Enhancing Privacy and Security of Customer Data Management in E-Commerce*

Linas Ablonskis[1,*,†], Rimantas Butleris[2,†], Audrius Lopata[1,†], Giedrius Sirevičius[3,†], Simonas Stepanovas[3,†] and Mindaugas Vasiljevas[3,4,†]

[1] *Department of Information Systems, Kaunas University of Technology, Studentų g. 50, LT-51368 Kaunas, Lithuania.*

[2] *Centre of Information Systems Design Technologies, Kaunas University of Technology, Studentų g. 50, LT-51368 Kaunas, Lithuania.*

[3] *UAB ICYBIT Savanorių pr. 321C, LT-50120 Kaunas, Lithuania.*

[4] *Department of Software Engineering, Kaunas University of Technology, Studentų g. 50, LT-51368 Kaunas, Lithuania.*

Abstract

We present a novel system for enhancing privacy and security of personally identifiable customer data in the domain of e-commerce. This is achieved by storing authentic customer data in a trusted third party, called a mediator, and replacing it with synthetic, yet convincing substitutes in the systems of e-commerce participants. Our system of customer data storage and access is designed to be retrofittable onto existing e-commerce platforms.

Keywords

e-commerce, personally identifiable information, privacy, security, obfuscation

1. Introduction

Privacy and security of personally identifiable customer data is an important feature of modern e-commerce systems. From a psychological and marketing perspective, there is a body of research indicating that customers value their privacy and prefer to use systems that do the same [1][2][3][4]. From a legal perspective, there is the EU General Data Protection Regulation [5] mandating strong safeguards for personal data privacy and security. In addition, there is the EU Data Governance Act [6] that defines intermediaries for sensitive data management on behalf of third parties.

It therefore becomes increasingly important to embed privacy-preserving designs into new e-commerce systems and to retrofit privacy-preserving designs onto existing e-commerce systems. In this paper we present a system for enhancing privacy and security of personally identifiable customer data in the domain of e-commerce, designed to be retrofittable onto existing e-commerce platforms. This system represents one possible embodiment of the method disclosed by [7].

The rest of this paper is organized as follows. Section 2 gives a review of related works. A description of our solution is provided in Section 3. Section 4 contains an example of how our solution might be applied in the context of an e-commerce system. Conclusions are drawn in Section 5.

* *IVUS2025: Information Society and University Studies 2024, May 15, Kaunas, Lithuania*

[1*] Corresponding author.

[†] These authors contributed equally.

linas.ablonskis@ktu.lt (L. Ablonskis); rimantas.butleris@ktu.lt (R. Butleris); audrius.lopata@ktu.lt (A. Lopata); mindaugas.vasiljevas@ktu.lt (M. Vasiljevas);

2. Related works

In the domain of e-commerce, privacy is usually equated with a disconnect between the digital and physical identity of a customer. It is preferable to maintain this disconnect in all e-commerce activities. There are works related to fully pseudonymous identities, with some of the latest being [8][9][10][11]. These kinds of systems are, however, not suited for usual e-commerce activities, for a full disconnect between digital and physical identity in e-commerce is often infeasible to achieve due to legal, technical or process constraints.

There are various identity management approaches, generally seen as falling into one of three domains: centralized, federated and user-centric [12]. Recent works add a self-sovereign identity domain [13], which, in principle, is equivalent to user-centric identity management with additional control given to the user. Some examples of centralized identity management include [14][15][16], and of federated identity management include [17][18][19][20]. User-centric identity management can be represented by works [21][22][23]. In this regard, our solution is at the intersection of federated and user-centric identity management approaches.

To limit disclosure of user data to authorized parties only, the following techniques are used: a) access control [24][25][26], b) obfuscation [27], and c) encryption [28][29]. Our solution uses all of these techniques in combination. To ensure proper use and tracking of disclosed sensitive data, the following techniques are employed: a) disclosure logging, b) disclosure policies, including sticky policies [30], and c) cryptographically signed disclosure contracts. Our solution uses all of these techniques, excluding the use of sticky policies.

3. Description of the system

Our method is primarily concerned with the obfuscation of personally identifiable information (PII) of a customer under an e-commerce domain with business-to-customer (B2C) orientation.

Under a typical B2C-oriented e-commerce domain, there are two distinct types of parties: a customer and a service provider. A service provider can be a merchant, a courier or some other party involved in the customer service provision chain. Service providers have to store, process and exchange customer-related information. We introduce an additional party, called a mediator, that is responsible for the obfuscation and management of the PII of a customer in relation to one or more service providers.

To avoid or minimize storage of the PII of a customer on the systems of a service provider, we propose for PII on the service provider side, where possible, to be stored and consumed obfuscated. By obfuscated we mean that authentic values of PII attributes of a customer, such as their name, surname, address, phone number, email address, etc., would be replaced with synthetic, yet normally appearing substitutes. The authentic values would be stored on the mediator side, encrypted in a manner that prevents any one party, including the mediator, from being able to decrypt them on their own. In addition, the authentic values would be accessible through cryptographically signed digital contracts only. Finally, such obfuscation is to be applied per service-provider, whereby the same physical customer would have different synthetic values of obfuscated attributes in the systems of different service providers.

Let us consider a service provider that wishes to adopt our scheme and start using the services of the mediator. The first step would be to obfuscate the PII of every customer already present on the service provider system, as shown in Figure 1.

Authentic values of PII attributes are stored encrypted under a symmetric encryption scheme with split keys. For every new service provider joining the system, two keys are generated: a service provider key and a mediator key for that service provider. These keys are combined to derive a composite key for encryption / decryption of data at rest w.r.t. that specific service provider. The mediator discards the service provider key and composite key at the end of every

interaction, data at rest in the mediator system cannot be accessed without the active participation of both the service provider and the mediator.

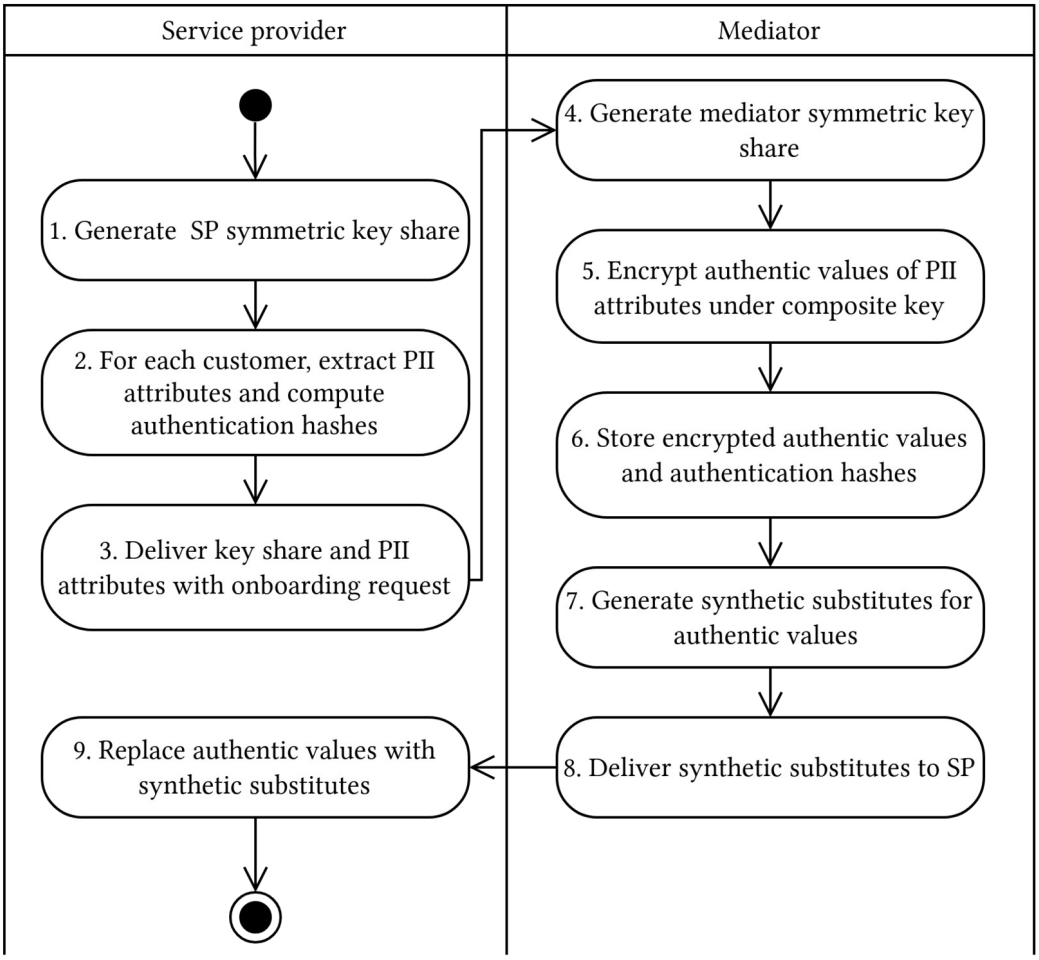

Figure 1: Service provider (SP) onboarding process.

For every customer within the service provider database, a set of PII attribute name-value pairs are extracted. The name-value pairs are sent to the mediator. The mediator encrypts the values under the composite key and stores them internally, with reference to the service provider. The original values are then discarded. The mediator also generates synthetic substitute values for each name-value pair that are convincing enough to be used for obfuscation. The synthetic values are also generated in such a way that makes them mathematically unrelated to their authentic counterparts. The substitute name-value pairs are then sent back to the service provider. The service provider replaces the original PII of the customer with substitutes and deletes the original PII.

For PII fields that are used to identify the customer within service provider systems, the mediator also stores cryptographic hashes of the original values. These hashes are needed for the mediator to be able to map authentic identifiers of the customer on the service provider systems to the synthetic identity without having access to the service provider key. This becomes necessary whenever the customer tries to authenticate with the service provider system using their authentic identifier.

All communication between the parties of our scheme must be conducted over secure communication channels that may be implemented using standard public key infrastructure (PKI) primitives together with a secure communication protocol, such as HTTPS [31]. The mediator

system also contains a separate secure environment that is used to perform cryptographic operations and to work with the shares of symmetric cryptographic keys, as shown in Figure 2.

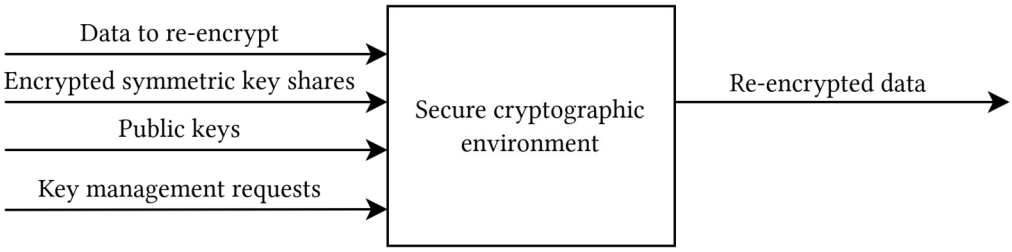

Figure 2: Secure cryptographic environment of the mediator.

Mediator key shares for symmetric encryption must never leave the secure cryptographic environment. In addition, to protect the key shares and any authentic values of PII attributes supplied by service providers from being intercepted in the intermediate nodes of the system, an additional layer of PKI-based hybrid encryption must be employed, e.g. as in [32]. That is to say, all sensitive input addressed to the secure cryptographic environment must be encrypted under the public key of the mediator, and all sensitive output produced by the secure cryptographic environment must be encrypted under the public key of the addressed service provider.

To allow for a customer to authenticate with the platform of the onboarded service provider using their authentic identifier, an authentication shim in the service provider platform would be needed. Such authentication shim would supply a hash of the login name of the customer to the mediator and ask for the corresponding synthetic login name to be resolved. The mediator is able to perform such resolution by comparing the provided hash to the hashes stored in its database with reference to the onboarded service provider.

If a service provider needs to update an authentic value of any attribute of the customer inside the mediator database, they may do so by repeating the onboarding process in part and in relation to that customer. In particular, the steps relating to the generation of symmetric key shares would be skipped, whereas every other step would follow the same procedure as shown in Figure 1.

Use of synthetic substitutes of PII attributes of a customer is not always possible within the domain of e-commerce due to legal, technical or process constraints. Accordingly, at some point a service provider might need to de-obfuscate some of the PII attributes of a specific customer. The procedure of retrieval of authentic values of attributes from the mediator is shown in Figure 3.

The service provider submits a request to the mediator for the retrieval of authentic values of select PII attributes of a customer having a specific synthetic identifier. The synthetic identifier of the customer uniquely identifies that customer in relation to the service provider inside the mediator system.

The request is shaped as a timestamped and cryptographically signed structured document. It contains a set of attributes of those authentic values that are being requested, the reason for the request and a set of terms and conditions (T&C) chosen from a predefined set supported by the mediator. The service provider also supplies its share of the symmetric encryption key needed to decrypt the authentic values inside the mediator database. The mediator verifies the timestamp, content and validity of the signature of the request and logs it. It then reconstitutes the full symmetric encryption key by combining the service provider key share with the mediator key share. The mediator then retrieves the authentic values of requested attributes, decrypts them and sends them to the service provider. The service provider key share, the composite key and the decrypted authentic values are then immediately discarded by the mediator.

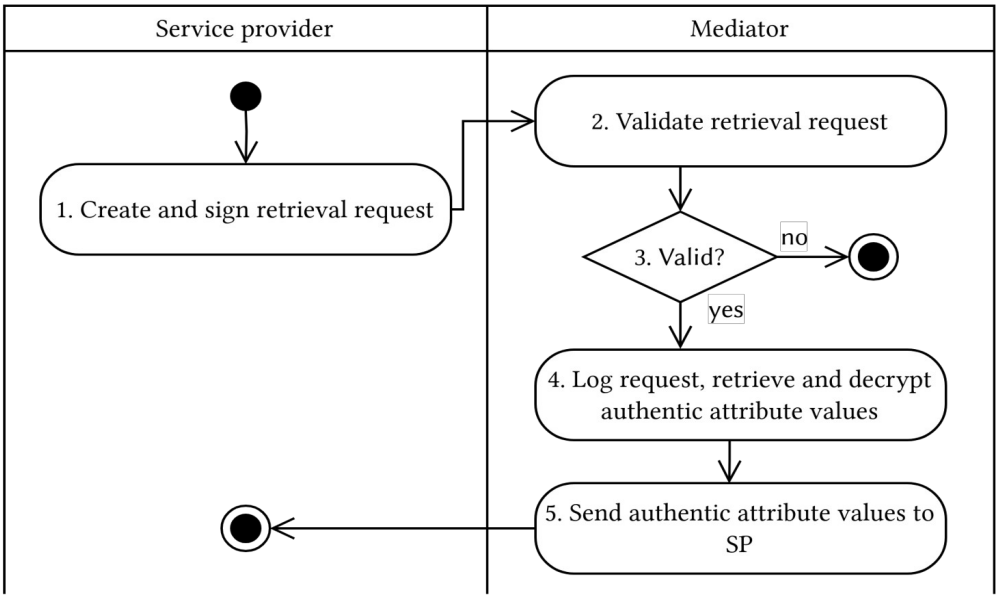

Figure 3: Retrieval of customer authentic attribute values.

The above-described requirements for the request document help to create a non-repudiable log of access to the authentic data of a customer by the systems of a specific service provider. They also help enforce proper use of authentic data in the implementation of the service provider system, since the implementer will have to consciously think about the T&C of use when shaping the request to the mediator system. In addition, authentic values of the PII attributes of the customer are only observed by those parts of the service provider system that are in absolute need of them to function. Consequently, all of the other parts of the system may continue to operate using synthetic data instead. Finally, if authentic data needs not be saved, it is to be discarded immediately after use.

Within the e-commerce domain, customer data might also have to be exchanged between various service providers, e.g. between a merchant and a courier. The scheme for service provider to service provider data exchange is shown in Figure 4.

If both the source and target service providers have a specific customer onboarded, the problem of data exchange is converted into a problem of synthetic substitute mapping between the obfuscation domains of the involved service providers. The mediator is able to solve the latter problem by comparing hashes of the values of customer identifying attributes in both domains. Once a customer synthetic substitute is remapped from the source service provider to the target service provider, the target service provider becomes able to work with synthetic data of the customer as well as to access the authentic data of the customer as described above.

If the target service provider does not have the specific customer onboarded, customer data is then transferred into the domain of the target service provider. In order to perform the transfer, both service providers have to supply their key shares to the mediator. The mediator then reconstitutes the composite keys and employs them to copy the authentic data from the storage domain of the source service provider to that of the target service provider. The synthetic substitutes of the attributes are also created and hashes of identifying attributes are transferred as well. In principle, the above is a repetition of the customer onboarding process, wherein the provider of the authentic values of PII attributes is the source service provider. Upon onboarding the customer with the obfuscation domain of the target service provider, the target service provider may continue to work with the customer data as usual.

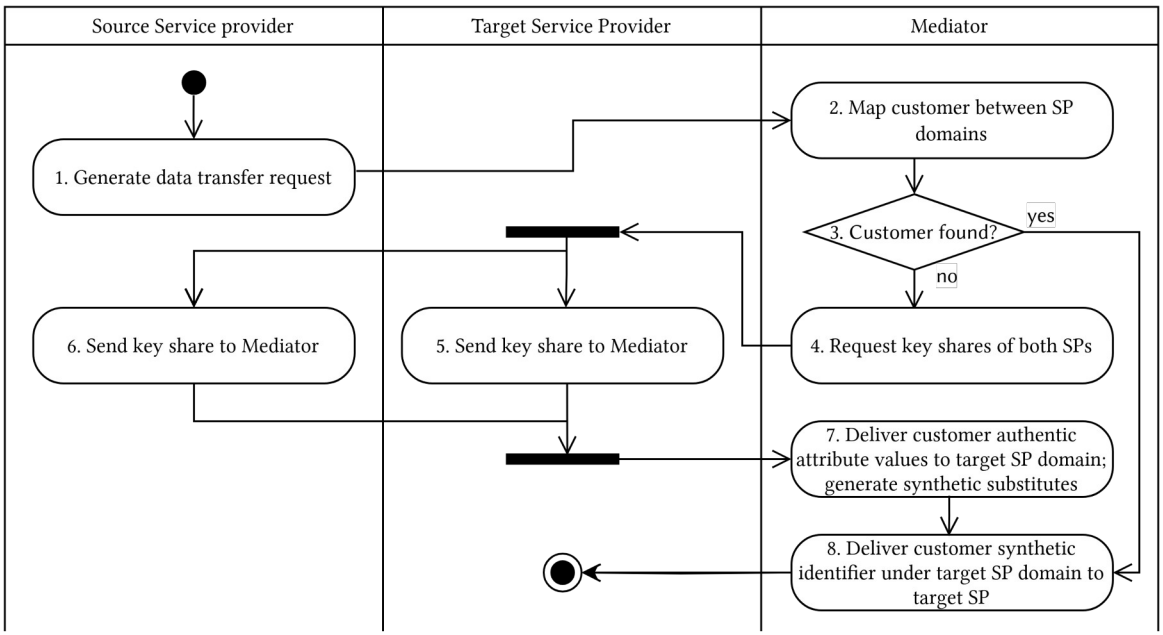

Figure 4: Transfer of customer data between service providers.

4. Motivating example

In this section we present a motivating example that describes how our system might be implemented within the context of a merchant platform in practice.

4.1. Merchant system database

We illustrate the usability of our system on a hypothetical merchant system based on the TPC-DS schema [33][34]. A modified version of the TPC-DS schema is presented in Figure 5. Notably, we consider the *Customer* and related tables only. We further introduce attributes *c_password_hash*, *c_salt*, and *c_telephone_number* so as to illustrate some of the everyday business cases of the merchant system. For simplicity, we discard attributes that are irrelevant to our study.

We consider that the merchant system chooses to obfuscate a minimum set of attributes. The minimum set of attributes in the TPC-DS schema would comprise: *c_first_name*, *c_last_name*, *c_login*, *c_email_address*, *c_telephone_number*, *c_birth_day*, *c_birth_month*, *c_birth_year*, *ca_street_number*, *ca_street_name*, and *ca_zip*. All attributes chosen from TPC-DS meet the requirement of their ability to directly identify the customer in different contexts. We exclude the primary key *c_customer_sk* and foreign keys *c_current_cdemo_sk*, *c_current_hdemo_sk*, *c_current_addr_sk*, *c_first_shipto_date_sk*, *c_first_sales_date_sk*, *c_last_review_date_sk* from the minimum set of attributes due to their inability to be used as direct identifiers outside of the context of TPC-DS without auxiliary data. There may be doubt as to whether *c_login* is an identifier outside of TPC-DS. Nevertheless, we opt to include it, for logins are usually created by customers themselves or comprise other global identifiers (e.g. email addresses), therefore, one cannot be sure if they are used only within the current context. We further exclude attributes *ca_city*, *ca_country* and *c_birth_country* from the set in order to increase the utility of some of the business analytics queries. Finally, we exclude both *c_password_hash* and *c_salt* from the set.

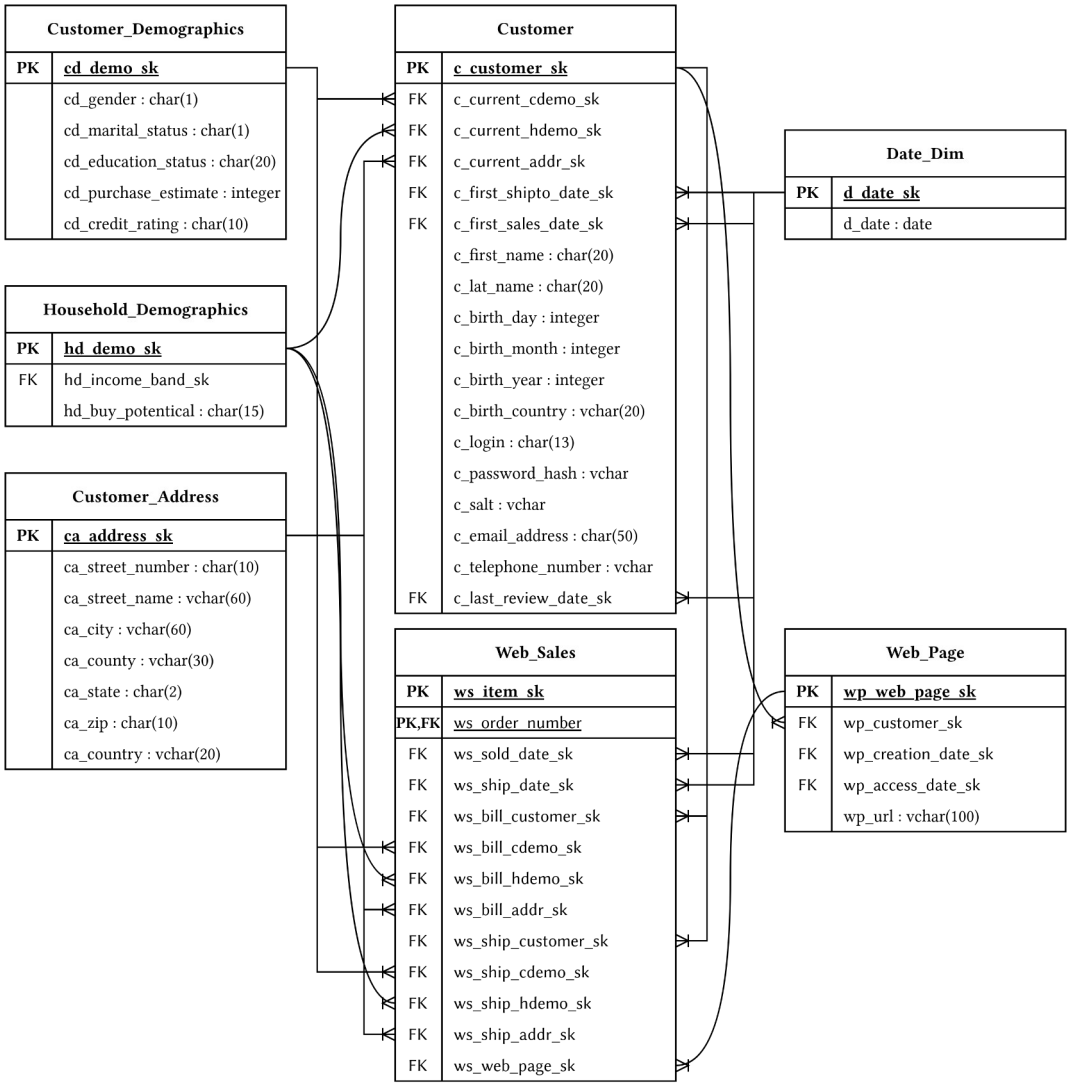

Figure 5: Merchant system database based on TPC-DS.

4.2. Mediator system database

The mediator system database is depicted in Figure 6. Customer attributes are stored in the table *User_Attributes_Authentic*, wherein the 3-tuple (*identifier*, *type*, *value*) is unique. The field *type* refers to the customer attribute name in the merchant database. The field *value* is the value of the attribute *type* within the merchant database. The *value* is encrypted using AES256-GCM [35], such that the encryption key is divided into parts and is distributed between the mediator and the merchant in accordance with an SSS scheme [36]. The attributes *fingerprint_mediator* and *fingerprint_organisation* store fingerprints of key shares of the (2,2)-threshold scheme being used.

Merchant information is stored in the table *Organisations*. The attribute *identifier* in this table serves as the primary key. The attribute *fingerprint* stores the fingerprint of the merchant key share. A record in the table *Organisations* is created when the merchant onboards onto the mediator system.

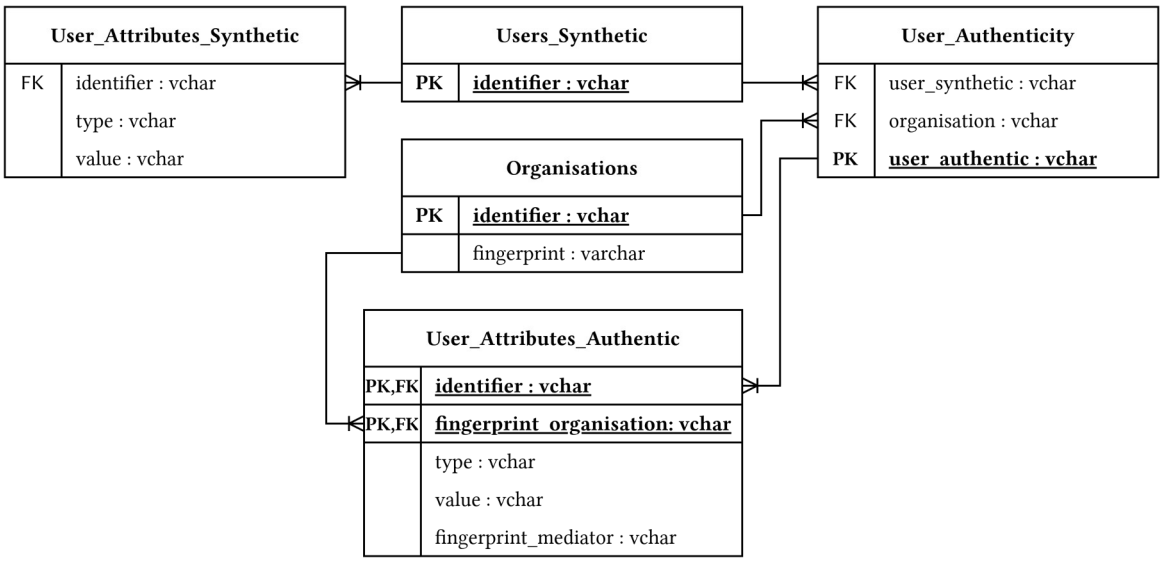

Figure 6: Mediator system database.

The table *User_Authenticity* maps merchants, authentic users and synthetic users, such that the 2-tuple (*organisation*, *user_authentic*) and the 2-tuple (*organisation*, *user_synthetic*) are unique. The attribute *user_authentic* is a primary key, output by a unique identification function. The unique identification function is executed when the merchant onboards a user onto the mediator system.

The synthetic user identifiers are stored in the table *Users_Synthetic*. They are obtained by supplying synthetic user attributes as input to the aforementioned unique identification function.

The synthetic user attributes are stored in the table *User_Attributes_Synthetic*. The 3-tuple (*identifier*, *type*, *value*) is unique. The *type* corresponds to the user attribute name in the merchant database. The *value* stores a randomly, but convincingly, generated synthetic value for that attribute.

4.3. Business decision support

Use of the mediator system does not preclude business analytics, however, some of the business decision support queries might involve de-obfuscation of some attributes. In general, the more attributes are obfuscated, the more often de-obfuscation operations will be involved in the queries. Therefore, the ability to execute business decision support queries without de-obfuscation depends on the level of obfuscation employed. On the other hand, certain business activities such as direct marketing would always involve de-obfuscation, because customer contact information belongs to the minimum set of attributes to be obfuscated.

TPC-DS Benchmark Standard Specification Version 3.2.0 describes 99 queries (i.e. business questions), 47 of which involve customer data [37]. Only 5/47 of the aforementioned queries of the TPC-DS benchmark would require de-obfuscation under the assumption that the minimum set of attributes was obfuscated. Those queries are B.18, B.19, B.24, B.30 and B.45. These queries require de-obfuscation because they need to retrieve customer data in relation to authentic time frames or locations.

5. Conclusions

We have presented a novel system for enhancing the privacy and security of customer PII within the domain of e-commerce. Our system offers several benefits: for the customer - increased privacy and security; for the service provider - stronger compliance with customer data protection regulations.

Retrofitting or implementing our system onto a system of a service provider will naturally partition the service provider system into parts that require strict security and parts where security can be relaxed. If authentic customer data has to be saved in the service provider system, for example, as part of an invoice, it may be stored in a secured and restricted environment that is inaccessible to less restricted environments. All other parts of the service provider system will either work with synthetic data or discard authentic data immediately after use. As a result, parts of the service provider system that use synthetic customer data would be unable to observe and leak authentic customer data.

Finally, with regards to obfuscation, we continue to provide all of the attributes of the customer in a form that appears to be normal, such that they would continue to be compatible with any legacy systems that were in need of such attributes. Therefore, implementation of our approach onto existing systems would not necessitate a total overhaul of customer representations across all of the subsystems on the service provision chain.

Acknowledgments

This paper presents the preliminary results of the research project financed by the European Union funds for the years 2021-2027 under the Economic Transformation and Competitiveness Development Programme of the Ministry of Economy and Innovation of the Republic of Lithuania, which manages the 2022-2030 development programme, Progress Measure No. 05-001-01-05-07 "Create a consistent system for promoting innovative activities" of the activity "Promote innovation supply" sub-activity "Invest in new APV product development activities and create conditions for researchers to participate in R&D activities of companies, promote intellectual property, early pilot production of new products created, and market preparation" (Central and Western Lithuania region).

The intellectual and technical concepts contained herein are patent-pending with reference to International Patent Application No. PCT/EP2024/078458.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] H. Nissenbaum, Protecting privacy in an information age: The problem of privacy in public, The ethics of information technologies, Routledge, 2020. 141-178.

[2] N. Gerber, P. Gerber, M. Volkamer, Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior, Computers & security 77, 2018, pp. 226-261.

[3] W.N. Price, I. G. Cohen, Privacy in the age of medical big data, Nature medicine 25.1, 2019, pp. 37-43.

[4] S. Zheng, et al, User perceptions of smart home IoT privacy, Proceedings of the ACM on human-computer interaction 2. CSCW, 2018, pp. 1-20.

[5] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2025, URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504

[6] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), 2025, URL: https://eur-lex.europa.eu/eli/reg/2022/868/oj/eng

[7] G. Sirevicius, S. Stepanovas, Mediation of Information Perception, 2025, International Patent Application No. PCT/EP2024/078458, World Intellectual Property Organization (WIPO).
[8] A. Lehmann, Scrambledb: Oblivious (chameleon) pseudonymization-as-a-service, Proceedings on Privacy Enhancing Technologies, 2019.
[9] J. Camenisch, A. Lehmann, Privacy-preserving user-auditable pseudonym systems, In 2017 IEEE European Symposium on Security and Privacy (EuroS&P), 2017, pp. 269-284, IEEE.
[10] A. V. Kayem, N. J. Podlesny, C. Meinel, A. Lehmann, On Chameleon Pseudonymisation and Attribute Compartmentation-as-a-Service, In SECRYPT, 2021, pp. 704-714.
[11] S. Casacuberta, J. Hesse, A. Lehmann, SoK: oblivious pseudorandom functions, In 2022 IEEE 7th European Symposium on Security and Privacy (EuroS&P), 2022, pp. 625-646, IEEE.
[12] D. Pöhn, W. Hommel, An overview of limitations and approaches in identity management, In Proceedings of the 15th International Conference on Availability, Reliability and Security, 2020, pp. 1-10.
[13] A. Tobin, D. Reed, The inevitable rise of self-sovereign identity, The Sovrin Foundation, 2016, URL: https://sovrin.org/wp-content/uploads/2017/06/The-Inevitable-Riseof-Self-Sovereign-Identity.pdf
[14] H. Roßnagel, A mechanism for discovery and verification of trust scheme memberships: The LIGHTest Reference Architecture. In Open Identity Summit 2017, 2017, pp. 81-92.
[15] U. Gerber, Authentication and Authorization for Constrained Environments, PhD thesis, Master's thesis, University of Zurich, 2018.
[16] Z. K. Zhang, M. C. Y. Cho, Z. Y. Wu, S. W. Shieh, Identifying and authenticating IoT objects in a natural context, Computer 48(08), 2015, pp. 81-83.
[17] M. Isaakidis, H. Halpin, G. Danezis, UnlimitID: Privacy-preserving federated identity management using algebraic MACs, In Proceedings of the 2016 ACM on Workshop on Privacy in the Electronic Society, 2016, pp. 139-142.
[18] M. Kroschewski, A. Lehmann, C. Özbay, OPPID: Single Sign-On with Oblivious Pairwise Pseudonyms, Cryptology ePrint Archive, 2024.
[19] E. Dauterman, D. Lin, H. Corrigan-Gibbs, D. Mazières, Accountable authentication with privacy protection: The Larch system for universal login, In 17th USENIX Symposium on Operating Systems Design and Implementation (OSDI 23), 2023, pp. 81-98.
[20] S. Hammann, R. Sasse, D. Basin, Privacy-preserving openid connect, In Proceedings of the 15th ACM Asia conference on computer and communications security, 2020, pp. 277-289.
[21] M. Takemiya, B. Vanieiev, Sora identity: Secure, digital identity on the blockchain, In IEEE 42nd annual computer software and applications conference (compsac), 2018, vol. 2, pp. 582-587, IEEE.
[22] A. Othman, J. Callahan, The Horcrux Protocol: A Distributed Mobile Biometric Self-sovereign Identity Protocol, Selfie Biometrics: Advances and Challenges, 2019, pp. 355-377.
[23] P.J. Lu, L.Y. Yeh, J.L. Huang, An privacy-preserving cross-organizational authentication/authorization/accounting system using blockchain technology, In 2018 IEEE International Conference on Communications (ICC), 2018, pp. 1-6, IEEE.
[24] P. Pappachan, R. Yus, S. Mehrotra, J. C. Freytag, Sieve: A middleware approach to scalable access control for database management systems, arXiv preprint, 2020, arXiv:2004.07498.
[25] K.D. Albab, I. Sharma, J. Adam, B. Kilimnik, A. Jeyaraj, R. Paul, M. Schwarzkopf, K9db: Privacy-Compliant Storage For Web Applications By Construction, In 17th USENIX Symposium on Operating Systems Design and Implementation (OSDI 23), 2023, pp. 99-116.
[26] I. Makhdoom, M. Abolhasan, J. Lipman, M. Piccardi, D. Franklin, PrivySeC: A secure and privacy-compliant distributed framework for personal data sharing in IoT ecosystems, Blockchain: Research and Applications, 2024.
[27] A. Deshpande, Sypse: Privacy-first Data Management through Pseudonymization and Partitioning, In CIDR, 2021.

[28] L. Tsai, H. Gross, E. Kohler, F. Kaashoek, M. Schwarzkopf, Edna: Disguising and Revealing User Data in Web Applications, In Proceedings of the 29th Symposium on Operating Systems Principles, 2023, pp. 434-450.
[29] L. Tsai, Flexible Privacy via Disguising and Revealing, PhD thesis, Brown University, 2024.
[30] M. Beiter, M. C. Mont, L. Chen, S. Pearson, End-to-end policy based encryption techniques for multi-party data management, Computer Standards & Interfaces, 2014, pp. 689-703.
[31] HTTP Over TLS, 2025, URL: https://datatracker.ietf.org/doc/html/rfc2818
[32] Cryptographic Message Syntax (CMS), 2025, URL: https://datatracker.ietf.org/doc/html/rfc5652
[33] R.O. Nambiar, M. Poess, The Making of TPC-DS, In VLDB, 2006, pp. 1049-1058.
[34] M. Barata, J. Bernardino, P. Furtado, An overview of decision support benchmarks: TPC-DS, TPC-H and SSB, New Contributions in Information Systems and Technologies: Volume 1, 2015, pp. 619-628.
[35] B. Buhrow, K. Fritz, B. Gilbert, E. Daniel, A highly parallel AES-GCM core for authenticated encryption of 400 Gb/s network protocols, In 2015 International Conference on ReConFigurable Computing and FPGAs (ReConFig), 2015, pp. 1-7. IEEE.
[36] C. L. Corniaux, H. Ghodosi, An entropy-based demonstration of the security of Shamir's secret sharing scheme, In 2014 International Conference on Information Science, Electronics and Electrical Engineering, 2014, pp. 46-48. IEEE.
[37] Transaction Processing Performance Council (TPC), TPC BENCHMARK ™ DS Standard Specification Version 3.2.0., 2025, URL: https://www.tpc.org/TPC_Documents_Current_Versions/pdf/TPC-DS_v3.2.0.pdf