DATA COMPARISON: APPLICATION OF PSI AND ZERO-KNOWLEDGE SCHEME

Solomon SARPONG[1]

[1]*Department of Physical and Mathematical Sciences, University of Environment and Sustainable Development, Somanya, Ghana*

---***---

Abstract - *The availability of electronic devices and the ubiquitousness of Internet services has made the interaction between; persons, electronic devices and persons and electronics devices inevitable. These interactions may lead to the exchange and or comparison of information. This paper has proposed privacy-preserving protocols that will enable persons with set of information communicate securely. The protocols in this paper are secure against honest-but curious and malicious attacks.*

Key Words: **privacy-preserving, zero-knowledge, unlinkability, cardinality, server privacy.**

1.INTRODUCTION

The availability of smartphones and the ubiquitousness of Internet services and its application to every facet of our lives has made privacy issues very important. Albeit, the society is increasingly dependent on and at the same time fearful of the availability of information. Data is being shared between devices at increasing rate. How do two or more parties without mutual trust exchange or compute the intersection between their sensitive information? It is in lieu of this that these protocols have been proposed: a). the use of third parties [1]–[3]; b). distributed technique [4], [5]; c). hybrid technique [6], [5], [7]–[12]

The need to exchange or compute of the intersection of the information brings to the fore the need for limited (privacy-preserving) sharing of sensitive information. Among these limited sharing of sensitive information protocols is the private set intersection (PSI) techniques. PSI finds application in scenarios where two parties wish to compute an intersection of their respective sets of items without revealing to each other any other information [13], [14]–[16]

We are in the information age and hence, the sharing of information is inevitable even though not done willingly at times. Usually, one of the parties will be seeking the information and the other maybe willing to share the information or is compelled to. The execution of the information sharing protocols such that, each of the parties get to know only what s/he is supposed to know is a problem in real world.

Such situation is encountered in aviation where the U.S. Department of Homeland Security (DHS) maintains a dynamic database of suspected terrorists (TWL: Terror Watch List). For every flight, DHS must perform privacy preserving set-intersection operation between its TWL and the passenger flight manifest of the airline to know if they have some names in common.

As a contribution to knowledge, this paper seeks to propose protocols that will enable persons with set of information exchange or find the intersection between their sets. At the end of the protocol, each of the parties get to know only what they are supposed to know but nothing else. The protocols are secure against honest-but curious and malicious attacks. The rest of the paper are as follows: there is discussion on private set intersection and zero-knowledge proof in sections 2 and 3 respectively; related works is in section 4; the proposed protocol is in section 5 and the conclusion is in section 6.

2. PRIVATE SET INTERSECTION (PSI)

How do two rival companies find the intersection of their private set of information such that: (i). only the intersection is known to each of them; (ii). apart from the intersection that is known, no other information is revealed. In such scenarios, PSI protocols and its variants becomes important. PSI is a cryptographic protocol that enables two parties, each with private set, compute the intersection of their sets such that apart from the intersection, no other information is revealed. In PSI protocol, each of the parties has a set of items and their aim is to compute the intersection of their sets items without leaking any other information.

Let S_A be the private set of company A and S_B the private set of company B. Their wish is to jointly compute the intersection of their private set such that both companies A and B know only $S_A \cap S_B$ but nothing more. A problem with PSI is that the inputs S_A and S_B can be chosen arbitrarily by the entities in the protocol [17], [18][13], [19], [20] [16], [21]–[23]

In order to prevent a dishonest party from inserting attributes s/he does not possess so as to know more information about the other, the attributes are certified/authorised. The use of authorized private set intersection (APSI) and its variants [24] [16], [23], [14], [15], [25] ensures that the attributes used by each party is certified. Hence, in APSI protocols only certified attributes are used. The certification of the attributes prevents dishonest behaviour by any malicious party.

PSI with data transfer (PSI-DT) is a variant of PSI. In PSI-DT, each item in the set has a database attached to it. Another variant of PSI is PSI cardinality, PSI-CA. In protocols using PSI-CA, only the magnitude of the intersection rather than its content is revealed. A variant of the PSI-CA is the authorised PSI-CA, where client input must be authorised by a mutually trusted authority.

The algorithms in PSI are *setup* and *interaction*. In the *setup,* all the parameters in the protocol are selected. In the *interaction,* there is protocol between the client and server. At the end of the protocol, the client obtains the output. Furthermore, APSI has three components – client, server and (offline) CA involved. APSI has three algorithms – *setup, authorize* and *interaction.* In the *setup,* global/public parameters are selected. *Authorize* allows the CA to issue authorizations (signatures) for each of the attributes of the client. In the *interaction*, the client obtains the intersection from the protocol between the attributes of the client and server.

2.1 Security Properties

Correctness, server and *client* privacy, and server and client unlinkability are the security requirements in PSI. *Correctness* – a PSI is correct when at the end of the protocol, the client outputs of the intersection between the attributes of server and client. Also, *correctness* in APSI entails a client outputting the intersection between the authorized attributes of the server and client. In *server privacy*, the client gets to know only the size of the intersection. Also, in *client privacy* the server gets to know only the size of the intersection. In both instances, a malicious server or client gets to know nothing. *Server privacy* in APSI entails the client getting to know only the size of the intersection between the authorized attributes of the server and client. *Client unlinkability* – a malicious server cannot identify any two instances of interaction are executed on the same inputs by the client. Also, in *server unlinkability*, the same input by the server cannot be distinguished by a client. [26]–[28].

3. ZERO-KNOWLEDGE PROOF (ZKP)

ZKP is a cryptographic protocol that enables one party (*prover*) to prove the knowledge of an information without disclosing it to the other party (*verifier*). That is a *'prover'* can prove the knowledge of certain information to a *'verifier'* without revealing the information. ZKP was proposed by [29]. In ZKP, properties such as; completeness, soundness, and zero-knowledge must be satisfied. The *completeness property* – a honest verifier will be convinced of the truth of a statement by an honest prover. The *soundness property* – except with some small probability, no cheating prover can convince the honest verifier that a false statement is true. The *zero knowledge property* states if the statement is true, no cheating verifier learns anything other than this fact [30], [31]–[36].

In ZKP, there is a *verifier*, *V* and a *prover*, *P*. A proof system (*P, V*) for a function is zero knowledge if for a verifier V^* there exists an efficient probabilistic algorithm S^* such that for every *x* such that $f(x) = 1$, the following random variables are computationally indistinguishable:

- The output of V^* after interacting with *P* on input *x*.

- The output of S^* on input *x*.

Hence, at the end of a zero-knowledge protocol the *verifier* does not know anything that s/he could not have known on his/her own.

4. RELATED WORKS

There was the application of Zero-knowledge in password authentication [37]; a P2P authentication; Diffie-Hellman key exchange; hash-based secure access control for mobile RFID systems. In order to secure and obtain mutual authentication between election authorities and voters, zero-knowledge based Diffie-Hellman key exchange algorithm was used, [38]. [39] proposed a protocol that integrates zero knowledge and Diffie-Hellman key exchange in order to authenticate and preserve the privacy between a network and persons trying to access it. [40] elucidates the uses of ZKP systems applications such as blockchain, zk-SNARKs, Zcash cryptocurrency.

In [14] there was the proposition of efficient protocols for private set intersection (PSI and APSI). They further postulated that, the choice between the usage of PSI or APSI depends on whether there is the need authorization and/or unlinkability. [41] proposed PSI protocols that has time and communication complexity which is linear to the size of the input sets. This protocol enables sampling from the intersection. [25] proposed PSI and authorized PSI (APSI) protocols that are secure in the malicious model under standard cryptographic assumptions. Their protocols have both linear communication and computational complexities. [15] designed an efficient cryptographic protocol for PSI-CA. Their protocol shows the size of the intersection set only and it is secure in the presence of both semi-honest and malicious adversaries.

5. THE PROTOCOL

This protocol helps users find similarity between two documents or the intersection between two sets of attributes. In the first part of the protocol, the initiator (company *A*) initiates the protocol to check the similarities between his/her documents and that of company *B*. If the initiator observes that there are enough common attributes between the two documents, then then proceed to exchange the common attributes.

This matchmaking protocol is based on variant of [15], [25]. In these protocols, to prevent malicious participants from manipulating the input set, APSI protocol is adopted so that all the input sets will be authorised. The input set of each member in the protocol is authorised by a mutually trusted certification authority, CA. The CA authenticates the attributes in the input set but does not take part in the protocol.

Algorithm 1 is a variant of [15]. This proposed APSI-CA protocol is secure against a malicious sever and semi-honest client. Let company *A* initiate the protocol to find the similarity between the document of company *B*. Assume company *A* has input $A = \{a_1, \dots, a_v\}$, which is certified by the certification authority by issuing $\sigma_i = H(a_i)^d$. Hence, the input becomes $\{(a_1, \sigma_1), (a_2, \sigma_2), \dots, (a_v, \sigma_v)\}$. Let the input of company *B* be, $B = \{b_1, b_2, \dots, b_n\}$ the certification authority issues $\mu_i = H(b_i)^d$, and after authentication, the input becomes $\{(b_1, \mu_1), (b_2, \mu_2), \dots, (b_n, \mu_n)\}$. A random number $R_A \leftarrow Z_{N/_2}$ is chosen by company *A*. Company *A* calculates $\alpha_i = (\sigma_i)^{2R_A}$, and broadcasts $\{\alpha_1, \alpha_2, \dots, \alpha_v\}$. Company *B* on receiving $\{\alpha_1, \alpha_2, \dots, \alpha_v\}$ computes, returns $\{t\alpha_{l_1}, \dots, t\alpha_{l_v}\}$ and $\{mb_1, \dots, mb_n\}$ to company *A*. Company *A* computes tb_j and further computes the absolute intersection $|\{t\alpha_{l_1}, \dots, t\alpha_{l_v}\} \cap \{mb_1, .\ \ .., mb_n\}|$. The absolute intersection computed enables company *A* know the upper limit of the number of attributes the document has in common with company *B*. A sufficient upper limit of common attributes with company *B* leads to the quest to know what those attributes are. The two companies execute Algorithm 2 to know the actual attributes they have in common.

Algorithm 2 is also a variant of [42]. This enables the companies to know the attributes they have in common. Company *A* computes M_i, N_i and $\pi = ZK\{R_{Ai}, i = 1, \dots, v | M_i^{2e}/N_i^2 = (g^e/g')^{2R_{Ai}}\}$ and sends to company *B*. There is the use of zero knowledge proof between the companies to verify π. The inability of company *A* to prove to company *B* the validity of π terminates the protocol. When π is valid, company *B* computes and sends Z, $\{M_i'\}, \{T_{Bj}\}$ and $\pi' = ZK\{R_B | Z = g^{2eR_B}, \forall i, M_i' = (M_i)^{2eR_B}\}$ to company *A*.

Using zero knowledge proofs, company *B* shows company *A* the validity of π'. However, the protocol is terminated on the inability of company *B* to prove the validity of π'. Company *A* computes $T_{Ai} = H_2(K_{Ai},\ ha_i,\ a_i)$ and sends to company *B*. Both compute and output $I_i \in A \cap B \text{ if there exits } i, j \text{ such that } T_{Ai} = T_{Bj}$. The intersection I_i enables them know the attributes they have in common. In order to make sure no one cheated, they send I_i to the certification authority, CA. The CA passes the protocol as successful if the verification of I_i from each of them is the same.

5.1 Security

The CA in this matchmaking protocol does not take part in it. The CA certifies the attributes used and monitors the protocol hence monitoring the protocol against cheating. Inputting security parameter *k* enables the CA to compute the RSA modulus $N = pq$ where $p = 2p' + 1, q = 2q' + 1$ and picks random elements g and g' such that $\langle -1 \rangle * \langle g \rangle \equiv \langle -1 \rangle * \langle g' \rangle \equiv Z_N^*$. Assume (e, d) are the RSA exponents: where *e* is a small prime,

$d = e^{-1} mod\ \emptyset(N)$ and g is a generator of QR_N. The (*p*, *q*, *d*) are the secret keys of the CA and the public parameters are $(N, e, g, g',\ H, H', H_1(\ \ ), H_2(\ \ ))$. The CA also finds hash functions such that, $H: \{0,1\}^* \rightarrow Z_N^*, H': \{0,1\}^* \rightarrow \{0,1\}^k, H_1: \{0,1\}^* \rightarrow Z_N^*$ and $H_2: Z_N^* * Z_N^* * \{0,1\}^* \rightarrow \{0,1\}^k$.

The CA certifies the input attributes by computing: $\text{certified input element} = H(\text{input element})^d mod\ N$.

The security of algorithms in this protocol evaluated based on these parameters: *Correctness*, *privacy and efficiency*. *Correctness* – company *A* possesses (a_i, σ_i) and company *B*, $\left(b_{j,} \mu_j\right)$. If: (1) for genuine σ_i the CA signature on a_i and (2) $a_i = b_j$, hence, $ha_i = hb_j$, and $ta_{l_i} = H'((\sigma_i)^{2eR_AR_B}) = H'((ha_i)^{2R_AR_B}) = tb_j$. *Privacy* – at the end of the protocol, both learn the size of the intersection whilst an adversary learns nothing. *Efficiency* – the protocol in this paper incurs linear computation and communication complexity for both parties. Client performs $O(w)$ modular exponentiations and server does $O(w+v)$ modular exponentiations.

Also, algorithm 2 [42] is secure in a malicious model in ROM under the RSA and DDH assumptions. With the hardness of RSA and DDH problems, and π, π' are zero-knowledge proofs, hence the protocol is a secure computation in ROM. The protocols in this paper is secure against semi-honest attacks. Furthermore, the protocol is secure against be external attacks. As an added security, the protocol is symmetric hence the output of the protocol is known by companies *A* and *B* simultaneously.

6. CONCLUSIONS

The use of smartphones and popularity of mobile social networks is on the ascendency. It has become imperative for practical and secured protocols that will enable users to interact effectively. This protocol uses certified sets and zero knowledge proofs that enables users to adequately find similarities between documents or attributes without leaking any private information. The certification of the private sets helps to reduce honest-but curious behaviours from the protocol users. On the other hand, the use of zero-knowledge helps the preservation of the secrecy of the private set of attributes during the protocol. Hence, this protocol can help users to find similarities between their documents or attributes without leaking any other information.

REFERENCE

[1] N. Eagle and A. Pentland, “Social Serendipity: Mobilizing social software,” *IEEE Pervasive Computing, Special Issue: The Smartphone*, pp. 28–34, 2005.

[2] J. Kjeldskov and J. Paay, “Just-for-us: a context-aware mobile information system facilitating sociality,” in *Proceedings of the 7th international conference on Human computer interaction with mobile devices \& services*, 2005, pp. 23–30.

[3] A.-K. Pietiläinen, E. Oliver, J. LeBrun, G. Varghese, and C. Diot, “Mobiclique: middleware for mobile social networking,” in *Proceedings of the 2nd ACM workshop on Online social networks*, 2009, pp. 49–54.

[4] M. Li, N. Cao, S. Yu, and W. Lou, “FindU: Privacy-preserving personal profile matching in mobile social networks,” *Proceedings - IEEE INFOCOM*, no. 1, pp. 2435–2443, 2011, doi: 10.1109/INFCOM.2011.5935065.

[5] R. Lu, X. Lin, and X. Shen, “SPOC: A secure and privacy-preserving opportunistic computing framework for mobile-healthcare emergency,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 3, pp. 614–624, 2013, doi: 10.1109/TPDS.2012.146.

[6] E. De Cristofaro, A. Durussel, and I. Aad, “Reclaiming privacy for smartphone applications,” *2011 IEEE International Conference on Pervasive Computing and Communications, PerCom 2011*, no. March, pp. 84–92, 2011, doi: 10.1109/PERCOM.2011.5767598.

[7] S. Sarpong and C. Xu, “A secure and efficient privacy-preserving matchmaking for mobile social network,” in *International Conference on Computer, Network Security and Communication Engineering (CNSCE)*, 2014, pp. 362–366.

[8] Y. Wang, T. T. Zhang, H. Z. Li, L. P. He, and J. Peng, “Efficient privacy preserving matchmaking for mobile social networking against malicious users,” in *Proc. of the 11th IEEE Int. Conference on Trust, Security and Privacy in Computing and Communications, TrustCom-2012 - 11th IEEE Int. Conference on Ubiquitous Computing and Communications, IUCC-2012*, 2012, pp. 609–615. doi: 10.1109/TrustCom.2012.142.

[9] Q. Xie and U. Hengartner, “Privacy-Preserving Matchmaking For Mobile Social Networking Secure Against Malicious Users,” 2011.

[10] S. Sarpong and C. Xu, “Privacy-preserving attribute matchmaking in proximity-based mobile social networks,” *International Journal of Security and its Applications*, vol. 9, no. 5, pp. 217–230, 2015, doi: 10.14257/ijsia.2015.9.5.22.

[11] S. Sarpong, C. Xu, and X. Zhang, “PPAM: Privacy-preserving Attributes Matchmaking Protocol for Mobile Social Networks Secure against Malicious Users,” 2016.

[12] S. Sarpong, C. Xu, and X. Zhang, “An Authenticated Privacy-preserving Attribute Matchmaking Protocol for Mobile Social Networks,” 2015.

[13] G. Ateniese, E. De Cristofaro, and G. Tsudik, “(if) size matters: Size-hiding private set intersection,” in

International Workshop on Public Key Cryptography, 2011, pp. 156–173.

[14] E. De Cristofaro and G. Tsudik, "Practical private set intersection protocols with linear complexity," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6052 LNCS, pp. 143–159, 2010, doi: 10.1007/978-3-642-14577-3_13.

[15] E. De Cristofaro, P. Gasti, and G. Tsudik, "Fast and Private Computation of Cardinality of Set Intersection and Union," in *Cryptology and Network Security*, J. Pieprzyk, A.-R. Sadeghi, and M. Manulis, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 218–231.

[16] E. Stefanov, E. Shi, and D. Song, "Policy-Enhanced Private Set Intersection: Sharing Information While Enforcing Privacy Policies," in *Public Key Cryptography -- PKC 2012*, M. Fischlin, J. Buchmann, and M. Manulis, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 413–430.

[17] L. Kissner, "Privacy-Preserving Set Operations," no. June, 2006.

[18] L. Kissner and D. Song, "Privacy-Preserving Set Operations," in *Advances in Cryptology -- CRYPTO 2005*, V. Shoup, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 241–257.

[[19] G. Ateniese, D. Francati, D. Nuñez, and D. Venturi, "Match Me if You Can: Matchmaking Encryption and Its Applications," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11693 LNCS, pp. 701–731, 2019, doi: 10.1007/978-3-030-26951-7_24.

[20] M. J. Freedman, K. Nissim, and B. Pinkas, "Efficient private matching and set intersection," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3027, pp. 1–19, 2004, doi: 10.1007/978-3-540-24676-3_1.

[21] D. Dachman-Soled, T. Malkin, M. Raykova, and M. Yung, "Efficient robust private set intersection," *International Journal of Applied Cryptography*, vol. 2, no. 4, pp. 289–303, 2012, doi: 10.1504/12.48080.

[22] C. Hazay and Y. Lindell, "Efficient protocols for set intersection and pattern matching with security against malicious and covert adversaries," in *Theory of Cryptography Conference*, 2008, pp. 155–175.

[23] E. Stefanov, E. Shi, and D. Song, "Policy-enhanced private set intersection: Sharing information while enforcing privacy policies," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, pp. 413–430. doi: 10.1007/978-3-642-30057-8_25.

[24] J. Camenisch and G. M. Zaverucha, "Private intersection of certified sets," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, pp. 108–127. doi: 10.1007/978-3-642-03549-4_7.

[25] E. De Cristofaro, J. Kim, and G. Tsudik, "Linear-Complexity Private Set Intersection Protocols Secure in Malicious Model *".

[26] J. Camenisch and G. M. Zaverucha, "Private Intersection of Certified Sets," in *Financial Cryptography and Data Security*, R. Dingledine and P. Golle, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 108–127.

[27] E. N. Zhang, J. Chang, and Y. U. Li, "Efficient Threshold Private Set Intersection," vol. 9, 2021, doi: 10.1109/ACCESS.2020.3048743.

[28] K. Zhang and R. Needham, "A Private Matchmaking Protocol," 2001, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.54.835&rep=rep1&type=pdf

[29] Goldwasser S., and Micali S., and Rackoff C., "The Knowledge Complexity of Interactive Proof Systems," *Society of Applied and Engineering Science*, vol. 18, no. 1, pp. 186–208, 1989.

[30] A. Mohr, "A survey of zero-knowledge proofs with applications to cryptography," *Southern Illinois University, Carbondale*, pp. 1–12, 2007, [Online]. Available: http://austinmohr.com/Work_files/zkp.pdf

[31] Rafael. Pass and Institutionen för numerisk analys och datalogi (Stockholm), *Alternative variants of zero-knowledge proofs*. 2004.

[32] D. Catalano, M. Di Raimondo, D. Fiore, and M. Messina, "Zero-knowledge sets with short proofs," *IEEE Trans Inf Theory*, vol. 57, no. 4, pp. 2488–2502, 2011, doi: 10.1109/TIT.2011.2112150.

[33] J. Gehrke, E. Lui, and R. Pass, "Towards privacy for social networks: A zero-knowledge based definition of privacy," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011, pp. 432–449. doi: 10.1007/978-3-642-19571-6_26.

[34] "CiteSeerX — Citation Query A Survey of Zero-Knowledge Proofs with Applications to Cryptography." Accessed: Mar. 13, 2021. [Online]. Available: http://citeseerx.ist.psu.edu/showciting?doi=10.1.1.95.3828.

[35] G. I. Simari, "A Primer on Zero Knowledge Protocols," *DisClosure*, pp. 1–12, 2002.

[36] J. Groth, R. Ostrovsky, and A. Sahai, "Perfect non-interactive zero knowledge for NP," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2006, pp. 339–358. doi: 10.1007/11761679_21.

[37] J. Kurmi and A. Sodhi, "A Survey of Zero-Knowledge Proof for Authentication," 2015. [Online]. Available: https://www.researchgate.net/publication/316492793

[38] M. K. Ibrahem and M. Khalel Ibrahem, "Robust Electronic Voting System Using Homomorphic Encryption Protocol and Zero-Knowledge Proof E-Election Using Homomorphic Encryption and Zero-Knowledge Proof View project Electronic Voting System View project Robust Electronic Voting System using Homomo," *International Journal of Enhanced Research in Science*, vol. 5, no. January 2016, pp. 2319–7463, 2016, [Online]. Available: https://www.researchgate.net/publication/327859875

[39] M. Khalel Ibrahim, "Modification of Diffie-Hellman Key Exchange Algorithm for Zero Knowledge Proof," 2012. [Online]. Available: www.pdffactory.com

[40] J. Hasan, "Overview and Applications of Zero Knowledge Proof (ZKP)," 2019. [Online]. Available: www.IJCSN.org

[41] T. Beauregard and J. 2021 Xia, "Private set intersection: problems on sampling from the intersection," pp. 1–34, 2021.

[42] E. De Cristofaro, J. Kim, and G. Tsudik, "Linear-Complexity Private Set Intersection Protocols Secure in Malicious Model," in *Advances in Cryptology - ASIACRYPT 2010*, M. Abe, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 213–231.