A Privacy-Preserving Online Medical Pre-diagnostic Scheme in Cloud Environment

Wei Guo (Visiting Student), Jun Shao, Rongxing Lu*, Yining Liu, Ali Ghorbani

Contact Email: rlu1@unb.ca

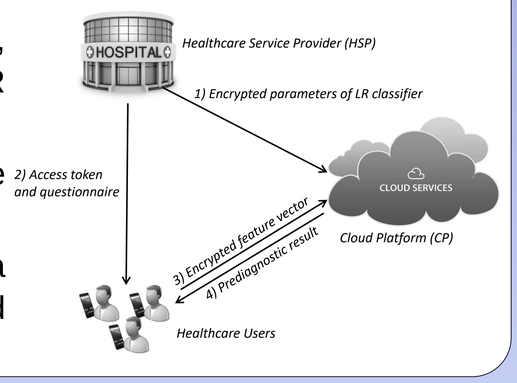
Canadian Institute for Cybersecurity (CIC), University of New Brunswick (UNB)

ABSTRACT

The online medical pre-diagnosis is an effective approach to address the shortage of health workers in the remote and underprivileged area. With its assistant, the limited medical resources will be available for the distant rural communities. However, the flourish of pre-diagnostic system still faces many challenges including information confidentiality and privacy preservation. In this work, we present a privacy-preserving online medical pre-diagnostic scheme, which is based on the logistic regression (LR) classification. With our scheme, healthcare service provider can securely outsource its LR classifier to the cloud platform, and meanwhile the cloud can process the user's sensitive health information without privacy disclosure.

System Model

- **Healthcare Service Provider (HSP)** owns a LR classifier for medical pre-diagnosis, which is considered as a private asset. Hence, HSP will encrypt the parameters of LR classifier before outsourcing it to the cloud platform.
- **Healthcare user** requests the pre-diagnostic service by submitting his encrypted feature vector, which contains sensitive data on his physical symptom.
- **Cloud platform (CP)** uses the encrypted parameters of LR classifier to construct a privacy-preserving medical pre-diagnostic service, which can evaluate users' diseased risk based on their encrypted feature vectors..



Design Goals

In order to achieve a privacy-preserving medical pre-diagnostic service, our scheme should fully guarantee the following two objectives:

- Confidentiality of HSP's LR classifier
- Privacy of user' feature vector

1. System Initialization

HSP outsources the LR classifier by following two steps:

- Firstly, HSP encrypts the parameters of LR classifier before sending it to CP:

$$\Gamma = g^{\gamma} h^{r_0};$$
$$B_1 = g^{\beta_1} h^{r_1}, B_2 = g^{\beta_2} h^{r_2}, \cdots, B_l = g^{\beta_l} h^{r_l};$$
$$\Theta = g^{\theta} h^{r_{l+1}};$$

- Secondly, HSP enumerates all possible value of feature vector. For each vector $\vec{x}$, if $LR(\vec{x}) > 0$, HSP computes $e(g,g)^{LR(\vec{x})}$ and add it into a Bloom filter $BF(m,k)$

BGN Homomorphic Encryption

- Given two ciphertexts $C_1 = g_0^{\mu_1} h^{r_1} \in G$ and $C_2 = g_0{}^{\mu_2} h^{r_2} \in G$, the ciphertext $C \in G_T$ of the product $\mu_1 \mu_2$ can be computed by

$$C = e(C_1, C_2) = e(g_0^{\mu_1} h^{r_1}, g_0{}^{\mu_2} h^{r_2})$$
$$= e(g_0, g_0)^{\mu_1 \mu_2} e(g_0, h)^{\mu_1 r_2 + \mu_2 r_1 + q r_1 r_2}$$

- Addition in G_T: Given two ciphers $C_1 = e(g_0, g_0)^{\mu_1} e(g_0, h)^{r_1}$ and $C_2 = e(g_0, g_0)^{\mu_2} e(g_0, h)^{r_2}$ in G_T, we can calculate the cipher $C \in G_T$ of the addition $\mu_1 + \mu_2$ by

$$C = C_1 C_2$$
$$= e(g_0, g_0)^{\mu_1} e(g_0, h)^{r_1} e(g_0, g_0)^{\mu_2} e(g_0, h)^{r_2} = e(g_0, g_0)^{\mu_1 + \mu_2} e(g_0, h)^{r_1 + r_2}$$

2. Feature Vector Submission

- Healthcare user generates the feature vector $\vec{x} = \langle x_1, x_2, \ldots, x_l \rangle$ and encrypts them before submitting to the cloud platform.

$$X_i = \bar{g}^{x_i} h^{\bar{r}_i} (i = 1,2, \ldots, l)$$

- Healthcare user also needs to compute the assistant data

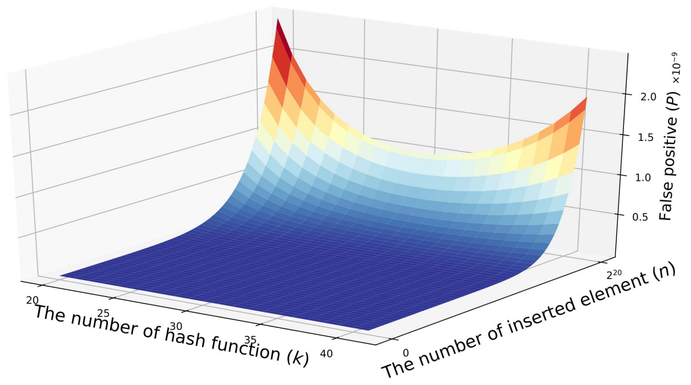
$$D = e(\bar{g}, A_0) \cdot \prod_{i=1}^{l} e(X_i, A_i) \cdot e(\bar{g}^{-1}, A_{l+1})$$

3. Medical Pre-diagnosis

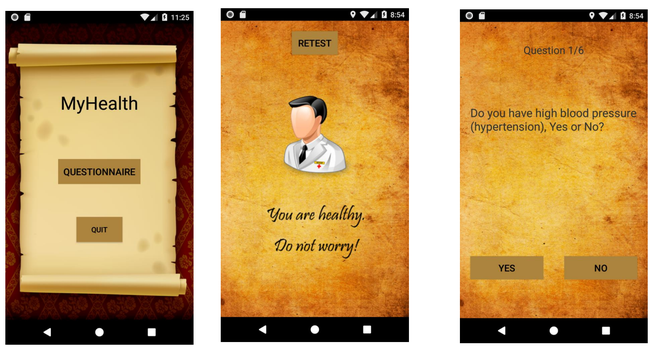
- CP provides the privacy-preserving medical pre-diagnostic service to evaluate user's health status based on his encrypted feature vector. Specially, the service's core algorithm is a secure LR classifier, which is built by encrypted parameters of LR classifier.

$$SecLR(\vec{X}, D) = \frac{e(\bar{g}, \Gamma) \cdot \prod_{i=1}^{l} e(X_i, B_i) \cdot e(\bar{g}^{-1}, \Theta)}{D}$$

- Then, CP checks whether the result is included in $BF(m,k)$.



Performance



Conclusion

- In this poster, we proposed a privacy-preserving online medical pre-diagnostic scheme in the cloud environment, where both the confidentiality of LR classifier and privacy of user's feature vector are protected.

Future work

- For future work, we will explore more classifications for online medical pre-diagnosis, like SVM, Naïve Bayes, and Decision tree, and balance the communicational and computational costs in the pre-diagnostic process.

Acknowledgement: This project is supported by URF Grant (no. 124419).