INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2020, VOL. 66, NO. 3, PP. 449-458
Manuscript received February 29, 2020; revised July, 2020. DOI: 10.24425/ijet.2020.131898

Towards an Auditable Cryptographic Access Control to High-value Sensitive Data

Krzysztof Kanciak, and Konrad Wrona

***Abstract*—We discuss the challenge of achieving an auditable key management for cryptographic access control to high-value sensitive data. In such settings it is important to be able to audit the key management process - and in particular to be able to provide verifiable proofs of key generation. The auditable key management has several possible use cases in both civilian and military world. In particular, the new regulations for protection of sensitive personal data, such as GDPR, introduce strict requirements for handling of personal data and apply a very restrictive definition of what can be considered a personal data. Cryptographic access control for personal data has a potential to become extremely important for preserving industrial ability to innovate, while protecting subject's privacy, especially in the context of widely deployed modern monitoring, tracking and profiling capabilities, that are used by both governmental institutions and high-tech companies. However, in general, an encrypted data is still considered as personal under GDPR and therefore cannot be, e.g., stored or processed in a public cloud or distributed ledger. In our work we propose an identity-based cryptographic framework that ensures confidentiality, availability, integrity of data while potentially remaining compliant with the GDPR framework.**

***Keywords*—Identity-based encryption, GDPR**

I. Introduction

IT is often claimed that the 21st century is an age of data. The data is one of the most important assets, enabling information superiority for military operations, protection of citizens by the governments and continuous improvement in health care and other public services. Although data mining and artificial intelligence have been around for several decades, attracting fluctuating research and commercial interest, it is in the last decade that the data, and particular the personal data, became one of the most sought for assets - both in context of commercial activities and the intelligence gathering.

Despite of all this potential for contributing to greater good, the data can be also used for malicious purposes - and indeed the immediate impact of this malicious use on individuals and society can be even greater than long-term opportunities promised by so-called *data science*. Identity theft, manipulation of public opinion, and activity of companies such as Cambridge Analytica are all good examples how data can be misused. Moreover, the high profile leaks by insiders, such as Bradley Manning and Edward Snowden, and through national-sponsored hackers, such as the Sandworm group [1], underline the need for strengthening defence-in-depth measures and ensuring confidentiality and integrity of data at rest, in transit and in use. We need to develop a comprehensive approach for protection of high-value sensitive data that meets both stringent requirements related to defence against sophisticated and powerful adversaries as well as stringent legal requirements introduced by the recent privacy regulations. In the rest of this paper we primarily focus on a civilian use case related to protection of personal data in context of the European GDPR regulation, however our discussion is also applicable to governmental and military applications.

A. Protection of personal data

During the last decade we could observe an increasing awareness among regulators and non-governmental organizations about importance of protection of privacy - and inadequacy of the existing norms and laws in this context. This has resulted in development of several legally-binding, as well as voluntary, frameworks focusing on protection of personal information - the most prominent of them being the European Union General Data Protection Regulation [2]. The GDPR by design gives individuals better control over their personal data (any information relating to an identified or identifiable natural person) and establishes a single data protection regulation across the EU, covering an easier access to personal data, a right to rectification, a clearer right to erasure (right to be forgotten), a new right to data portability, a right to consent, and a right to be informed about a breach of one's personal data. The GDPR generally follows a binary approach to personal data – the data is either personal or not. If data is considered to be personal data, the full weight of the GDPR's regulatory regime applies to any entity processing that information. In this context, an encrypted personal data falls in a kind of a grayscale - although ciphertexts can be mostly considered as indistinguishable in context of the concerned subjects, they can still be matched to the concrete subject if an appropriate key is available.

There is a fundamental contradiction between development of new business models, that often utilize a wide spectrum of data sources, and the ability to enforce an effective protection of personal data. The increased operational burden and potential liability, introduced by legal frameworks such as

This work was supported by the project SEMACITI financed by the Polish Ministry of Defence as a part of Kosciuszko Programme.

K. Kanciak is with Military University of Technology, Warsaw, Poland.
K. Wrona is with NATO Cyber Security Centre, The Hague, The Netherlands and with Military University of Technology, Warsaw, Poland.

This work was presented at the International Scientific Conference Mathematical Cryptology & Cybersecurity (MC&C 2020), Warsaw, 16-17.01.2020.

the GDPR, can discourage and hamper innovation in services and new business model. With the GDPR, even organizations without a physical market presence in the EU may still be required to comply with the GDPR if the organization offers paid or unpaid goods or services to individuals located in the EU or if the organization is collecting personal data about individuals within the EU. In addition, if an organization works with suppliers or partners that operate in the EU, they will expect the organization to comply with the GDPR in order to limit their own risk. The GDPR compliance is considered a requirement to conduct business with EU data subjects.

It is worth mentioning that year by year (or data theft by data theft) individuals are becoming increasingly aware of the threats of data breaches, as well as of commercial value of their personal data. The GDPR provides a framework for protecting personal data across the European Union and aims to give back the control over personal data to individuals. As a consequence, an adequate balance between personal data protection and data-based business models has to be achieved. We think that modern cryptography offers tools that make it possible.

B. Cryptographic solutions to personal data protection

We believe that cryptography provides a key solution to protect privacy of individuals while preserving opportunities for innovation in IT services. In our work we discuss a set of cryptographic techniques that ensure confidentiality, availability, integrity of data and at the same time are in our opinion compliant with data protection framework. Moreover, thanks to use of cryptography, consumers could in a potentially much more fine-grained and dynamic way decide with whom, when and what data they share.

Personal data encryption has a potential to become extremely important for preserving innovation in business models in a data-driven economy, while protecting subjects' privacy, especially in the context of widely deployed modern monitoring, tracking and profiling capabilities, used both by governmental institutions and by high-tech companies. We support an opinion that *for achieving balance between data protection and not restricting the growth of the digital economy would be for the European Data Protection Board (EDPB) to maintain an up-to-date list of proven encryption technologies. Any personal data that was encrypted with those technologies vetted by the EDPB would be considered "not personal" for any parties who did not have the encryption key.* [3]. In such scenario, any personal data that is encrypted using cryptographic solutions vetted by the EDPB would be considered not personal. In this article we propose suitable candidates for such cryptographic solutions and discuss some of challenges related to their use in practice.

In order to make our discussion more concrete, let us discuss some examples of business use cases where individuals willingly provide their consent to disclose some of their personal data. In 1998 auto insurance company Progressive launched in the United States so called Snapshot program [1], the first large scale telematic-based tracking program. Progressive used the program to incentivize good driving habits by offering discounts to *safe drivers*. However, in 2013, the company started using the collected data also to penalize the *bad drivers*. Quasi overnight, the application of the collected data has changed - from awarding to penalizing. And although from a technical perspective one could argue that this is alike, clearly from the perspective of the users, it has a completely different impact.

[1] https://www.progressive.com/auto/discounts/snapshot/

C. Privacy-preserving storage

Privacy-preserving storage of personal data in a public repository, such as a public cloud or a distributed ledger, enables also new channels of communication between organizations, including financial institutions, and their users or clients. Various EU regulatory provisions require that a company must provide certain information to a client in writing, either on paper or using another *durable medium*. Such record-keeping requirements are included, for example, in the revised Markets in Financial Instruments Directive (MiFID II). Therefore, durable medium has become one of the biggest regulatory challenges for the financial institutions and banks are analyzing various durable medium options and looking into innovative alternatives. One of frequently discussed technologies include distributed ledger technology. However, most of such blockchain-based notary-like solutions are not compliant with the restrictions placed on personal data processing by the GDPR.

D. Fine-grained release of information

Many of the current approaches to protecting access to personal data are largely binary - when a legal agreement with a service provider is acknowledged by the user, the provider gains access to all future (and sometimes even some past) personal data of interest. This approach is questionable, as the user has to rely on and trust the provider with respect to technical measures taken to protect his data, as well as to adhering to an intended use of the data. Of course, there is a legal framework in place to protect user, but this does not address the fact that once privacy regarding some data is lost, e.g. due to a successful cyber attack or commercial misuses, it cannot be regained.

Cryptographic access control has a potential to offer much more fine-grained control over release of personal data. In particular, using approaches such as attributed-based access control, relatively complex access control policies can be integrated into the key management process. This, combined with individual encryption of specific data elements, enable implementation of fine-grained approach to control and release of personal data.

Our proposed framework uses a combination of mechanisms in order to ensure a holistic protection of data. Identity-based encryption (IBE) is used to preserve confidentiality, whereas integrity and accountability are protected by a blockchain. Availability is ensured by a decentralized high-performance storage system.

As our goal is to make our work understandable to a broad spectrum of readers with varying background, we provide in

the following sections a concise introduction to all cryptographic, legal and operational concepts that are required to understand our framework.

E. Related work

We observe growing number of articles about managing sensitive data in distributed environments and we share emerging opinion that simple encryption mechanisms do not cover all desired by GDPR regulation requirements. This subject is gaining a lot of attention recently, but we are not aware of any ready-made approach or system in the world of cryptography that solves all issues formulated above without any doubts. In our work we cover various areas, such as the GDPR compliance [4], [5], data protection, cryptography and distributed ledger technology. However, from a functional perspective the most related work is in the area of blockchain-based encryption schemes and privacy-preserving solutions [6]–[12] such as Identity-Based Puncturable Encryption, decentralized, fine-grained key management [13]–[15], decryption delegation rights management [16]–[18], and threshold schemes [19]–[21]. In our work we focus on permissioned blockchain solutions like Hyperledger blockchain technologies ecosystem [22]. However, our solution could be also extended to achieve larger decentralization by using a public ledger - our inspiration here is from the work performed within the NuCypher project [2] and Umbral threshold proxy re-encryption scheme [23], [24]. NuCypher allows data owner delegating decryption rights to recipient through a re-encryption process performed by a set of semi-trusted proxies. When threshold value of proxies perform re-encryption, recipient is able to combine these independent re-encryptions and decrypt the original message using his private key.

II. Cryptography and legal data protection requirements

An obvious and potentially effective approach to protecting personal data is encryption. Data-intensive systems, such as financial systems or IoT solutions, may collect various data that can be considered personal, at the same time introducing a business need to store and process collected *big* data in a convenient and cost-effective way. But the question if an encrypted data should be treated as personal or not gets convoluted due to the GDPR's inclusion of two other concepts: anonymization and pseudonymization.

The GDPR considers *anonymized* data as data, which has been irreversibly stripped of any way of identifying the underlying individual, even by the organization that did the anonymizing. From crypto perspective it is more like hashing than encryption. Encryption assumes ciphertext-plaintext mapping existence. Thus encryption is definitely not locking the data up and throwing away the key.

Pseudonymization, on the other hand, involves *the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information*. While it does provide additional safeguards, pseudonymized data, unlike anonymized data, is still unequivocally considered personal data under the GDPR, as noted in Recital 26. Under that same recital, anonymized data is no longer personal data because it has been transformed to unidentifiable. Article 4 of the GDPR states that personal data is 'any information relating to an identified or identifiable natural person ('data subject').' Thus, as long as there is a key to the encrypted data out there somewhere, the personal data can be traced back to the individual and the GDPR will apply.

Pseudonymized data, however, should be considered identifiable if it could be attributed to an individual. Pseudonymization is reversible and entity, which pseudonymized the data, possesses secret mapping to reverse the process. This is, from cryptographic point of view, more or less definition of encryption. In anonymization, no one has a secret key or the key is gone. In pseudonymization, the same party who pseudonymizes the data usually does have a key. However, with encryption, many of the parties who are processing the data, such as cloud storage providers, do not have the encryption key to unscramble that data. The encryption key stays with the generator or the end user of the data. The fundamental difference is: Who holds the metaphorical key?

A. Role of encryption in enforcement of the GDPR

Encryption is mentioned in the GDPR as one of the measures that can be taken to maintain security and to prevent processing in infringement of the regulation, but it is not as strong category as pseudonymization or anonymization. The GDPR states that companies need to take into account the state-of-the-art, risks and severity for the rights and freedoms of the data subjects when implementing encryption. This encryption should ideally always cover the entire communication in an end-to-end manner and it is nowadays becoming common practice.

The simple conclusion is that encrypted data is considered as personal and cannot be processed, e.g., on a public blockchain. This differs from an approach commonly taken by cryptographers and in secure protocol design, where an objective is to ensure that an encrypted data is, generally speaking, indistinguishable from a random string of bits and thus may be transported over public channels. Nevertheless, because of mostly reversible character of encryption, it is often argued that encryption is from a legal perspective more conceptually similar to pseudonymization than anonymization and therefore encrypted data would also be considered as a personal data under the GDPR [3]. We think that the question of how encrypted data would be viewed under the GPDR is still an open one. We believe it is important to create discussion forum with participants with legal and cryptographic background in order to explore this important topic more in-depth.

B. Identity-based encryption for personal data protection

We believe that identity-based encryption (IBE) may offer a suitable answer to the GDPR personal data protection requirements. In identity-based approach publicly known string (e.g email address of recipient) is needed to encrypt message, but the instantiation of the decryption key is separated in time until

[2]https://www.nucypher.com/

the moment when recipient makes request to so-called Key Generation Center (KGC). The complete process is described in more details in Section IV. But, the most important fact from the privacy perspective is that the secret key, required for decryption of the data, does not exist until the recipient asks for its generation. We think this very specific setting may solve the problem of possessing keys to pseudonymized data (that was formulated in paragraph above) and allows considering IBE encrypted data as not personal. Proposed approach has several advantages:

- key recovery in IBE setting is fully controlled by KGC and the key does not exist up to successful authorization of secret message recipient;
- key recovery process fulfills policy implemented in organization (including the right to be forgotten);
- organization is able to implement policies that will ensure proper data processing like private key expiration period or forgotten user policy;
- message recipient does not need private key before he receives message (PKI case).

In [25] we can find same considerations: *there is an ongoing debate surrounding whether data typically stored on a distributed ledger, such as public keys and transactional data qualify as personal data for the purposes of the GDPR. Specifically, the question is whether personal data that has been encrypted or hashed still qualifies as personal data*. This document points that the holder of the encryption key can still re-identify each data subject through decryption given that the personal data is still present in the dataset that has been encrypted. As a consequence, encrypted data remains personal data — at least for the holder of the key able to identify such data. The Article 29 Working Party [3] indeed clarified in its opinion on cloud computing that although encryption 'may significantly contribute to the confidentiality of personal data if implemented correctly' it does not 'render personal data irreversibly anonymous' (Directive 95/46/EC. Recital 26). In [26] it was suggested that 'sufficiently well-encrypted data, where the provider has no access to the key, should not be 'personal data', and similarly with sufficiently anonymised data'. This implies that there is a fundamental distinction from a legal perspective between scenarios, where parties may have access to the decryption key and the scenarios where it could be ensured that such access does not exist.

IBE allows decryption key creation to be postponed to the decryption moment. Until the key is created the risk of linking the data to the owner is reduced - and would require successful attack on a used cryptographic mechanism or its implementation. In our work we consider blockchain-based IBE decryption key creation that meets the requirement of auditability.

III. Auditable key and data management

As discussed earlier, a critical assumption in our approach is non-existence of the decryption key before generation of such key is requested by an authorized user. However, the meaning of *non-existence* requires some more discussion. In particular, in a naive implementation of an IBE, the key generation authority has to be treated as a trusted-third party, as it can generate, without user's knowledge, private keys related to his identity. A same naive implementation could include an identifier of a recipient of an encrypted document in a metadata attached to the encrypted object. In such a scenario, it would be easy for a compromised - or misbehaving - key generation authority to generate the decryption and to decrypt any chosen object encrypted under the scheme. Moreover, such a scenario does not provide a strong and auditable privacy protection. In this case, although the decryption key might not exist in the theory at the moment when data is encrypted, in practice it can be easily generated on-demand by an authority without a control or knowledge of a user - this can result effectively in a *Big Brother* scenario that, although most probably welcomed by many governments and organizations, is unacceptable from a citizen's perspective.

In order to prevent such a systematic failure, the implemented IBE scheme needs to meet specific requirements that are discussed in the following sections.

A. Anonymization of encrypted objects

The encrypted objects cannot provide any clues that would enable disclosure of the identity required for decryption. The idea of fully anonymous IBE ciphertexts was explored, e.g., in [27] and [28]. The *committed blind anonymous IBE* scheme has been proposed in [29] and has an advantage of fulfilling both our requirement for anonymity of the ciphertext and escrow protection, as discussed in the next section.

B. Preventing unilateral generation of decryption keys

In order to ensure an auditable cryptographic access control, we would like to prevent any entity from unilateral generation of the decryption keys. This simple to formulate requirement is rather difficult to achieve in practice. For example, in IBE the ability to generate the decryption key is removed from the end user, and assigned to the key generation center (KGC) instead. Unfortunately, in simple IBE schemes KGC needs to be a trusted party, as it is able to generate any decryption used in the system.There are two main approaches that can be implemented in order to reduce and control the inherent key escrow characteristics of the IBE schemes. They both rely on splitting of some aspects of the key generation authority between multiple entities:

1) Separation of the identity authentication and key generation actions between two or more parties [30],
2) Dividing of the secret key required for generation of the private decryption key between multiple entities [31], [32].

In both cases the ability of key escrow and uncontrolled decryption and access to private data is significantly reduced, as performing such illegitimate activity would require collusion of several independent parties.

Separation of authentication and key issuing tasks between two parties - an identity certifying authority and a key generating authority has been proposed in [30]. In this approach, identity certifying identity issues a certificate, that can be verified

[3] https://ec.europa.eu/newsroom/article29/news.cfm

by the key generation authority. In order to achieve increased privacy, it is assumed that the identity certifying authority does neither log identity of the user requesting certification nor the issued certificate - or any other information that could be used to connect these two.

Another method of avoiding key escrow relies on splitting the master secret key between multiple parties, by relaying on threshold schemes. This approach was first suggested in [31], where implementation using verifiable secret sharing [33] was proposed. The concept was further extended in [32].

Recently, an approach called registration-based encryption (RBE) [34] has been proposed in order to solve the key escrow problem in the IBE. The solution relies on replacing private key generator with a public *key accumulator*. This work has been further extended in [35], where first efficient implementation of RBE has been proposed, and in [36], where a verifiable RBE, providing succinct proofs of registration and non-registration, has been proposed.

Trusted third party don't have to be given the ability to decrypt ciphertexts intended for users in so called certificate-less encryption schemes [37], [38]. It's a form of asymmetric encryption that eliminates the disadvantages of traditional PKI based public-key encryption scheme, i.e. doesn't require digital certificates or a public-key infrastructure, and identity-based encryption i.e. it remains secure against any third party attacks including KGC. These schemes use two pairs of keys. One of them is a classic private-public key pair but the public key is not signed by any kind of Certificate Authority. Second pair of keys is ID-based and consists identity as a public component and and the associated identity-based private key supplied by a key generation centre beforehand. To encrypt the message use of both public components is needed (identity and classic public key. Decryption requires both secret values - classic private key generated by the party of communication and the id-based private key supplied by the key generation centre. This setting implies the key generation centre cannot break the confidentiality of a transmitted message as it does not know the classic private key of any party. The problem is that classic public keys still have to be somehow publicised and identity cannot be easily changed (KGC is still needed to generate private keyassociated with identity) which means granularity and expressiveness of this approach is very limited.

C. Unique decryption key for each object

In order to enable a fine-grained cryptographic access control and a selective release of sensitive data, it is important to ensure that each data object is encrypted with a separate key.

D. Durable log of key generation requests

It is important to verify, for which data objects a decryption key has been generated - and thus which objects require different handling and protection. Every operation (addition, removal, key instantiation) is logged on blockchain. Decrypted objects in our PoC are deleted or re-encrypted with 'new identity' which translates into a changed period of validity (explained in Section IV-B).

IV. Identity-based encryption

The IBE is not a new idea - it has almost 20 years of research history and it is an established encryption method. For example, the Boneh-Franklin IBE scheme, which uses techniques from pairing-based cryptography, is described in RFC 5091, published in 2007 []. The concept of identity-based cryptography has been introduced in [39]. Its primary innovation was the use of user identity attributes, such as email addresses or phone numbers, instead of digital certificates, for encryption and signature verification. The original motivation for IBE was to simplify certificate management in email systems. This feature significantly reduces the complexity of a cryptography system by eliminating the need for generating and managing users' certificates. It also makes it much easier to encrypt messages to non-enrolled users, since messages may be encrypted for users before they interact with any system components.

Identity-based cryptosystems do not instantiate secret keys before first encrypted message. Main advantage of IBE is that it allows a user to encrypt a message without a need for pre-distribution of encryption keys. An IBE is a public key scheme, where the public key is derived directly from the user identity, while a trusted third party, called the Key Generation Center (KGC), generates the corresponding private key. It means that there is a way for a sender to encrypt secret message to a receiver without establishing symmetric secret key or having public key of receiver. The sender needs only identity of receiver to encrypt his message. Decryption key is being instantiated at a distant time when receiver is authenticated by the KGC. This means in turn that before the user makes request to KGC for his decryption key, the key does not exist - and therefore one can argue that encrypted data is not pseudonymized but anonymized.

A fully functional IBE scheme has been proposed in [31]. The scheme is characterized by a chosen-ciphertext security in the random oracle model, assuming a variant of the computational Diffie-Hellman problem [40].

When Alice sends an email to Bob, she can simply encrypt the message using the public key string. There is no need for Alice to obtain Bob's public key certificate. When Bob receives the encrypted email he contacts a third party, the PKG. Bob authenticates himself to the PKG and obtains his private key from the PKG. Bob can then read his email.

A. Mathematical background

There are two main methods for implementing IBE schemes. Bilinear parings on ellipitc curves groups and lattices. Bilinear pairings based schemes have much shorter keys and ciphertexts and these are important in context of distributed computing and communication with public or permissioned distributed ledgers. But these schemes are susceptible to quantum attacks due to Shor's algorithm. On the other hand lattice based IBEs such like Gentry-Peikert-Vaikuntanathan [41] and Agraval-Boneh-Boyen [42] and their extensions have quantum algorithm resistance i.e. there are no quantum algorithms for solving hard computational problems for lattices.

Below we describe an IBE scheme using pairings on elliptic curves. Pairings map pairs of points on an elliptic curve into the multiplicative group of a finite field. First identity-based encryption scheme was published in 2001 by Boneh and Franklin's [31]. A comprehensive report on pairing-based cryptography has been published by NIST in 2015 [43].

A point on the elliptic curve E is an ordered pair (x, y), where $x, y \in F_q$ is satisfying equation $y^2 = x^3 + ax + b$ (so called the short Weierstrass equation) where a and b are chosen so that $4a^3 + 27b^2 \neq 0$. Points on elliptic curve form a finite Abelian group structure. It means there is a way to "add" two points on an elliptic curve, and always get another point. To get possibility of "points multiplication" we can add them repeatedly and the discrete logarithm problem in such groups is believed to be hard, making it ideal for cryptography.

A pairing is a function e that takes a pair of two points on an elliptic curve and outputs an element in a finite field. The pairings we consider are bilinear which means that the pairing map preserves the additive structure of the elliptic curve, and carries it over into the finite field. We can say a pairing (or map) $e : G_1 \times G_1 \to G_T$ is bilinear if:

- G_1 and G_T are groups of the same order q
- For all elements $a, b \in Z_q^*, g \in G_1 \quad e(g^a, g^b) = e(g, g)^{ab}$
- The map is non-degenerate (i.e., if $G_1 = \langle g \rangle$, then $G_T = \langle e(g, g) \rangle$)
- e is efficiently computable.

Now let (G_1, G_T) be a pair of bilinear groups where g is random generator of G_1 and there is efficintly computable pairing $e : G_1 \times G_1 \to G_T$. The Decisional Bilinear Diffie Hellman problem is to decide, given a tuple of values (g, g^a, g^b, g^c, T),where a, b, c are selected randomly from Z_q^*, whether $T = e(g, g)^{abc}$ or if T is random element of G_T. So that the difference of probabilities described in figure below is negligible: $|Pr(Exp(0) = 1) - Pr(Exp(1) = 1)| \leq \epsilon$

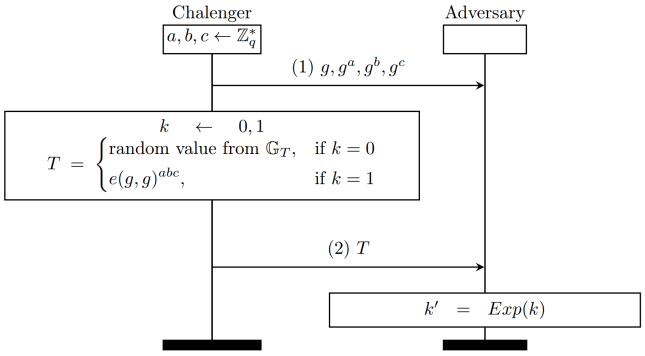

Fig. 1. Decisional Bilinear Diffie Hellman

Although in theory pairings exist for any elliptic curve, in practice there are curves whose pairings are not suitable for cryptographic applications. There is a specific parameter of each elliptic curve that can be calculated known as the embedding degree k. In order to efficiently implement pairings for use in cryptography, we need k to be small (less than 100). However, k is usually about the same size as q, which is at least 160 bits — way too big. But there are two common ways to find pairing-friendly elliptic curves. The first is to use what are known as supersingular elliptic curves, which always have $k \leq 6$. The second way is to use a technique called the complex multiplication (CM) method to construct certain families of elliptic curves with small k. In brief, the security of identity-based encryption is based on the property that the particular bilinear maps chosen are one-way functions, since the Bilinear Diffie-Hellman problem is reducible (algorithmically equivalent) to the discrete-log or inverse operation for these bilinear maps [44].

In order to actually implement any pairing-based cryptographic protocol, it is necessary to choose a specific pairing function e. The two most commonly used pairings are the Weil and Tate pairings. With the goal of speeding up computation, researchers have discovered several new pairings. These include the Ate, Eta, reduced Tate, twisted Ate, and R-Ate pairings among others. For further details go to [43] survey paper that reviews all details. It is ok to treat pairings as a "black box" and build various cryptographic schemes making use of assumed properties of the pairings.

B. Definition of identity

As pointed in [31] identity tag may have an arbitrary form. This means, that it can be used to encode an auxiliary information. For example, a traditional PKI certificate includes a preset expiration date. In an IBE system key expiration can be implemented by including the minimum validity date in the identity. An illustration of this approach is having message encrypted by sender using the example identity: "jonh@company.com —— 2020", so that John - the receiver can use his private key during 2020 only and next year John needs to obtain a new private key from the KGC. Hence, we get the effect of annual private key expiration. But, unlike in the case of traditional PKI systems, sender of a message does not need to obtain a new certificate from John every time John refreshes his private key - he can just include the new validity date in the identity. Granularity and expressiveness of this approach to formulation of the IBE public keys is practically unlimited.

However, this approach relies on an assumption that the KGC is able to correctly interpret the syntax and semantics of a public key and would issue a private decryption key only if all conditions included in the public key are met.

As a practical example, in our proof-of-concept RFC 5091 IBE implementation, the identity is a JSON document with predefined parameters.

C. Advantages and disadvantages of the key escrow

The inherent key escrow feature, provided by basic IBE schemes, opens some specific implementation possibilities. For example, the decryption process can take place on a third party server – and while this looks like a huge disadvantage in crypto world, it could significantly improve usability and reliability for commercial applications by not requiring any

client-side installation and client-side keys storage and management. This is important in some scenarios, e.g., for bank customers where private key loss could result in their personal bankruptcy - although these consequences are obvious fact in cryptocurrencies world, it is difficult to imagine that a regular bank customer would accept a similar risk. Moreover, financial, public government institutions require policies specified, procedures, acts of law and their interpretations, sensitivity levels, targets of evaluation, information processing standards, risk management etc. Thus server-side decryption, although not explicitly compliant with the GDPR, can support many practical and regulatory requirements.

D. Basic IBE flow

A basic IBE flow involves the following parties: the sender S, the recipient R and key generation center KGC that can extract private keys from public keys. In our proof-of-concept there are three more ancillary parties: distributed storage (reponsible for availability), distributed ledger (integrity, auditability) and authorization provider which is pluggable component responsible for handling authentication, ensuring that users are actually identified in a required way (like phone authorization number, email verification link, but also national ID with ability of signing documents or future blockchain based identity providers). Prior to generation of the private key, it is also critical to perform an appropriate authentication of the requester, however the details of such authentication process are outside of the IBE schemes.

A standard IBE scheme consists of four algorithms:

- **Setup**: Prior to the actual communication, a one-time setup must be performed in order to initialize the system. This process generates the public parameters that must be distributed among all parties, and the master secret which should only be deployed to the Key Generation Center. The master secret allows the KGC to extract private keys. Technically `setup` takes a security parameter n and returns chosen elliptic curve, chosen hash function with consistent with n output size and a product of secret s and a random point P on the curve. The system parameters include a description of a finite message space M, and a description of a finite ciphertext space C
- **Encrypt**: The sender uses its identifier to encrypt the plaintext message. Afterwards, the ciphertext is transmitted to the recipient. `Encrypt` phase takes as input parameters, ID, and $m \in M$. It returns a ciphertext $c \in C$
- **Extract**: When the recipient receives the ciphertext, it queries the KGC for its private key with their public key. Given the identity of the recipient is confirmed, the private key is extracted and sent back. This phase takes as input parameters, master-key, and an arbitrary $ID \in \{0,1\}^*$, and returns a private key d. Here ID is an arbitrary string that will be used as a public key and d is the corresponding private decryption key. The Extract algorithm extracts a private key from the given public key.
- **Decrypt**: Possessing the private key pair of the public key, the recipient attempts to decrypt the received ciphertext. If the private key matches the public key that was used during encryption, the process succeeds, resulting in the plaintext. `Decrypt` phase takes as input parameters, $c \in C$, and a private key d. Then returns $m \in M$.

The greatest advantage of IBE lies in the fact that neither the sender, nor the recipient needs to obtain each other's public keys in advance. When performing the encryption, they simply use publicly-known information.

E. Proxy re-encryption (PRE)

Identity-based proxy re-encryption allows a proxy to transform ciphertext from one public key to another, without learning anything about the underlying plaintext. A data producer encrypts a message m, with Alice's public key pk_A, resulting in ciphertext $c_A = encrypt(pk_A, m)$ — it's so called first-level ciphertext. Alice decides to delegate access to message m to Bob, who has the key pair (pk_B, sk_B). To do so, Alice creates a re-encryption key: $rk_{A\to B} = rekey(sk_A, pk_B)$. In re-encryption phase second-level ciphertexts is created $c_B = reencrypt(c_A, rk_{A\to B})$. In identity-based approach instantiation of the decryption key is separated in time until the moment when recipient makes request to KGC and from the GDPR perspective the secret key does not exist until the recipient asks for its generation. In identity-based proxy re-encryption scheme with fully controlled KGC we can go step further and ensure decryption keys instantiation only for re-encrypted ciphertext. It means that first-level ciphertexts (ciphertexts before re-encryption) will never be decrypted i.e can no longer be attributed to a specific data subject. So as IBE may offer decryption key postponement, PRE in very specific setting postpones ciphertext creation. From GDPR point of view first-level ciphertext

V. USE CASES

Below, we discuss two possible use cases for use of IBE-based auditable cryptographic access control for data management. They are focused on providing a legal compliance with privacy and banking regulations, while supporting business innovation and user's ability to control the data.

A. Durable medium

We have chosen to develop proof of concept that fulfills idea of durable medium. It is an important mechanism that banks are required to use in order to keep their customers informed about products, agreements, payment schedules, etc. In practice, the implementation of a durable medium is not straight forward. The EU Court of Justice has challenged the practice of sending messages to customers using banking mailboxes because such approach does not guarantee the immutability of the distributed documents. Thus, a durable medium has become one of the biggest regulatory challenges faced by the financial institutions. In particular, a durable medium needs to:

- allow information to be addressed personally to the recipient

- enable storage of information that is accessible for future reference and for a period of time adequate for the purposes of the information (storability)
- allow the unchanged reproduction of the information stored (reproduction)

Our threat model assumes that the adversary is computationally bounded, secure cryptographic hash functions exist, and there is a bilinear pairing in which the decisional bilinear Diffie-Hellman assumption holds. We need also blockchain BFT consensus assumptions. We assume that message recipients do not trust senders.

1) Blockchain-based document life-cycle management: Our proposed storage system is implemented in public cloud (we use major commercial service providers, i.e., Azure and AWS) and is based on Gaia[4]. Access control in a storage system is performed on a per-address basis. Every document committed to the storage system has its metadata blockchain transaction. Metadata contains hash of the document, its language, validity time, pointer to a transaction committing an earlier version of the document, publisher's and signer's identity, domain name, operation type (creation, update, erasure, reading) — all parameters to provide full document life-cycle management and none of them de-anonymizes the document recipient.

2) Key generation component: We investigated three variants of possible solution. First one is a centralized KGC (with a trusted third party) and is the closest one to the solutions currently implemented in the payment system infrastructure. Second variant adds a very simple threshold scheme to improve customers privacy and weaken identity management system reliance. Third one is the most complex one and implements federated multi-authority identity management system.

a) Centralized KGC: Every encrypted document can be downloaded by recipient. The private key for this particular document is instantiated at the time of authentication by KGC. KGC does not know the location of an encrypted file as it depends on blockchain transaction hash value so that it cannot bind the ciphertext to identity. The scheme is simple, but introduces challenges in terms of reliability and fairness due to single points of failure or compromise respectively.

b) Threshold scheme: Our threshold scheme is a hybrid approach using identity-based and symmetric encryption, inspired by [45]. Document is encrypted using a symmetric key that is divided between separate security domains (i.e. KGCs) that act as two separate authentication components. The recipient has to authenticate to KGCs, possibly using different methods of authentication, in order to collect parts of symmetric key required to decrypt the document. In this approach KGC fraud impersonation (with file location knowledge) becomes infeasible. But it has to be noted that intermediate symmetric encryption breaks assumption of key non-existence, which was critical in our approach in order to provide the GDPR compliance. It means that right after secret key is divided and encrypted with identity of the recipient, it has to be implicitly forgotten.

c) Federated multi-authority identity managers: Last tested approach is inspired by [46]. It is most secure, but the least applicable due to its complexity and low suitability to the business environment of the discussed case. This approach is inspired by [46] and assumes identity propagation between different but federated security domains which exchange messages containing users' authentication credentials. All federated parties need to have common interest which is not obvious in context of own customers data sharing. In this scheme, multiple established authorities can instantiate private keys associated with the identities under their control (i.e. in their security domains) independently. This scheme has additional authority setup phase that takes as inputs an identifier of an authentication provider (federated authority) and the global parameters and outputs a public key and a master secret key for a given authority. In the federated identity management schema, authorization providers from different security domains exchange messages containing authentication information and credentials characterizing users [47]. This approach can leverage existing conventional centralized authorities such as corporate directory services, certificate authorities, or domain name registries. Each separated security domain serve as its own root of trust. Multi-authority identity management is an additional layer of identity management and is implementable in case of common interest or business case among federation members.

[4]https://github.com/blockstack/gaia

B. Electronic receipt

Second considered use case are digital receipts understood as machine-readable, electronic substitutes for their contemporary paper-based printed counterparts. While in Poland retail landscape is still not yet adopting digital receipts, other countries like Sweden or the United Kingdom indicate that the adoption of digital receipts is actually happening.

A practical problem is to provide functionality of digital receipts without over-complicating user's buying experience and without infringing user's data ownership. Currently, two main approaches are used in implemented solutions. First approach requires users to provide additional data, such as email or loyalty card, during payment process, which is in our opinion unnecessary complication of the buying process. Second approach (used by, e.g., Flux in the UK) passes receipt data to banking application, which violates user's right to privacy, as there is no need for a bank or a card provider to know user's shopping list as opposed to only the value of the payment. There is no need to share this valuable, and potentially sensitive, information with a bank.

Our proof of concept uses credit card number used during payment process, which is concatenated with some additional data to provide an identity utilized in an IBE encryption process[5]. Since, under specific circumstances described earlier in our paper, IBE encrypted data can be considered not

[5]In details it means consumers allowance for using PSD2 procedure to get his bank identificators (bank account number, credit cards numbers etc). PSD2 is an EU Directive 2007/64/EC, administered by the European Commission to regulate payment services and payment service providers throughout the European Union and European Economic Area. The Directive's purpose is to increase pan-European competition and participation in the payments industry also from non-banks, and to provide for a level playing field by harmonizing consumer protection and the rights and obligations for payment providers and users.

personal, the digital receipts can be stored until the moment of end consumer authentication in his IBE domain. Consumer releases his personal data for processing at the moment when a private key is instantiated. Until then, no private keys exist, which is auditable thanks to a blockchain-based key management (described in previous paragraph).

In our proposed solution shopping data remains entirely under consumers control – this also enables potential emergence of a new data market for shopping data. Banks are used as a authentication providers only which is possible under PSD2 (Payment Services Directive EU 2015/2366) regulation. Although our solution requires an update to firmware used in payment terminals, it does not change payment process from the customer's point of view.

VI. Conclusions

We have presented results of our initial investigation of use of IBE for implementing auditable and the GDPR-compliant data-management framework that provides decentralized storage and life-cycle management of personal data. Data storage, access, erasure, and auditability are ensured by a public blockchain log.

We have proposed three variants of a decentralized framework for auditable management of personal data while maintaining fairness and confidentiality. We demonstrated the feasibility of using the framework to address the data sharing needs of actual financial or public organizations in the context of a durable medium. To address all identified requirements introduced by the EU regulation framework, we propose a cryptographic framework using an IBE in combination with decentralized reliable storage, blockchain-based auditability and threshold secret sharing. We believe that such a framework can offer a suitable technology to protect the integrity and confidentiality of shared data and to ensure data-access accountability by generating a third-party verifiable audit trail for data accesses. The presented results are in our opinion promising, but initial and require further validation - both in context of technical scalability and performance as well as legal implications. We would like to encourage discussion within the cryptographic R&D community on how the recent advances could be used in order to address some practical issues introduced by increasingly complex regulations and supervision requirements.

References

[1] A. Greenberg, *Sandworm - A new era of cyberwar and the hunt for the Kremlin's hackers*. Doubleday, 2019.

[2] EU, "General Data Protection Regulation 2016/679," 2016.

[3] J. Gresham, "Is encrypted data personal data under the gdpr?" Available online at: https://iapp.org/news/a/is-encrypted-data-personal-data-under-the-gdpr/, 3 2019.

[4] S. Garg, S. Goldwasser, and P. N. Vasudevan, "Formalizing data deletion in the context of the right to be forgotten," Cryptology ePrint Archive, Report 2020/254, 2020, https://eprint.iacr.org/2020/254.

[5] D. Derler, S. Ramacher, D. Slamanig, and C. Striecks, "I want to forget: Fine-grained encryption with full forward secrecy in the distributed setting," Cryptology ePrint Archive, Report 2019/912, 2019, https://eprint.iacr.org/2019/912.

[6] J. B. Bernabe, J. L. Canovas, J. L. Hernández-Ramos, R. T. Moreno, and A. F. Skarmeta, "Privacy-preserving solutions for blockchain: Review and challenges," *IEEE Access*, vol. 7, pp. 164 908–164 940, 2019.

[7] L. Widick, I. Ranasinghe, R. Dantu, and S. Jonnada, "Blockchain based authentication and authorization framework for remote collaboration systems," *2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, pp. 1–7, 2019.

[8] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "Medrec: Using blockchain for medical data access and permission management," 08 2016, pp. 25–30.

[9] A. W. Dent, "A brief introduction to certificateless encryption schemes and their infrastructures," in *Public Key Infrastructures, Services and Applications*, F. Martinelli and B. Preneel, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 1–16.

[10] H. Shafagh, L. Burkhalter, A. Hithnawi, and S. Duquennoy, "Towards blockchain-based auditable storage and sharing of iot data," in *Proceedings of the 2017 on Cloud Computing Security Workshop*, ser. CCSW '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 45–50. [Online]. Available: https://doi.org/10.1145/3140649.3140656

[11] G. Zyskind, O. Nathan, and A. Pentland, "Decentralizing privacy: Using blockchain to protect personal data," *2015 IEEE Security and Privacy Workshops*, pp. 180–184, 2015.

[12] X. A. Wang, F. Xhafa, Z. Zheng, and J. Nie, "Identity based proxy re-encryption scheme (ibpre+) for secure cloud data sharing," in *2016 International Conference on Intelligent Networking and Collaborative Systems (INCoS)*, Sep. 2016, pp. 44–48.

[13] M. Egorov and M. Wilkison, "Nucypher KMS: decentralized key management system," *CoRR*, vol. abs/1707.06140, 2017. [Online]. Available: http://arxiv.org/abs/1707.06140

[14] A. Sonnino, M. Al-Bassam, S. Bano, and G. Danezis, "Coconut: Threshold issuance selective disclosure credentials with applications to distributed ledgers," *CoRR*, vol. abs/1802.07344, 2018. [Online]. Available: http://arxiv.org/abs/1802.07344

[15] V. Reniers, D. V. Landuyt, P. Viviani, B. Lagaisse, R. Lombardi, and W. Joosen, "Analysis of architectural variants for auditable blockchain-based private data sharing," in *SAC '19*, 2019.

[16] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, "Improved proxy re-encryption schemes with applications to secure distributed storage," *IACR Cryptology ePrint Archive*, vol. 2005, p. 28, 2005.

[17] D. Nuñez, I. Agudo, and J. López, "Proxy re-encryption: Analysis of constructions and its application to secure access delegation," *J. Netw. Comput. Appl.*, vol. 87, pp. 193–209, 2017.

[18] D. Nuñez, I. Agudo, and J. Lopez, "Proxy re-encryption: Analysis of constructions and its application to secure access delegation," *Journal of Network and Computer Applications*, vol. 87, 03 2017.

[19] E. Kokoris-Kogias, E. C. Alp, S. D. Siby, N. Gailly, L. Gasser, P. Jovanovic, E. Syta, and B. Ford, "Verifiable management of private data under byzantine failures," Cryptology ePrint Archive, Report 2018/209, 2018, https://eprint.iacr.org/2018/209.

[20] A. N. Amroudi, A. Zaghain, and M. Sajadieh, "A verifiable (k,n,m)-threshold multi-secret sharing scheme based on ntru cryptosystem," *Wireless Personal Communications*, vol. 96, pp. 1393–1405, 2017.

[21] B. Rajabi and Z. Eslami, "A verifiable threshold secret sharing scheme based on lattices," *Information Sciences*, vol. 501, 11 2018.

[22] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. D. Caro, D. Enyeart, C. Ferris, G. Laventman, Y. Manevich, S. Muralidharan, C. Murthy, B. Nguyen, M. Sethi, G. Singh, K. Smith, A. Sorniotti, C. Stathakopoulou, M. Vukolic, S. W. Cocco, and J. Yellick, "Hyperledger fabric: A distributed operating system for permissioned blockchains," *CoRR*, vol. abs/1801.10228, 2018. [Online]. Available: http://arxiv.org/abs/1801.10228

[23] D. Nuñez, "Umbral: a threshold proxy re-encryption scheme," 2018, https://raw.githubusercontent.com/nucypher/umbral-doc/master/umbral-doc.pdf.

[24] M. Egorov, D. Nuñez, and M. Wilkison, "Nucypher : A proxy re-encryption network to empower privacy in decentralized systems," 2018.

[25] European Parliamentary Research Service Scientific Foresight Unit, "Blockchain and the general data protection regulation: Can distributed ledgers be squared with european data protection law?" Available online at: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634445/EPRS_STU(2019)634445_EN.pdf, 7 2019.

[26] M. Finck, "Blockchain and the general data protection regulation. can distributed ledgers be squared with european data protection law?" 2019.

[27] C. Gentry, "Practical Identity-Based Encryption Without Random Oracles," in *EUROCRYPT Adv. Cryptol.*, vol. 4004, 2006, pp. 445–464.

[28] X. Boyen and B. Waters, "Anonymous hierarchical identity-based encryption (Without random oracles)," in *Adv. Cryptol. - CRYPTO*, 2006.

[29] J. Camenisch, M. Kohlweiss, A. Rial, and C. Sheedy, "Blind and anonymous identity-based encryption and authorised private searches on public key encrypted data," in *Int. Work. Public Key Cryptogr.*, S. Jarecki and G. Tsudik, Eds., 2009, pp. 196–214.
[30] S. S. M. Chow, "Removing Escrow from Identity-Based Encryption," in *Int. Work. Public Key Cryptogr.* Springer, 2009, pp. 256–276.
[31] D. Boneh and M. Franklin, "Identity-Based Encryption from the Weil Pairing," *SIAM J. Comput.*, vol. 32, no. 3, pp. 586–615, 2003.
[32] A. Kate and I. Goldberg, "Distributed Private-Key Generators for Identity-based Cryptography," in *Int. Conf. Secur. Cryptogr. Networks*, 2010.
[33] P. Feldman, "A practical scheme for non-interactive verifiable secret sharing," *28th Annu. Symp. Found. Comput. Sci.*, pp. 427–438, 1987. [Online]. Available: http://ieeexplore.ieee.org/document/4568297/
[34] S. Garg, M. Hajiabadi, M. Mahmoody, and A. Rahimi, "Registration-based encryption: Removing private-key generator from IBE," in *Proc. TCC*, 2018, pp. 689–718.
[35] S. Garg, M. Hajiabadi, M. Mahmoody, A. Rahimi, and S. Sekar, "Registration-Based Encryption from Standard Assumptions," in *Public-Key Cryptogr. - PKC*, 2019, pp. 63–93.
[36] R. Goyal and S. Vusirikala, "Verifiable Registration-Based Encryption," IACR, Tech. Rep., 2019. [Online]. Available: https://eprint.iacr.org/2019/1044
[37] S. Chatterjee and P. Sarkar, *Identity-Based Encryption*. Springer, 2011.
[38] A. W. Dent, "A brief introduction to certificateless encryption schemes and their infrastructures," in *Proc. of the European Public Key Infrastructure Workshop (EuroPKI 2009)*, F. Martinelli and B. Preneel, Eds. Springer, 2010, pp. 1–16.
[39] A. Shamir, "Identity-based cryptosystems and signature schemes," in *Workshop on the theory and application of cryptographic techniques*. Springer, 1984, pp. 47–53, dostep online: http://discovery.csc.ncsu.edu/Courses/csc774-S08/reading-assignments/shamir84.pdf.
[40] X. Boyen and L. Martin, "The Boneh-Franklin BF Cryptosystem," IETF, Tech. Rep. RFC 5091, 2007.
[41] C. Gentry, C. Peikert, and V. Vaikuntanathan, "Trapdoors for hard lattices and new cryptographic constructions," Cryptology ePrint Archive, Report 2007/432, 2007, https://eprint.iacr.org/2007/432.
[42] S. Agrawal, D. Boneh, and X. Boyen, "Efficient lattice (h)ibe in the standard model," in *Advances in Cryptology – EUROCRYPT 2010*, H. Gilbert, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 553–572.
[43] D. Moody, R. C. Peralta, R. A. Perlner, A. R. Regenscheid, A. L. Roginsky, and L. Chen, "Report on pairing-based cryptography," NIST, Tech. Rep., 2015.
[44] Y. Yacobi, "A note on the bi-linear diffie-hellman assumption," in *ryptology ePrint Archive, Report 2002/113*, 2002.
[45] E. Kokoris-Kogias, E. C. Alp, S. D. Siby, N. Gailly, L. Gasser, P. Jovanovic, E. Syta, and B. Ford, "Verifiable management of private data under byzantine failures," Cryptology ePrint Archive 2018/209, 2018.
[46] Y. Rouselakis and B. Waters, "Efficient statically-secure large-universe multi-authority attribute-based encryption," ePrint Archive 2015/016, 2015.
[47] D. Chadwick, "Federated identity management," in *Foundations of Security Analysis and Design V SE - 3*, 2009, p. 96–120.