Hindawi
Complexity
Volume 2021, Article ID 6651662, 8 pages
https://doi.org/10.1155/2021/6651662

Research Article

NN-QuPiD Attack: Neural Network-Based Privacy Quantification Model for Private Information Retrieval Protocols

Rafiullah Khan,[1] **Mohib Ullah**,[1] **Atif Khan**,[2] **Muhammad Irfan Uddin**,[3] **and Maha Al-Yahya**[4]

[1]*Institute of Computer Science & Information Technology, The University of Agriculture, Peshawar, Pakistan*
[2]*Department of Computer Science, Islamia College Peshawar, Peshawar, KP, Pakistan*
[3]*Institute of Computing, Kohat University of Science and Technology, Kohat, Pakistan*
[4]*Department of Information Technology, College of Computer and Information Sciences, King Saud University, P. O. Box 145111, 4545 Riyadh, Saudi Arabia*

Correspondence should be addressed to Rafiullah Khan; rafiyz@gmail.com

Received 24 December 2020; Revised 15 January 2021; Accepted 19 January 2021; Published 2 February 2021

Academic Editor: Furqan Aziz

Web search engines usually keep users' profiles for multiple purposes, such as result ranking and relevancy, market research, and targeted advertisements. However, user web search history may contain sensitive and private information about the user, such as health condition, personal interests, and affiliations that may infringe users' privacy since a user's identity may be exposed and misused by third parties. Numerous techniques are available to address privacy infringement, including Private Information Retrieval (PIR) protocols that use peer nodes to preserve privacy. Previously, we have proved that PIR protocols are vulnerable to the QuPiD Attack. In this research, we proposed NN-QuPiD Attack, an improved version of QuPiD Attack that uses an Artificial Neural Network (RNN) based model to associate queries with their original users. The results show that the NN-QuPiD Attack gave 0.512 Recall with the Precision of 0.923, whereas simple QuPiD Attack gave 0.49 Recall with the Precision of 0.934 with the same data.

1. Introduction

Web search engines (WSEs) have become an essential tool to find topic-specific information due to exponential growth in information and communication technology. To give the most relevant results to the user, WSE maintains his/her profile [1]. The user profile carries the user's web search queries; however, it may contain sensitive information about the user, such as health condition, gender, political affiliation, and religious affiliations [2]. WSEs usually publish privacy policies to inform the users about the usage of their profile data. Most of the time, the terms and conditions are vague, and the user profile may be exposed and misused by third parties, leading to serious privacy concerns [3]. Such an incident happened in 2007 when America Online (AOL) published the web search log of users [4] and in 2005 when the department of justice asked Google to submit their web search log [5].

Numerous techniques are available to address privacy infringement. These techniques include query scrambling [6], profile obfuscation [7], proxy services [8], and Private Information Retrieval (PIR) protocols [9–12]. In the query scrambling technique, the user query is transformed into diverse minor questions and later posted to the WSE, while in the profile obfuscation technique, the user query is posted to the WSE with fake queries. In the proxy-based approach, the user submits his/her query to the WSE through the proxy server, whereas, in PIR protocols, a group of users submit queries on behalf of each other to hide their identity. According to the literature, PIR protocols provide better privacy to WSE users as compared to other techniques [1–13]. Some studies indicate that PIR protocols are vulnerable to machine learning attacks [13, 14], especially QuPiD Attack [1, 3].

QuPiD Attack is a machine learning-based attack that quantifies the privacy provided by the PIR protocols using the user's history and machine learning algorithm. Previously, we tested the performance of QuPiD Attack with ten popular machine learning algorithms that belong to different

families of classification algorithms such as rule-based, tree-based, and lazy-learner algorithms. Moreover, we used the Topic Score feature vector for training and testing of the QuPiD Attack. The Topic Score feature vector comprises a set of numeric values of 10 major topics acquired from the uClassify service [1–13]. According to the results, QuPiD Attack associated 40% anonymized queries with the correct user with 70% Precision.

In this paper, we proposed the NN-QuPiD Attack, a Neural Network-based Query to Profile Distance attack that measures the privacy provided by the famous Private Information Retrieval protocol Useless User Profile (UUP). NN-QuPiD Attack uses a user's profile or web history, and a Neural Network-based machine learning algorithm identifies the user of interest queries in an anonymized web search log. The experiments are conducted with Multilayer Perceptron and Long Short Term Memory (LSTM) and Bidirectional Long Short Term Memory (BiLSTM) classification algorithms with a benchmark dataset released by AOL. Moreover, experiments are also conducted with the IBk (Instance-Based K-Nearest Neighbours) classification algorithm, as IBk performed well in the previous QuPiD Attack version [1]. The experiments are conducted with the Topic Score and query string feature vectors. The results show that the QuPiD Attack's performance with LSTM and BiLSTM is far better with query strings or textual data than the Topic Score in terms of Precision and Recall. The model formed with BiLSTM gave 0.512 Recall with the Precision of 0.923, whereas IBk gave 0.49 Recall with the Precision of 0.934 with the same data. The results show that the QuPiD Attack performance can be improved further using Artificial Neural Network techniques with fine-tuned parameters. Moreover, researchers can effectively use QuPiD Attack as a privacy evaluation mechanism for future private information retrieval protocols.

The rest of the paper is organized as follows: in Section 2, we discuss the state of the art attacks and mechanisms used to evaluate private information retrieval protocols and other privacy preservation tools. In Section 3, we describe the proposed NN-QuPiD Attack. In Section 4, we discuss the experimental setup and dataset, selected classification algorithms, and performance evaluation metrics. In Sections 5 and 6, we discuss experimental results and conclusions.

2. Related Work

According to the literature, PIR protocols provide better privacy to web users [1–13]. Some studies indicate that PIR protocols are vulnerable to machine learning attacks [13, 14]. This section contains a brief discussion about machine learning-based attacks that evaluate the PIR protocols' performance. Peddinti and Saxena [15] proposed an adverse model to assess the performance of both Tor (the onion routing, a proxy-based technique) and TMN (Track Me Not, a profile obfuscation technique) in terms of privacy. They use users' history and machine learning algorithms to train the model and to classify the incoming query. According to the results, the accuracy of their proposed model is 48.88 in case of TMN and 25.95% in case of Tor.

Similarly, Petit et al. proposed SimAttack [16] to evaluate the performance of Tor, TMN, and GooPIR (profile obfuscation technique) in terms of privacy. They proposed n-gram [17] and Dice coefficient [16, 18] based function for model training and testing. According to the results, their proposed model's accuracy in Tor, TMN, and GooPIR was 36.8%, 46.9%, and 35.4%, respectively. Gervais et al. [19] proposed privacy evaluation only for the profile obfuscation model. Similarly, Basla et al. [20] proposed a dummy queries classification, semantic classification, and profile filtering-based privacy evaluation model for profile obfuscation-based solutions.

Khan et al. proposed QuPiD (Query Profile Distance) attack [1, 13] purely for the evaluation of the PIR protocols. QuPiD Attack uses a user's history and machine learning algorithm to build a prediction model that can associate the anonymized query with the correct user. They evaluate the performance of QuPiD attacked with ten well-known machine learning algorithms that belong to Bayesian, rule-based, tree-based, metaheuristic, and lazy-learner families. They used the "Topic Score[1]" feature vector acquired from the uClassify service to build a prediction model. The Topic Score feature vector comprises a set of numeric values of 10 major topics acquired from uClassify service. The uClassify service classifies the textual data (query string) into ten major topics, i.e., Science, Computer, Society, Arts, Health, Sports, Recreation, Business, Games, and Home. According to the results, the IBK performed better by associating 43.4% queries with the correct user with 78% Precision. In this work, we enhanced QuPiD Attack's power by introducing Neural Network algorithms for building the prediction model.

3. NN-QuPiD Attack

PIR protocols provide privacy to the user by creating a group of users and shuffling their queries among the other group members. Due to this shuffling process, the users' queries will never register with their true originator in the web search log and thus provide privacy to the user. We have solved this problem by proposing the Session Window technique [1]. Session Window technique is used to find all the possible time-based sessions (web search log entries) where our User of Interest (UoI) appeared. The whole procedure of the Session Window is illustrated in Figure 1.

In the proposed NN-QuPiD (Neural Network-based Query Profile Distance) Attack model, a web search engine is assumed to be an entity interested in user's (UoI) original queries for accurate profiling. Moreover, it is also assumed that the web search engine already has the user's real history "PU" (i.e., before using any PIR protocol). The proposed model is divided into two major phases, i.e., Model Building Phase and Testing Phase. In the Model Building Phase, first, we applied preprocessing techniques to the user's history "PU" for data cleansing and forwarded to the next step. We used the Recurrent Neural Network with Long Short Term Memory technique to build the classification model in the next step. In Testing Phase, we took an anonymized log of the User of Interest (UoI) and applied the Session Window

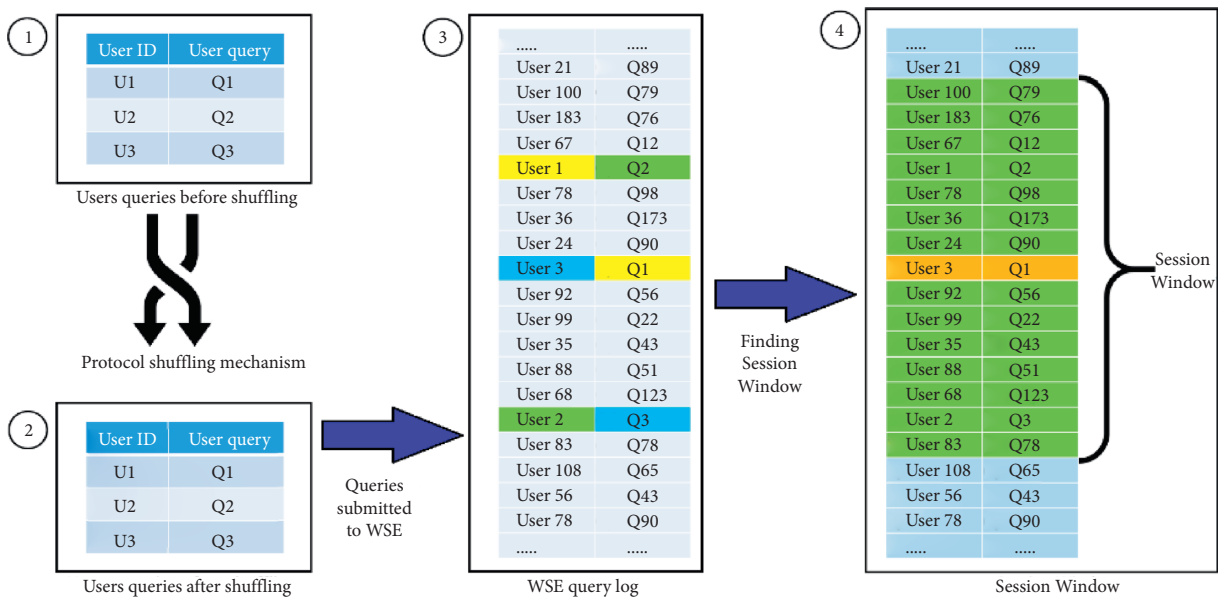

Figure 1: Queries entry in the weblog and Session Window [1].

technique to find all sessions of UoI. After finding all sessions, data preprocessing techniques are used for data cleansing, and then data is given to the classification model for label prediction. The working of the proposed NN-QuPiD Attack is depicted in Figure 2.

4. Methodology

4.1. AOL Dataset and User Selection. We used AOL web search logs for testing our proposed NN-QuPiD Attack. AOL web search log comprises more than 20 million web search queries submitted by 6.5 million users for three months (March to May) in 2006. Although the AOL weblog is relatively old, we are still forced to use it as it is the only available benchmark [1] dataset with our required features for our research. AOL weblog is composed of five attributes, i.e., User ID, query, query time and date, the rank of the clicked item, and URL. March and April data is used as training data or user of interest profile (PU), while the rest of the data (month of May) is used for testing purposes.

We have selected active users for experimentation from the AOL weblog. Active users submitted at least 300 web search queries for more than two months in the entire period. The analysis found that, out of 6.5 million users, only 3.29% (21,407) users are active in the AOL dataset. Then, we randomly select 100 active users from those 3.29% users to keep the cumulative distribution of queries intact. The details of the dataset are given in Table 1.

4.2. Feature Vector Extraction. The AOL web search log dataset comprises five attributes, i.e., User ID, query, query time and date, the rank of the clicked item, and URL. However, we have selected three significant attributes for experimentation, i.e., User ID, query, and query time and date. In the previous QuPiD Attack, we have used the query score attribute from uClassify for model building; however, the query score attribute did not perform well in Neural Networks. Therefore, we have used the user's query and ID as primary attributes to build the prediction model in NN-QuPiD Attack. For the user query, we used the "Word2Vec" tool from the "Affective Tweets" [21] Weka package to produce Word Embedding [22] and then used the Dl4MlpClassifier package [23] to train the model using LSTM and BiLSTM.

4.3. Classification Algorithm. Our previous study has used ten off-the-shelf classification algorithms that belong to various classification families such as rule-based, tree-based, and lazy-learner ones. The results showed that the prediction model built with IBk performed well by associating more than 45% percent of queries with the correct user. However, due to Artificial Neural Network (ANN) family algorithms' promising results, this research aims to test QuPiD Attack's performance with ANN algorithms.

An Artificial Neural Network is a computing system inspired by neurons' simplification in an animal brain [24]. ANN is based on a network of artificial neurons or nodes like a biological brain. Each node receives a signal, processes it, and can send the signal to other neurons connected to it. In ANN, signal at connection is a real number, and each neuron's output is calculated using some nonlinear function of the sum of its input. Usually, neurons are aggregated into layers, and each layer may perform different transformations to the input.

ANN can be classified into six major categories: Radial basis function Neural Network, Feedforward Neural Network, Recurrent Neural Network (RNN), Kohonen Self-organizing Neural Network, Modular Neural Network, and Convolutional Neural Network (CNN). Usually, for problems like the identification of picture, CNN is used. In contrast, for the issues such as sequence to sequence translations (speech or handwriting recognition), RNN is

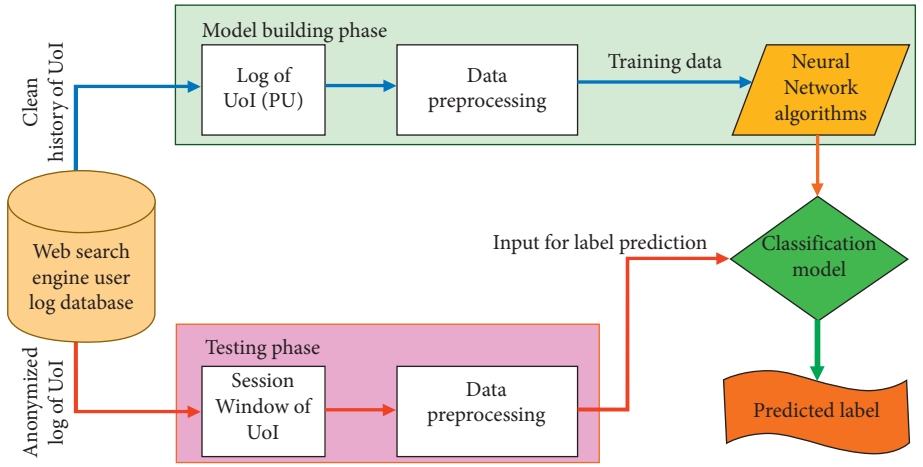

Figure 2: Operation of the NN-QuPiD Attack model.

Table 1: Dataset properties.

Period	March 01 to May 31, 2006
Total no. of users	657,426
Total no. of queries	36,389,567
Selected users	100 active users
Total queries by selected users	175,911
Training queries	116,101
Testing queries	59,809

used with Long Short Term Memory (LSTM). Unlike the simple ANN, in Recurrent Neural Network, a concept of Long Short Term Memory (LSTM) is used, which has the feedback mechanism. This feedback mechanism allows the network to process an entire sequence of the data [25].

This research used two well-known Artificial Neural Network algorithms Multilayer Perceptron and Recurrent Neural Network (RNN), to build the prediction model. Moreover, we also compare the performance of RNN with Long Short Term Memory (LSTM) and Bidirectional Long Short Term Memory (BiLSTM). The parameters used RNN are shown in Table 2.

4.4. Performance Evaluation Metrics. The selected Recurrent Neural Network's performance is evaluated using standard machine learning metrics, i.e., Precision, Recall, and F-Measure. Precision shows the number of correctly identified associations; Recall shows how many corrected associations are placed correctly, while F-Measure is the harmonic mean of the Precision and Recall. Precision, Recall, and F-Measure are mathematically represented in the following equations:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}},$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}, \quad (1)$$

$$F - \text{Measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Table 2: Parameters Recurrent Neural Network (RNN) for NN-QuPiD Attack.

Type	Recurrent Neural Network, Multilayer Perceptron
Layer	LSTM and BiLSTM
Activation function	Activation ReLU
Gate activation function	Activation sigmoid
No. of Epochs	5, 10, 15, 20, 25, 30, 50, 100, and 150

"True Positive" shows the positive associations that are correctly identified as positive. "False Positive" indicates the number of associations falsely identified as correct, and "False Negative" indicates the number of associations falsely identified as negative.

5. Results and Discussion

This study's primary aim is to improve and evaluate the QuPiD Attack's performance with Neural Networks due to their promising performance in other classification problems. Previously, QuPiD Attack's performance was tested with ten classification algorithms belonging to different families and it was found that IBk performed well out of all selected algorithms [1]. Therefore, in this study, we improved the QuPiD Attack's performance by using RNN with LSTM and BiLSTM layers with different Epochs numbers. For experimentation, we took 100 active users from the AOL web search log and evaluated NN-QuPiD Attack's performance in terms of Precision, Recall, and F-Measure.

Initially, we conducted the experiments with a basic Feedforward ANN algorithm Multilayer Perceptron. The model's performance is tested with both Topic Score feature vector and query string feature vector under 0.1 to 1.0 learning rates. The results show that the QuPiD Attack's performance with Multilayer Perceptron was a complete disappointment in both Precision and Recall with the Topic Score feature vector scenario. However, we had hoped for better results as ANN has been used in various tasks

Table 3: Performance of NN-QuPiD Attack with Multilayer Perceptron algorithm under different learning rates.

Learning rate	Topic Score		Query strings	
	Precision	Recall	Precision	Recall
$L = 0.1$	0.364	0.207	0.45	0.311
$L = 0.2$	0.369	0.211	0.41	0.328
$L = 0.3$	0.362	0.204	0.402	0.306
$L = 0.4$	0.377	0.206	0.399	0.292
$L = 0.5$	0.341	0.204	0.42	0.275
$L = 0.6$	0.359	0.198	0.37	0.284
$L = 0.7$	0.33	0.192	0.35	0.247
$L = 0.8$	0.34	0.195	0.38	0.245
$L = 0.9$	0.351	0.194	0.401	0.254
$L = 0.10$	0.358	0.195	0.388	0.234

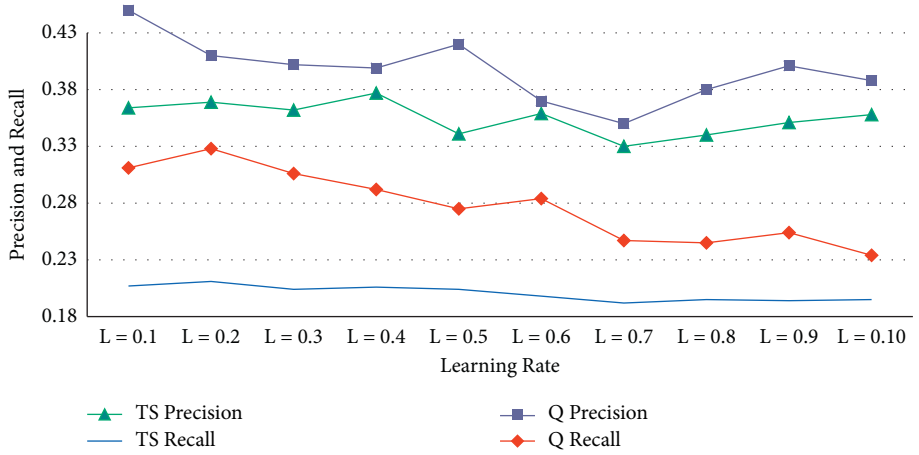

Figure 3: Performance of NN-QuPiD Attack with Multilayer Perceptron algorithm.

effectively. Even with different learning rates, the maximum Recall we got was 0.211, with a 0.2 learning rate.

In contrast, the Recall rate for the rest of the learning rate fluctuates between 0.192 and 0.207. However, Multilayer Perceptron based model performed better in the case of user query string feature vector scenario associating 32.8% queries with the correct user with 41% Precision. The results of the experiments are shown in Table 3 and Figure 3.

TS shows the Topic Score feature vector scenario, and Q shows the user query feature vector scenario.

For further improvement in QuPiD Attack with ANN, we then considered Recurrent Neural Networks with LSTM and BiLSTM configuration under various numbers of Epochs between 5 and 150. Initially, we conducted experiments on the Topic Score feature vector. With the Topic Score feature vector, QuPiD Attack was able to associate 33.2% queries with the correct user with 38.2% Precision with BiLSTM configuration. With LSTM configuration, QuPiD Attack was able to associate 22.5% queries with the Precision of 37.2%, whereas IBk associated 48.1% queries with the correct user with the Precision of 78% with the same configuration. The results of the RNN based QuPiD Attack with the Topic Score feature vector-based experiments are shown in Table 4 and Figure 4.

QuPiD Attack performed better with BiLSTM and Topic Score feature vector as compared to the Multilayer Perceptron algorithm. In the next experiment, we considered user queries to build the prediction model instead of the Topic Score to study a model's performance with query strings. The results show that the QuPiD Attack's performance with LSTM and BiLSTM is far better with query strings or textual data compared to the Topic Score in terms of Precision and Recall. The model built with BiLSTM gave 0.512 Recall with the accuracy of 0.923 at 100 epochs, and LSTM gave 0.47 Recall with the Precision of 0.932 at 150 epochs. However, IBk gave 0.49 Recall with the Precision of 0.934 with the same data and query string feature vector. The NN-QuPiD Attack results with query string feature vector are shown in Tables 5 and 6 as well as Figures 5 and 6.

From the results, it is concluded that the performance of LSTM is slightly inferior to IBk, but BiLSTM gave better results, and it can be improved further by the fine-tuning of other parameters and functions used to build the prediction model.

Table 4: Performance of NN-QuPiD Attack with RNN algorithm and Topic Score feature vector.

Classification algorithm	Precision	Recall
RNN LSTM	0.372	0.225
RNN BiLSTM	0.382	0.332
IBK	0.78	0.481

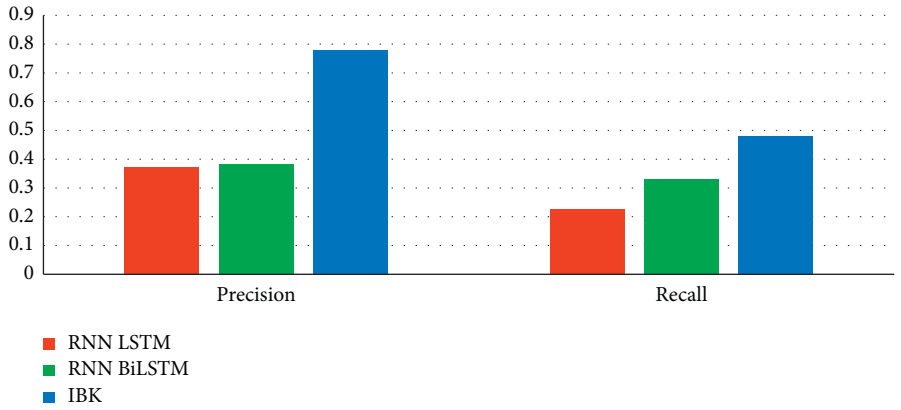

Figure 4: Performance of NN-QuPiD Attack with RNN algorithm and Topic Score feature vector.

Table 5: Performance of NN-QuPiD Attack with RNN algorithm and query string feature vector under different Epochs.

Epochs	RNN LSTM		RNN BiLSTM	
	Precision	Recall	Precision	Recall
5	0.936	0.432	0.933	0.451
10	0.95	0.457	0.938	0.452
15	0.936	0.462	0.945	0.468
20	0.936	0.463	0.92	0.485
25	0.936	0.462	0.928	0.479
30	0.934	0.467	0.945	0.493
50	0.932	0.459	0.925	0.493
100	0.934	0.466	0.923	0.512
150	0.932	0.47	0.931	0.51

Table 6: Performance comparison of RNN and IBK algorithms based NN-QuPiD Attack with query string feature vector.

Classification algorithm	Precision	Recall
RNN LSTM	0.932	0.47
RNN BiLSTM	0.923	0.512
IBK	0.934	0.49

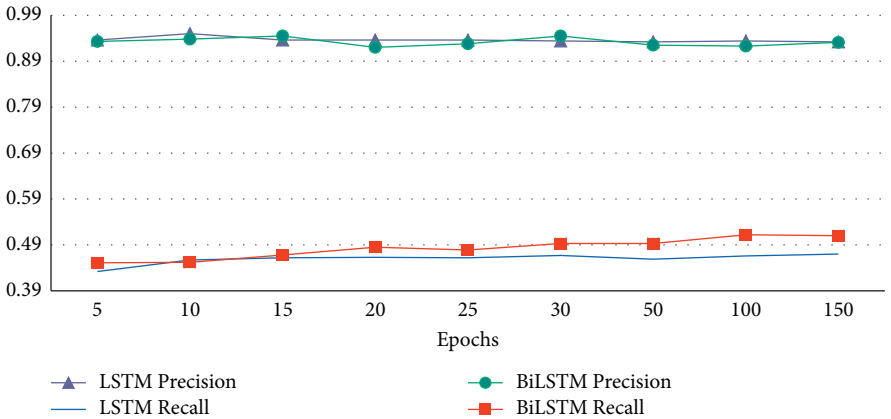

Figure 5: Performance of NN-QuPiD Attack with RNN algorithm and query string feature vector under different Epochs.

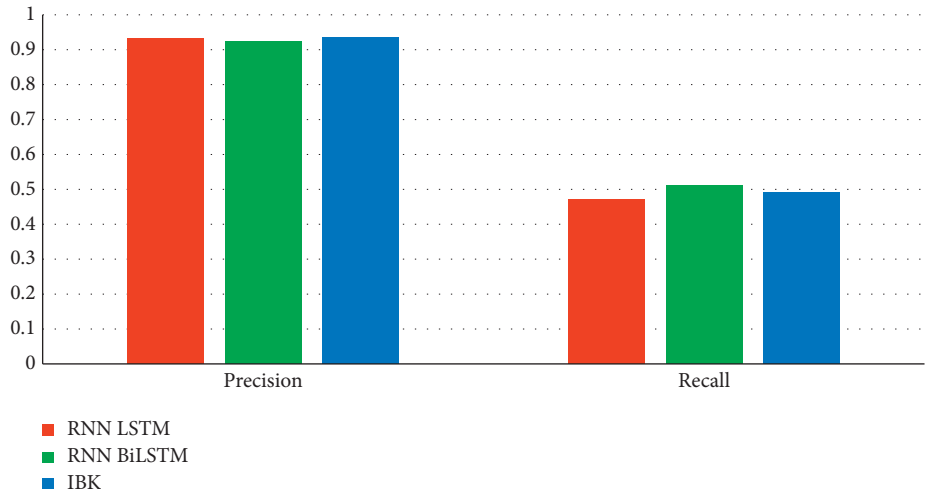

Figure 6: Performance comparison of RNN and IBK algorithms based NN-QuPiD Attack with query string feature vector.

The Recall of the attack shows how many queries are correctly associated with the correct user, while Precision shows how much machine results are close to the actual values. According to the results, the performance of BiLSTM is better when the machine is trained with query strings. Recall's rate shows that attack based on BiLSTM can associate more than 50% anonymized queries with the correct user. The performance of the Multilayer Perceptron (MPL) algorithm and both numeric and textual data are found deprived as compared to BiLSTM. Upon investigation, it was found that BiLSTM can handle textual and sequential data efficiently due to its bidirectional feeding process.

In the previous QuPiD Attack version, we tested the attack's performance with ten machine learning algorithms belonging to tree, rule, lazy-learner, metaheuristic, and Bayesian families. The model was trained using the Topic Score feature vector, and the performance of the IBk algorithm was found better with a 43.4% average Recall. In this research, we present the Neural Network-based QuPiD attack model, and we tested its performance with both Topic Score feature vector and query string feature vector. The Topic Score feature vector's preference with the IBk algorithm is better as IBk uses the K-Nearest Neighbour method to classify the data. However, the RNN BiLSTM based QuPiD Attack's performance is better with query string feature vector; then, the IBk is suitable for numeric data.

6. Conclusion

Controlling private data is becoming increasingly important in today's world, especially in web searches, as users' queries can be used to infringe their privacy by third parties. This paper presents NN-QuPiD Attack: a Neural Network-based QuPiD Attack that measures the privacy provided by the famous Private Information Retrieval protocol Useless User Profile (UUP). NN-QuPiD Attack uses a user's profile or web history, and a Neural Network-based machine learning algorithm identifies the user of interest queries in an anonymized web search log. The experiments are conducted with the AOL benchmark dataset, while Multilayer Perceptron and Recurrent Neural Network (RNN) (LSTM and BiLSTM) are selected as classification algorithms. Moreover, experiments are also conducted with the IBk classification algorithm, as IBk performed well in the previous QuPiD Attack version.

We conducted the experiments with a basic Feedforward ANN algorithm, Multilayer Perceptron, using Topic Score and query string feature vector. The results show that the QuPiD Attack's performance with Multilayer Perceptron was found deprived both in terms of Precision and Recall with the Topic Score feature vector scenario. Next, we conducted experiments with Recurrent Neural Networks with LSTM and BiLSTM configuration under various epochs. The results show that the QuPiD Attack's performance with LSTM and BiLSTM is far better with query strings or textual data compared to the Topic Score in terms of Precision and Recall. The model built with BiLSTM gave 0.512 Recall with the Precision of 0.923, whereas IBk gave 0.49 Recall with the Precision of 0.934 with the same data.

The results show that the QuPiD Attack performance can be improved further using Artificial Neural Network techniques with fine-tuned parameters. This situation is alarming for currently available PIR protocols as they cannot provide adequate privacy to the users in QuPiD Attack. Therefore, it is recommended that future researches in the privacy preservation area should also consider the fact that PIR protocols are vulnerable to QuPiD Attack. Moreover, QuPiD Attack can be effectively used by the researchers as a privacy evaluation mechanism for future private information retrieval protocols.

Data Availability

The data used to support the findings of this study are available at http://www.radiounderground.net/aol-data/.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to thank the Researchers Supporting Project no. RSP-2020/286, King Saud University, Riyadh, Saudi Arabia, for supporting this project.

References

[1] R. Khan, A. Ahmad, A. O. Alsayed, M. Binsawad, M. A. Islam, and M. Ullah, "QuPiD attack: machine learning-based privacy quantification mechanism for pir protocols in health-related web search," *Scientific Programming*, vol. 2020, Article ID 8868686, 11 pages, 2020.

[2] R. Khan, M. A. Islam, M. Ullah, M. Aleem, and M. A. Iqbal, "Privacy exposure measure: a privacy-preserving technique for health-related web search," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 6, pp. 1196–1204, 2019.

[3] R. Khan, *On the effectiveness of private information retrieval protocols*, Ph.D. dissertation, Department of Computer Science, Capital University of Science and Technology, Islamabad, Pakistan, 2020.

[4] E. Adar, "User 4xxxxx9: anonymizing query logs," in *Proceedings of Query Log Analysis Workshop, International Conference on World Wide Web*, Alberta, Canada, May 2007.

[5] A. Petit, *Introducing privacy in current web search engines*, Ph.D dissertation, Université de Lyon, France, 2017.

[6] A. Arampatzis, G. Drosatos, and P. S. Efraimidis, "Versatile query scrambling for private web search," *Information Retrieval Journal*, vol. 18, no. 4, pp. 331–358, 2015.

[7] V. Toubiana, L. Subramanian, and H. Nissenbaum, "Trackmenot: enhancing the privacy of web search," arXiv: 1109.4677, 2011.

[8] S. B. Mokhtar, A. Boutet, P. Felber, M. Pasin, R. Pires, and V. Schiavoni, "X-search: revisiting private web search using intel sgx," in *Proceedings of the 18th ACM/IFIP/USENIX Middleware Conference*, pp. 198–208, Las Vegas, Nevada, December 2017.

[9] M. Ullah, M. A. Islam, R. Khan, M. Aleem, and M. A. Iqbal, "ObSecure Logging (OSLo): a framework to protect and evaluate the web search privacy in health care domain," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 6, pp. 1181–1190, 2019.

[10] C. Romero-Tris, A. Viejo, and J. Castellà-Roca, "Multi-party methods for privacy-preserving web search: ssurvey and contributions," in *Advanced Research in Data Privacy*, pp. 367–387, Springer, Berlin, Germany, 2015.

[11] M. Ullah, R. Khan, and M. A. Islam, "Poshida, a protocol for private information retrieval," in *Proceedings of 2016 Sixth International Conference on Innovative Computing Technology (INTECH)*, pp. 464–470, Dublin, Ireland, August 2016.

[12] M. Ullah, R. Khan, and M. A. Islam, "Poshida II, a multi group distributed peer to peer protocol for private web search," in *Proceedings of 2016 International Conference on Frontiers of Information Technology (FIT)*, pp. 75–80, Islamabad, Pakistan, December 2016.

[13] R. Khan and M. A. Islam, "Quantification of PIR protocols privacy," in *Proceedings of 2017 International Conference on Communication, Computing and Digital Systems (C-CODE)*, pp. 90–95, Islamabad, Pakistan, March 2017.

[14] R. Khan, M. Ullah, and M. A. Islam, "Revealing PIR protocols protected users," in *Proceedings of 2016 Sixth International Conference on Innovative Computing Technology (INTECH)*, pp. 535–541, Dublin, Ireland, August 2016.

[15] S. T. Peddinti and N. Saxena, "Web search query privacy: eevaluating query obfuscation and anonymizing networks1," *Journal of Computer Security*, vol. 22, no. 1, pp. 155–199, 2014.

[16] A. Petit, T. Cerqueus, A. Boutet et al., "SimAttack: private web search under fire," *Journal of Internet Services and Applications*, vol. 7, pp. 1–17, 2016.

[17] G. Kondrak, "N-gram similarity and distance," in *Proceedings of International Symposium on String Processing and Information Retrieval*, pp. 115–126, Buenos Aires, Argentina, November 2005.

[18] A. Petit, *Introducing privacy in current web search engines*, Ph.D. thesis, Universität Passau, Germany, 2017.

[19] A. Gervais, R. Shokri, A. Singla, S. Capkun, and V. Lenders, "Quantifying web-search privacy," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pp. 966–977, Scottsdale, ARI, USA, November 2014.

[20] E. Balsa, C. Troncoso, and C. Diaz, "OB-PWS: obfuscation-based private web search," in *Proceedings of 2012 IEEE Symposium on Security and Privacy*, pp. 491–505, San Francisco, CA, USA, May 2012.

[21] F. Bravo-Marquez, E. Frank, B. Pfahringer, and S. M. Mohammad, "AffectiveTweets: a Weka package for analyzing affect in tweets," *Journal of Machine Learning Research*, vol. 20, pp. 1–6, 2019.

[22] B. Wang, A. Wang, F. Chen, Y. Wang, and C.-C. J. Kuo, "Evaluating word embedding models: methods and experimental results," *APSIPA Transactions on Signal and Information Processing*, vol. 8, 2019.

[23] S. Lang, F. Bravo-Marquez, C. Beckham, M. Hall, and E. Frank, "Wekadeeplearning4j: a deep learning package for weka based on deeplearning4j," *Knowledge-Based Systems*, vol. 178, pp. 48–50, 2019.

[24] Y.-Y. Chen, Y.-H. Lin, C.-C. Kung, M.-H. Chung, and I.-H. Yen, "Design and implementation of cloud analytics-assisted smart power meters considering advanced artificial intelligence as edge analytics in demand-side management for smart homes," *Sensors*, vol. 19, no. 9, pp. 2047–2073, 2019.

[25] I. N. Yulita, M. I. Fanany, and A. M. Arymuthy, "Bi-directional long short-term memory using quantized data of deep belief networks for sleep stage classification," *Procedia Computer Science*, vol. 116, pp. 530–538, 2017.