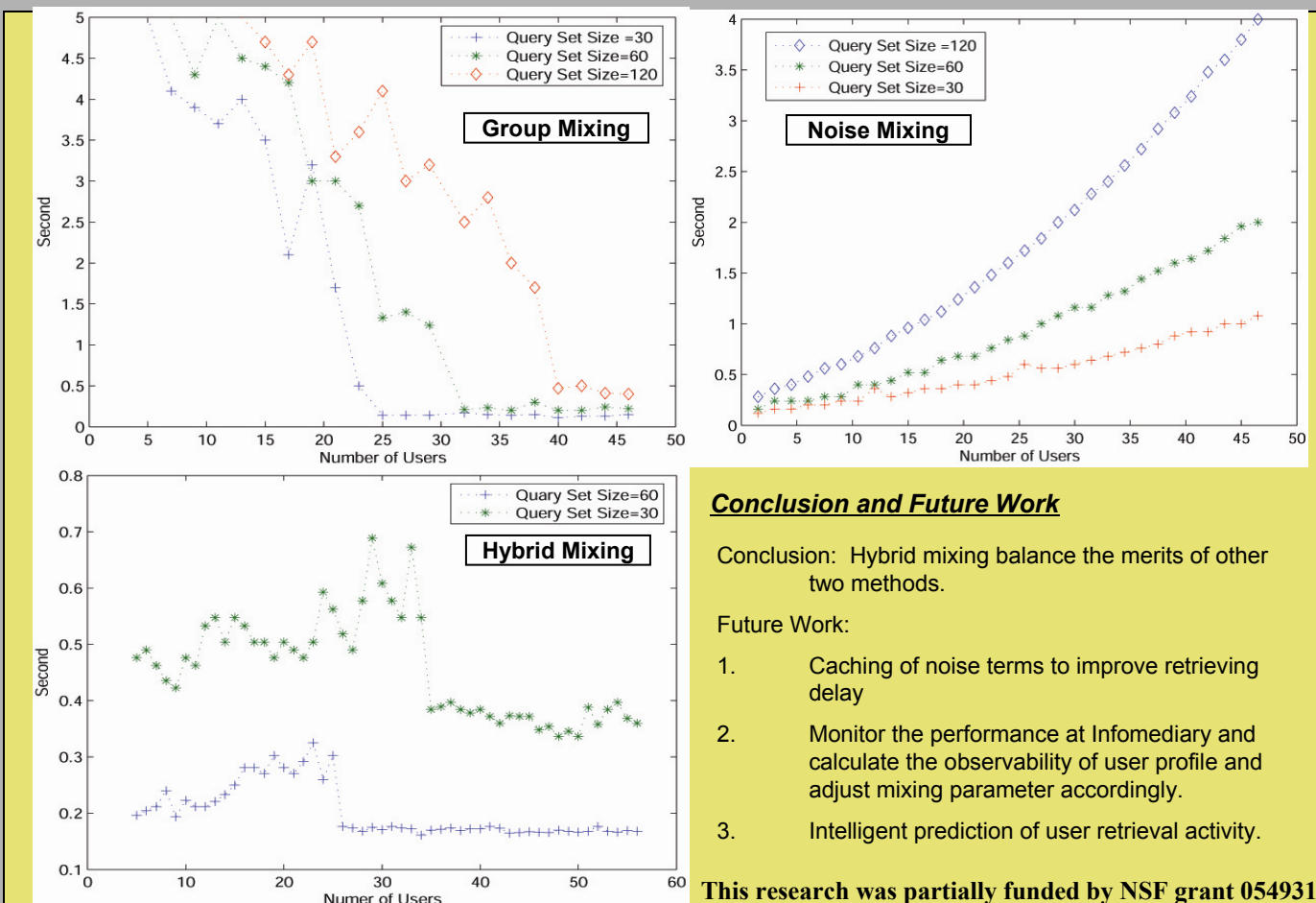
Experiment

An initial prototype system implementing the MedSIFTER model was developed in order to examine the effectiveness of health-related information retrieval. Experiments were conducted to evaluate the MedSIFTER's interaction with simulated clients under different query mixing strategies. A user profile is represented as weighted vector of medical terms. Average turnaround time are measured.

•*Group Mixing:* When users served simultaneously form a non-homogenous group in terms of user profile, all query terms from the group member are mixed together by union to form a minimal group query terms set to be sent out. Informediary will dispatch returned results to real initiator.

•*Noise mixing:* Adds noise query terms to each individual user query until the number of terms in the query set is sufficiently large. When the search results are returned to the Infomediary, it only forwards to the user the information relevant to the original query terms from the user.

•*Hybrid mixing:* When there are no enough users for the current retrieval session, the Infomediary first attempts to generate a small group query term set using group mixing. The remaining terms needed to form a large term set are complemented by noise terms.

Conclusion and Future Work

Conclusion: Hybrid mixing balance the merits of other two methods.

Future Work:

1. Caching of noise terms to improve retrieving delay

2. Monitor the performance at Infomediary and calculate the observability of user profile and adjust mixing parameter accordingly.

3. Intelligent prediction of user retrieval activity.