

# Selective Deletion in a Blockchain

Peter Hillmann, Marcus Knüpfer, Erik Heiland, and Andreas Karcher  
*Bundeswehr University Munich*

Werner-Heisenberg-Weg 39, 85579 Neubiberg, Germany

Email: {peter.hillmann, marcus.knuepfer, erik.heiland, andreas.karcher}@unibw.de

**Abstract**—The constantly growing size of blockchains becomes a challenge with the increasing usage. Especially the storage of unwanted data in a blockchain is an issue, because it cannot be removed naturally. In order to counteract this problem, we present the first concept for the selective deletion of single entries in a blockchain. For this purpose, the general consensus algorithm is extended by the functionality of regularly creating summary blocks. Previous data of the chain are summarized and stored again in a new block, leaving out unwanted information. With a shifting marker of the *Genesis Block*, data can be deleted from the beginning of a blockchain. In this way, the technology of the blockchain becomes fully transactional. The concept is independent of a specific block structure, network structure, or consensus algorithm. Moreover, this functionality can be adapted to current blockchains to solve multiple problems related to scalability. This approach enables the transfer of blockchain technology to further fields of application, among others in the area of Industry 4.0 and Product Life-cycle Management.

**Index Terms**—Blockchain, delete, erase, forgetting, oblivion, GDPR

## I. INTRODUCTION

The distributed ledger technology also known as the blockchain technology exists since 2008 [1]. It has become well-known through the crypto currencies Bitcoin and Ethereum. Satoshi Nakamoto solved the Byzantine Generals Problem alongside the invention of Bitcoin, which is a common problem in distributed systems [2]. The Ethereum network has extended the blockchain technology by the Turing completeness in 2015. It allows the execution of programs decentralized and to store so-called smart contracts in the blockchain. In this way, it is possible to bind crypto tokens to material goods instead of trading virtual goods. The increased public interest of the last years is still unbroken due to the high exchange rate of the stocks [3]. As a result more and more crypto currencies appeared. Nevertheless, the technology of blockchain is also of interest in other areas like Industry 4.0 [4] and Supply Chain [5]. In Estonia, the blockchain is used at national level to administer health data [6]. The reason for this are the unique properties, whose transactions is desirable in many areas. The advantages are decisive in terms of security, traceability and not requiring a third party. The underlying technology of the blockchain provides a method for untrusted participants to reach a consensus over a public untrusted network.

The most important properties of a blockchain are the trustworthiness and reliability based on the immutability of the data. By concatenating data sets in the form of blocks using the hash value, it is not possible to manipulate a block within the

blockchain. Thus the order as well as the content of the blocks are fixed. To ensure the absolute correctness of a blockchain, it has to be completely available, starting with the *Genesis Block*. Therefore the blockchain is often called unhackable. However, a change of the blockchain is possible, but becomes less likely with the increasing number of confirmation blocks.

Since the technology is still in the nascent stage, there are still problems with it:

- Blockchain attribute immutability in combination with illegal and unwanted content in a public blockchain
- Growth of the blockchain
- Laws with the right of erasure

The blockchain attribute immutability represents in these cases rather a problem than an advantage. Researchers discovered that the Bitcoin blockchain contains pornographic content [7]. This is possible because Bitcoin transactions allow a variety of non-Bitcoin related data to be stored alongside [8]. Since a full node must download the entire blockchain in order to work with the network, the illegal content is also downloaded and made available to others. As a result, some countries prohibit the operation of full nodes and make it liable to prosecution.

Further, the growing size of a blockchain makes the operation very difficult in long term. This problem becomes more intense as more people create transactions in a specific blockchain. Bitcoin, released in 2009, has almost reached a blockchain size of 300 GB [9]. Dependent on storage space and network bandwidth, the costs of operation increase significantly.

Last but not least, the right-to-erasure has to be respected, especially in case of personal data. Currently, there is no clear answer on how to deal with such problems.

With this paper, we present the first concept to delete unwanted data in a running blockchain. The blockchain will forget this information instead of a direct deletion to keep a trusted chain. This approach works independently of the network infrastructure and the consensus algorithm of a blockchain.

The remainder of this paper is structured as follows. In Section II, we describe a simplified scenario to identify the requirements of the problem. Section III discusses related work in relation to the requirements. The main part in Section IV presents the concept of extending the functionality of a blockchain to be fully transactional. Thereafter, Section V shows some experiments and practical usage, before Section VI concludes this paper.

## II. SCENARIO AND REQUIREMENTS

The General Data Protection Regulation (GDPR) was adopted in May 2016. The objective of this directive is to standardize the handling of personal data of European citizens both within and outside of the European Union. Among other things, especially the right to erasure was introduced in Art. 17 of the GDPR, also known as the right to be forgotten [10]. The German Blockchain Association published concerns for the treatment of blockchain technology under the GDPR law [11]. In addition to suggestions on the interpretation of the current law, they highlight future improvement possibilities. The right to erasure collides with the concept of the blockchain. Its immutability is not only not manipulable, but also does not allow intended changes in retrospect. Such a manipulation would also be desirable to the blockchain to delete illegal content saved on it [12].

To illustrate the problem we focus on the use case of logging and auditing. Logging describes the automated recording of system states, product data or other process information. Thus makes it possible to trace the actions by a computer system for humans. Especially in IT security, logging is important to detect weak points and to document security breaches or to investigate them afterwards. However, since log files can also be manipulated, it is mandatory that the authenticity of the log files is given. Otherwise it can not be used for a reasonable evaluation. Depending on the scenario, log entries should be stored for different lengths of time.

Based on this, we obtain the following requirements:

- **Authenticity:**  
The act of confirming the truth of an data entry.
- **Redundancy of data:**  
Storing the information multiple times to be fail-safe.
- **Delete on request:**  
A selective erasure of entries or blocks with less effort.
- **Scalability:**  
The system runs for a long time independent of the amount of users and the blockchain size.

In the following, we decide between deletion and forgetting of data. The difference is that deleting is on purpose while forgetting is a delayed procedure. Nevertheless, if a user of data systems do not follow the rules of erasure, any information published could be stored forever.

## III. RELATED WORK

To tackle the problem, different approaches have already been made and other technologies may useable to fulfil the requirements partially. The most promising techniques are the off-chain transactions, since interactions with the main chain are comparatively expensive, time-consuming and the whole network has to reach a consensus. During off-chain transaction, digital tokens are exchanged and moved without being recorded on the main chain [13]. In order to provide guarantee, *Binding Corporate Rules* can be used. But this approach does not solve the problem for global and public blockchains.

Another possibility is that not the private user data are stored in the blockchain, but only the hashes of the user data for possible verification. These could be shared off-chain, while the correctness can still be checked by the hashes. For example, payment channels and micropayment channels establish a connection between two parties or a closed community. This can use to perform any number of transactions without being held on the main chain [14, 15]. Theoretically, the data of a channel could be forgotten by all participants of the channel deleting the closed channel afterwards. The main chain is only involved in the beginning and extended with transactions at the end [16]. This also works by storing only asymmetrical encrypted data in the blockchain, while maintaining the keys off-chain. If the data needs to be retrieved from the ledger, the key has to be requested and the data must be decrypted first. The data can be effectively regarded as erased if the decryption key(s) are not available. This throttles the transactions per second problem as well as the memory scaling problem. Nevertheless, this does not fulfil the requirement of reducing the blockchain length, to shrink the data size and erase illegal content.

The approach of Proof-of-Burn can be used to transfer coins or data from one blockchain to another [17]. With a complete transfer of all important data, the old blockchain can be deleted. In context of a GDPR-compliant, it supports the right to erasure with data tokenisation. In a system, where tokens have to be generated and redeemed for each piece of data in order to process this data, a user requests to burn tokens associated to her account. Thus, the information embedded in the tokens are no longer usable. A selective deletion is not possible, especially of not account associated data. Furthermore the effort is very high.

The IBM identity mixer takes a different approach and does not try to disguise or delete the data [18]. It rather disguises the user before even storing the data [19]. However, one problem with this approach is that it is no longer possible to assign the data to a person. This excludes the legal way in case of misconduct.

Furthermore, there are many more approaches, which try to avoid the problem instead of providing an integrated solution. The simple solution of pruning locally stored parts does not solve the problem for the global, distributed blockchain [20]. Another possibility is a hard fork to a new blockchain [21] after unwanted content is stored. But this is very time inefficient as it can take place on every transaction. Approaches based on chameleon-hash functions leave the responsibility with the key owners and produce a lot effort [22, 23].

All in all, the presented techniques allow to store the data for example in hashed ways or in a way where the data subject is obfuscated. No presented solution solves the problem that the data on the distributed blockchain cannot be deleted, and thus the requirements for the use case are not fulfilled.

## IV. CONCEPT

In this paper, we present the first concept to create the possibility to delete information from a living blockchain

system. This allows the blockchain technology to become a full transactional datastore with high trust and all the other established properties. Furthermore, this approach ensures that all blockchain systems will remain scalable and manageable even if they run for a long time.

In order to fulfil the requirements, we extend the functionality of the blockchain by creating summary blocks at regular intervals. Afterwards, we present the concept, which gives the users of the blockchain the possibility to delete their data on request. This is done after a certain time or with delay. Therefore, we then describe the process to limit the length of the blockchain and to delete on request. This is followed by an evaluation, in which we discuss the improvements achieved as well as strengths and weaknesses.

### A. Background and Consensus Algorithm

A direct deletion of an individual entry or a block destroys the hash chain of a blockchain. This leads to inconsistency, corruption and the loss of trustworthiness. The information could no longer be completely traced back to the first block anymore. The first block is called *Genesis Block*. To be able to delete individual entries on request, the blockchain have to be able to forget or lose information with intent. Therefore a novel concept is necessary, whereas the action of deletion is approved by the trusted anchor nodes. These can be seen as the guardians of the blockchain. For the election of the group of these trusted nodes, several community based approaches can be applied [24]. This depends on the type of the blockchain: public, private, consortium, hybrid. For example, the trusted community could consist of a non-profit organisation or participated users, who have previously done transaction in the blockchain.

One of the main processes of the distributed system is to agree on a common consensus. The presented concept is based on this functionality, independent of the specific consensus algorithm. We rely on the outcome of the trustworthy common agreement of the system's nodes for the same knowledge of each nodes. The agreement on the same blockchain is usually done by some core nodes, called anchor nodes. These node manage the full copy of the blockchain and build the quorum. All reliability in a blockchain starts with the common agreement of the quorum on the *Genesis Block* of the blockchain at the beginning. Among other things, the block number  $\alpha$  addresses a specific block in the chain. It is used to execute a certain function on fixed blocks by each node. Through the consensus algorithm, it can be assumed that each node has the same level of information [25].

### B. Extension with summary blocks

In the first step, the blockchain is extended by a special block type. Such a block is called summary block  $\Sigma$ . It consists of deterministic information only and does not include external or additional information. A summary block is periodically created by each node itself, for example every tenth block. These blocks are integrated in the blockchain like normal blocks. Figure 1 shows an example of a summary block  $\Sigma$ .

It can be noted that the block number  $\alpha_\Sigma$  of the summary block is increased by one as normal blocks. The summary block has the same timestamp  $\tau$  as the block before. This is mandatory so that each node can calculate the summary block with its hash values by itself.

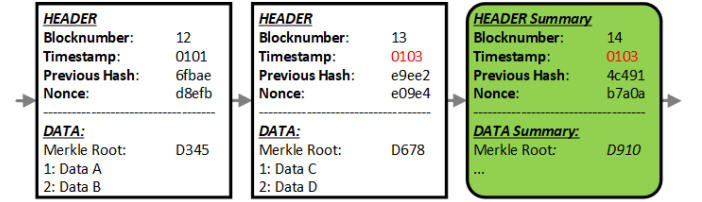


Fig. 1. Example of extending the blockchain with a summary block, visualized as green block with round corners.

As each anchor node has the same agreed blockchain, the summary block will be identical for each node. So this information can be used to check synchronisation by comparing the hash of its summary block. Beneficially, the block do not need to be propagated by itself. The selected consensus algorithm of a blockchain system has to be extended by this behavior.

One difficulty with summary blocks is that the nodes must create them themselves and still have exactly the same header. As all nodes work together with the same consensus algorithms, this should be always the case. In case of a failure, the hash of the blocks are different, which would result in a fork in the blockchain and thus split the network.

### C. Limiting the Sequences in a Blockchain

By inserting the summary blocks, the blockchain can be seen as separated into sequences.

A sequence  $\omega$  is a series of blocks including the summary block at the end of each sequence. The length  $l_n$  is defined as the number of blocks in a sequence  $\omega_n$ . A representation of the sequences is given in Figure 2. The lengths  $l$  of the individual sequences  $\omega_n$  is defined constant depending on the application, but do not have to be. Regardless of the configurable parameter, the distance between two summary blocks is described by  $\delta_l$ , the length  $l$  of a sequence  $\omega$ .

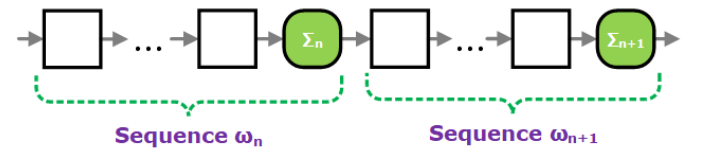


Fig. 2. Consideration of the sequences of a blockchain, defined by the summary blocks.

Now, we are able to reduce or shrink the length  $l_\beta$  of the blockchain  $\beta$  and therefore the size of it. The quorum agrees on a defined blockchain length  $l_{max}$ . Alternatively, another property can be used, for example the maximum number of sequences  $\omega$  in the living blockchain. If the blockchain grows larger than the specified length  $l_{max}$ , the oldest sequence will be merged into the next summary block. Therefore, the same

information of all single blocks of the first sequence is copied into the new summary block at the end of the chain. This includes the information of the old summary block to be deleted. The copied information consists of the data part of the block as well as the block number and timestamp. The *nonce* and *previous hash* of a block are not needed anymore. After that, the quorum build a consensus about redefining the *Genesis Block*, based on the blocks of the current chain. By a majority vote, the quorum determines the new first Block and the time of the changeover. The new *Genesis Block* will be the block after the summary block of the sequence to be deleted. A *Marker m* is used to indicate the shifting *Genesis Block*, holding the block number. As this block is part of the blockchain, it is a trusted anchor for the left blockchain part already approved by the anchor nodes. Afterwards, the old sequence can be cut off and deleted from the blockchain. As currently the same information is still in the blockchain, the chain is still valid and complete. Consequently, this procedure is based on sequences, not on single blocks. Figure 3 visualizes the process of summarizing the oldest sequence into the new summary block similar to the round-robin principle. The green colored and rounded blocks  $\Sigma$  are the summary blocks and the shaded blocks are noted to be deleted.

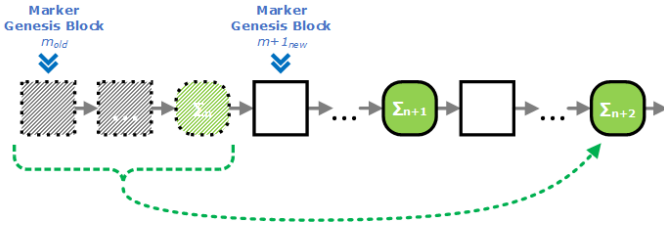


Fig. 3. Summarization after exceeding a specified chain length: The first sequence will be deleted after their content is copied into the new summary block.

The data part of a summary block has the following structure as in Figure 4, whereas the content is copied from the original blocks. So, the block number, the timestamp and the entry number are kept the same as initially integrated.

#### D. Deleting information on request

To enable the deletion of specific information, a participant needs to be granted the additional functionality to send a delete request. For this purpose, a user submits such a request to the blockchain system in form of a deletion entry or a deletion transaction. This follows the same procedure as normal entries and transactions. The associated data set to be deleted is referenced by the block number and the according entry number, in which the data set is stored. It may happen that an entry is located in a summary block. This must be taken into account by also considering the data records in summary blocks. The complexity of the procedure is linear and very low as a blocks are references directly by number. Regarding the deletion of information, the following two aspects of authorization and semantic cohesion need to be considered. Furthermore, the challenge of delayed deletion and the possibility of temporary entries is addressed.

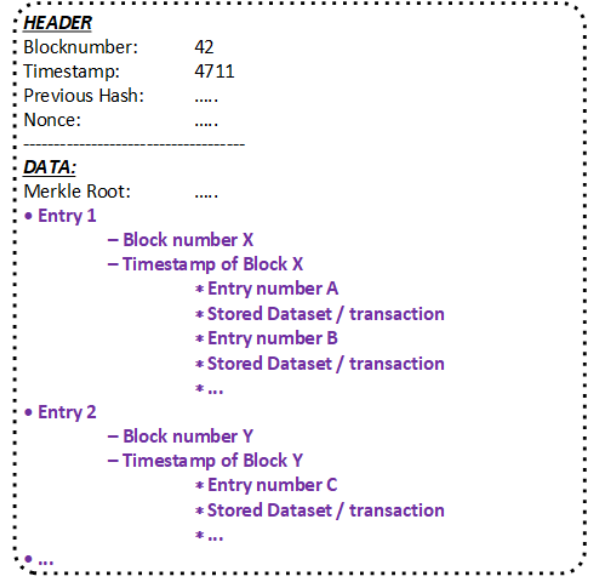


Fig. 4. Data structure of the summary blocks.

1) *Authorization*: To ensure that the user is authorized to have the information deleted, a deletion request must be signed with the client signature just like a normal entries. For authorization of privileges, it can be applied a role-based concept with corresponding user signatures. Particular reference is made to authentication approaches based on blockchains [26, 27]. For example, the anchor nodes of the quorum work together as a basis of trust and are jointly granted full administrative privileges. These receive a master signature. Based on a protocol for consensus-based voting, the quorum dictates the requirements and constraints for edition of the editing [28]. In comparison, a user is only allowed to submit delete requests for his own transactions. These can be identified by comparing the signature of the user and the stored signature of an data entry. The system checks if the signatures share the same key and the privileges are given according the role-based concept. According to the consensus of the anchor nodes, a deletion request is approved and entered in the blockchain.

2) *Semantic Cohesion*: The semantic correctness of the complete blockchain has to be proven by the quorum before a deletion takes place. A deletion request can only be granted, if further transactions do not rely on it. Otherwise, multiple transactions need to be revoked, which may involves additional parties. A deletion request of such a chain part of a transaction chain can be approved by the signatures of all dependent parties. So these entries become orphaned in the meaning of the content [29]. At this point, the quorum must ensure that the cohesion in the blockchain is maintained and remains valid after the deletion. This depends strongly on the use case and the specific scenario. Therefore, a deletion request has to be proven manually. An automatic approached could be designed based on the principle of *Bell-LaPadula model* or *Brewer-Nash Model*.

3) *Delayed Deletion*: In this case, deletion means that the information is not deleted immediately. For the moment, the specified data is marked to be deleted in the future. Subsequent incoming transactions based on this marked data are no longer permitted. A marked data set will be deleted when the sequence gets too long or too old, see Equation 1. This is the case when the marked data reaches the beginning of the blockchain (see Figure 5). As a result, the sequence is considered old enough and the information to be deleted is no longer included in the corresponding summary block.

$$\text{if } l_{\beta} > l_{max}, \text{ then } l_{\beta_{new}} = l_{\beta_{old}} - l_{\omega_1}, \quad (1)$$

(whereas  $\omega_1$  is the beginning of  $\beta$ )

Figure 5 shows how an entry is deleted with delay by the blockchain using a deletion request.

Since there is no need to keep deletion requests longer than necessary they will never be copied into a summary block. To prevent a long delay in deletion, a possibility is to extend the blockchain with empty blocks. This can be done, by regularly adding empty blocks after a time interval if no transaction has occurred.

To avoid shortening the blockchain too much, a minimum length or a minimum number of summary blocks can be specified for the remaining blockchain. Another criterion for ensuring that sufficient data remains in the blockchain is a minimum time span coverage. So, there are two possibilities to determine the length of the blockchain: by time and by (summary) block quantity. New sequence in a blockchain may be created, although the oldest sequence is not yet old enough to be deleted. This creates empty summary blocks. This case for example occur at the beginning of the blockchain. However, multiple sequences can also being combined in one summary block.

4) *Temporary entries*: The current design can be extended by temporary entries, which will be deleted automatically after a specific timestamp ( $\tau$ ) or block number ( $\alpha$ ) has been reached. Therefore, an entry is extended by an optional data field in which the maximum storage time is recorded (e.g. 42: Data T :  $\tau 8888$  or  $\alpha 4711$ ). If the blockchain exceeds the timestamp or block number given, the entry will not be transferred to the new summary block. So it is automatically removed from the blockchain without additional authorization needed. The system cleans up its own content and deletes the temporary information. Use cases for this functionality include log files of operating systems or products traced in a blockchain according to Industry 4.0 [30].

## V. EVALUATION

To evaluate our concept for selective deletion in a blockchain, we have implemented a prototype of the system. The implementation is realized as a client-server system using Python and Java. In order to be language independent and to get reusable components, the middleware technology of CORBA is used. The blockchain can be seen as a list of blocks maintained by server nodes, namely the anchor nodes. If the sequence reaches its defined length, a summary block will be

created automatically. This includes the check about blocks to be deleted.

To evaluate the concept, we focus on the use case of logging as mentioned in Section II. All logins to a terminal are logged to the blockchain. Therefore, the signature of each specific user login is stored in a block. In this way, the authentication of the user can be monitored and audited. The blockchain system is configured to create a summary block for every third block. The persons in the scenario are referred to as *ALPHA*, *BRAVO* and *CHARLIE*. In the following test setup, the specific signatures of each client prevent a user from deleting an entry of another user.

To visualize the blockchain, the entries are listed line by line. Each block has the following header structure: block number; timestamp; previous block hash; own block hash; optional data entry. An data entry is structured as follows: D stores data record; K holds the user; S poses as signature (here simplified). The structure of a data entry is specified beforehand by a YAML schema. For simplification, blocks starting with S are the summary blocks. These are created automatically and marked in blue.

Figure 6 shows the *Genesis Block 0* with previous hash *DEADB*. The first two summary blocks are empty, because there was nothing to add up from previous sequences yet. The maximum length of the blockchain is not reached, so nothing gets deleted. Each user has an entry in block 1,3 and 4.

Figure 7 shows a request on deletion in block number 6. The user *BRAVO* requests the deletion of its entry in block number 3 and data entry 1. Since the issuer of the entry to be deleted matches the user of the deletion request, the corresponding entry is not transferred to the summary block at the next action of deletion. Furthermore, the first and second sequence of the blockchain is merged into the last summary block. The entry from block number 3 is not copied. The maker for the *Genesis Block* is changed to block number 6. All information before block 6 is deleted.

Nevertheless, wrong request of deletions can be included in the blockchain, but these have no further effects.

Figure 8 presents the blockchain one cycle ahead. The deletion request entry of *BRAVO* is no longer stored in the current blockchain, as deletion entries are never transferred.

### A. Achieved Enhancements

Our concept has multiple advantages, offers additional possibilities, and solves the growing data problem:

- **Data Reduction:**

Deleting and forgetting unwanted content by not copying the information to the new summary block.

- **Corrections:**

Change information, which maybe submitted wrongly.

- **Reorganization:**

Possibility to reorganize information for better scalability.

- **Summarized Information:**

The ability to summarize coherent information. E.g., if a system logs an event several times, these logs can be stored summarized in the blockchain.



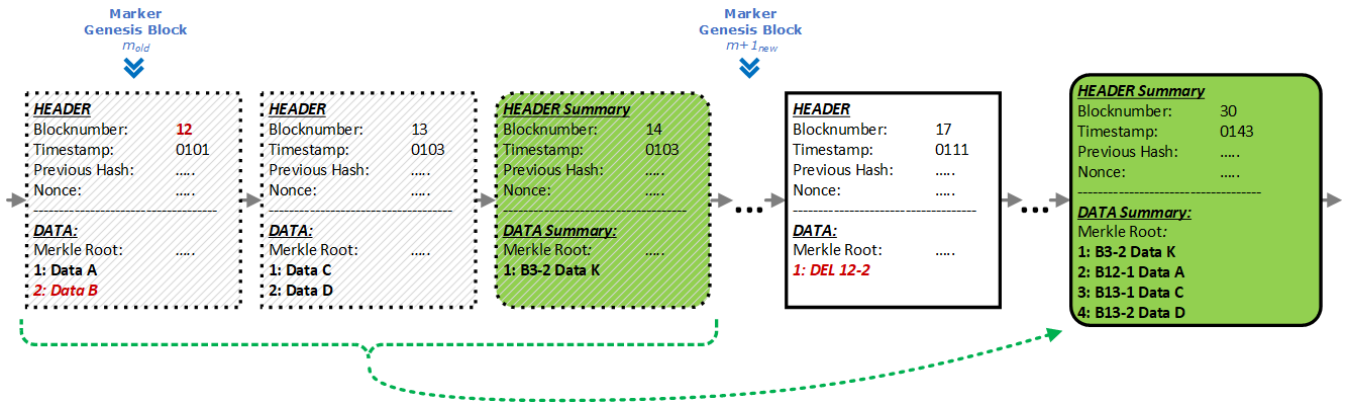


Fig. 5. Selective deletion on request. The colored marked entry is not copied into the new summary block due to the delete request. The first sequence will be deleted after their content is copied.

```
0,0,DEADB,2dfe9
1,1537034677447,2dfe9,c3950
-1: D:login@terminal1; K:ALPHA; S:login@terminal1::ALPHA
S: 2,1537034677447,c3950,82db6
3,1537034706696,82db6,acd66
-1: D:login@terminal3; K:BRAVO; S:login@terminal3::BRAVO
4,1537034729910,acd66,582b7
-1: D:login@terminal7; K:CHARLIE; S:login@terminal7::CHARLIE
S: 5,1537034729910,582b7,5426a
```

Fig. 6. Console output shows the current status of the content in the block after three logins.

```
6,1537034764184,5426a,2d40f
-1: B#:3; E#:1; K:BRAVO; S:4::BRAVO
7,1537034809070,2d40f,f5325
-1: D:login@terminal2; K:ALPHA; S:login@terminal2::ALPHA
S: 8,1537034809070,f5325,4d158
-1: I:(1, 1); D:login@terminal1; K:ALPHA; S:login@terminal1::ALPHA
-2: I:(4, 1); D:login@terminal7; K:CHARLIE; S:login@terminal7::CHARLIE
```

Fig. 7. The first two sequences are summarized into the latest summary block, whereby the entry to be deleted was not copied.

• **Recovery:**

In the case of cryptocurrencies, it offers the possibility to make lost coins usable again. It means not for a single user, but for the entire blockchain system to prevent a system shutdown in long term [31].

*B. Discussion: Strengths and Weaknesses*

To realize the possibility of deletion in a blockchain, certain trade-offs are made.

1) *Hamper 51 % attack:* The most known attack on a blockchain is the 51 % attack [32], where an attacker controls more than 50 % of the resources to create new blocks or

```
9,1537034863483,4d158,1a2e3
-1: D:login@terminal5; K:CHARLIE; S:login@terminal5::CHARLIE
10,1537034877547,1a2e3,0cc96
-1: D:login@terminal3; K:CHARLIE; S:login@terminal3::CHARLIE
S: 11,1537034877547,0cc96,eff78
-1: I:(7, 1); D:login@terminal2; K:ALPHA; S:login@terminal2::ALPHA
-2: I:(1, 1); D:login@terminal1; K:ALPHA; S:login@terminal1::ALPHA
-3: I:(4, 1); D:login@terminal7; K:CHARLIE; S:login@terminal7::CHARLIE
```

Fig. 8. Console output one cycle ahead without deletion request in the summary block.

manipulate the block order. The precise amount depends on the consensus algorithm. This attack is also the base for enhanced attacks such as double spending and feather forking. During the process of a blockchain, each newly added block is a confirmation of all previous blocks. Since the old sequence in which the information was stored is deleted, there is no possibility for new nodes to validate the correctness of the most current summary block. To counter this possibility, we make use of the summary blocks to store data more than once, or at least the Merkle root as reference for validity to reduce the amount of data. Therefore, the sequence to be deleted and the reference to a middle sequence, for example  $\omega_{l_\beta/2}$ , is stored in the new summary block, see Figure 9. So, each entry that is longer than  $l_\beta/2$  in the blockchain has at least  $l_\beta/2$  confirmations at each time. The trustworthiness is given, even if the oldest entry is combined in the latest summary block, This hampers the 51 % attack against a blockchain without summary blocks as for an attacker it is not enough to run the attack for only one block. The attacker has to run the attack for a least  $l_\beta/2$  number of blocks. The longer the blockchain is and the more summary blocks there are, and consequently the more trustworthy the chain is.

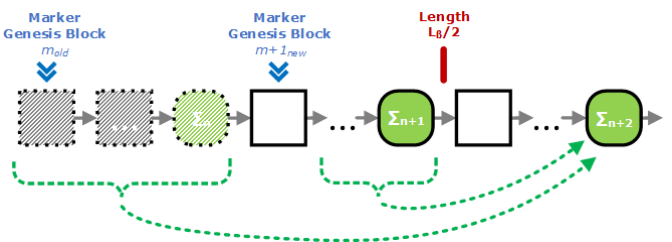


Fig. 9. Hampering the 51 % attack by storing data more than once in the blockchain

2) *Size of summary blocks:* By adding up the information in summary blocks, they become larger over time. The creation of these summary blocks can take a long time, depending on the amount of data to be copied and the storage performance. One possibility to counter this is to structure the information

logically and build packages. In addition, the copying of much information can be avoided by working with hash references. The data packets are stored separately and only linked in the blockchain, as with other off-chain approaches. Another opportunity is to summarize the content as often the result is mandatory to be stored, not every single transaction.

3) *Extending Consensus algorithm*: Any consensus algorithm can be extended by the described behavior. Nevertheless, it needs to be ensured that the participants do not evaluate the blockchain based on the length or the value of the block index as the chain can be shortened. This can be prevented if the nodes only accept a blockchain which is traceable from its current status quo.

4) *Isolating nodes or evil nodes*: Node isolation is an attack where an attacker tries to separate participants from the network by controlling partner nodes of the target. Known attacks are Eclipse attack [33] or Sybil attack [34]. Thus, an attacker can weaken or completely destroy the network piece by piece. Therefore, the blockchain system has to have some anchor nodes, whereas clients obtain the current status quo of the blockchain.

5) *Ascertaining the quorum and reaching a consensus*: The problem of quorum building is a long known problem in computer science. There are different methods to solve this problem, especially for the selection of anchor nodes for distributed systems [35]. Depending on the type of the blockchain, these procedures offer different advantages and disadvantages. A decision has to be made individually. This also applies to the selection of a suitable consensus algorithm. The presented approach in this publication is independent of the concrete characteristics of quorum selection and consensus algorithm.

## VI. SUMMARY

In this paper, we presented the first concept to delete entries in a blockchain while maintaining the characteristic attributes like trustworthiness and consistency. In particular, the requirements of laws to delete personal data can be met such as defined in the European Data Protection Act (GDPR). In summary, the concept is based on an extension of the behavior of the consensus algorithm. The blockchain is divided into sequences by regularly adding summary blocks. These summarize the old information of the first sequences and store the data at the current end of the blockchain. During this process, invalid, obsolete, or to be deleted entries are ignored. Afterwards, the marker for the *Genesis Block* is shifted to the following sequence. The first sequence can be forgotten and will be deleted. So, the blockchain only contains data that is relevant for the current and future time. This solves the problem of a growing blockchain size. Furthermore, the blockchain technology becomes full transactional. A deletion request requires a certain amount of time to be processed, which can be seen as a disadvantage. But it is mandatory to avoid illegal manipulations. In the end, the system is not as dynamic as classical database systems. Nevertheless, the block structure as well as the network are independent of

this solution. Also, the added functions are independent of the consensus itself and can be combined with different strategies. So, the approach is flexible adaptable to extend all blockchain systems as well as already running systems like Bitcoin.

Typical fields of application are not limited to the area of cryptocurrencies to cope with the increasing size of the blockchain. In the area of Industry 4.0, the production of a good can be recorded along the entire supply chain. As soon as the minimum best-before date has been exceeded or the data has expired, the new technology can be used to automatically clean up the blockchain. In the field of vehicle maintenance, the life cycle of each car can be documented centrally, so that manipulations are excluded, e.g. on the milage or accidents. After a vehicle is taken out of service, the blockchain as database is cleaned up to handle the data amount. In the future, we will implement the concept in upcoming blockchain frameworks like Hyperledger, Ethereum or the Enterprise blockchain: MultiChain.

## ACKNOWLEDGMENT

Special thanks goes to Simon Ostendorff for the development of the basic prototype and his passionate support for this research.

## REFERENCES

- [1] Satoshi Nakamoto. Bitcoin: A Peer-to-Peer Electronic Cash System. 2008. URL <https://bitcoin.org/bitcoin.pdf>. (Date last accessed 28-February-2020).
- [2] Leslie Lamport, Robert Shostak, and Marshall Pease. The Byzantine Generals Problem. 1982. URL <https://people.eecs.berkeley.edu/~luca/cs174/byzantine.pdf>. (Date last accessed 28-February-2020).
- [3] buybitcoinworldwide. Bitcoin Kurshistorie-Diagramm. 2020. URL <https://www.buybitcoinworldwide.com/de/preis/>. (Date last accessed 28-February-2020).
- [4] Martin Laabs and Sinisa Dukanovic. Blockchain in Industrie 4.0: Beyond cryptocurrency. 2018.
- [5] Kari Korpela, Jukka Hallikas, and Tomi Dahlberg. Digital Supply Chain Transformation toward Blockchain Integration. 2017. URL <https://scholarspace.manoa.hawaii.edu/handle/10125/41666>. (Date last accessed 28-February-2020).
- [6] Taavi Einaste. Gesundheitswesen: estnische Erfahrungswerte. 2018. URL <https://nortal.com/de/blog/blockchain-healthcare-estonia/>. (Date last accessed 28-February-2020).
- [7] Martin Holland. Bitcoin: Forscher finden Kinderpornographie in der Blockchain. 2018. URL <https://www.heise.de/newsticker/meldung/Bitcoin-Forscher-finden-Kinderpornographie-in-der-Blockchain-4000693.html>. (Date last accessed 28-February-2020).
- [8] Roman Matzutt, Jens Hiller, Martin Henze, Jan Henrik Ziegeldorf, Dirk Müllmann, Oliver Hohlfeld, and Klaus Wehrle. A Quantitative Analysis of the Impact of Arbitrary Blockchain Content on Bitcoin. 2018.

- URL <https://fc18.ifca.ai/preproceedings/6.pdf>. (Date last accessed 28-February-2020).
- [9] Blockchair. Blockchain size. 2020. URL <https://blockchair.com/ethereum/charts/blockchain-size>. (Date last accessed 28-February-2020).
- [10] Bundesamt für Sicherheit in der Informationstechnik. Right to erasure ('right to be forgotten'). 2013. URL <https://gdpr-info.eu/art-17-gdpr/>. (Date last accessed 28-February-2020).
- [11] Natalie Eichler, Silvan Jongerius, Greg McMullen, Oliver Naegele, Liz Steininger, and Kai Wagner. Blockchain, data protection, and the GDPR. 2018. URL [https://www.bundesblock.de/wp-content/uploads/2018/05/GDPR\\_Position\\_Paper\\_v1.0.pdf](https://www.bundesblock.de/wp-content/uploads/2018/05/GDPR_Position_Paper_v1.0.pdf). (Date last accessed 28-February-2020).
- [12] Samuel Gibbs. Child abuse imagery found within bitcoin's blockchain. *The Guardian*, 2018. URL <https://www.theguardian.com/technology/2018/mar/20/child-abuse-imagery-bitcoin-blockchain-illegal-content>. (Date last accessed 28-February-2020).
- [13] bitcoinwiki. Off-Chain Transactions. 2017. URL [https://en.bitcoin.it/wiki/Off-Chain\\_Transactions](https://en.bitcoin.it/wiki/Off-Chain_Transactions). (Date last accessed 28-February-2020).
- [14] bitcoinwiki. Payment channels. 2018. URL [https://en.bitcoin.it/wiki/Payment\\_channels](https://en.bitcoin.it/wiki/Payment_channels). (Date last accessed 28-February-2020).
- [15] Joseph Poon and Vitalik Buterin. Plasma: Scalable Autonomous Smart Contracts. 2017. URL <https://plasma.io/plasma.pdf>. (Date last accessed 28-February-2020).
- [16] Søren Debois, Alevtina Dubovitskaya, Avigdor Gal, Petr Novotny, Stefanie Rinderle-Ma, Stefan Schulte, Ludwig Stage, and Kaiwen Zhang. Data Technology to the Rescue: Digging the D in GDPR. 2018.
- [17] P4Titan. Slimcoin: A Peer-to-Peer Cryptocurrency with Proof-of-Burn. 2014. URL <https://www.chainwhy.com/upload/default/20180703/4ae7cee40462e7951f508b28dd1d9936.pdf>. (Date last accessed 28-February-2020).
- [18] IBM. Identity Mixer. 2018. URL [https://www.zurich.ibm.com/identity\\_mixer/](https://www.zurich.ibm.com/identity_mixer/). (Date last accessed 28-February-2020).
- [19] R. Matzutt, M. Henze, J. H. Ziegeldorf, J. Hiller, and K. Wehrle. Thwarting unwanted blockchain content insertion. page 364–370.
- [20] Martin Florian, Sebastian A. Henningsen, Sophie Beaucamp, and Björn Scheuermann. Erasing Data from Blockchain Nodes. pages 367–376.
- [21] G. Ateniese, B. Magri, D. Venturi, and E. Andrade. Redactable blockchain—or—Rewriting history in Bitcoin and friends. page 111–126.
- [22] Jan Camenisch, David Derler, Stephan Krenn, Henrich C. Pöhls, Kai Samelin, and Daniel Slamanig. Chameleon-hashes with ephemeral trapdoors. page 152–182.
- [23] Ivan Puddu, Alexandra Dmitrienko, and Srdjan Capkun.  $\mu$ chain: How to Forget without Hard Forks. Cryptology ePrint Archive, Report 2017/106, 2017.
- [24] Duncan Black. *The theory of committees and elections*.
- [25] Alyssa Hertig. 2x Called Off: Bitcoin Hard Fork Suspended for Lack of Consensus. 2017. URL <https://www.coindesk.com/2x-called-off-bitcoin-hard-fork-suspended-lack-consensus/>. (Date last accessed 28-February-2020).
- [26] Rohas Nagpal. Blockchain-based authentication of devices and people. 2018. URL <https://medium.com/blockchain-blog/blockchain-based-authentication-of-devices-and-people-c7efcfcf0b32>. (Date last accessed 28-February-2020).
- [27] A.H. Mohsin, A.A. Zaidan, B.B. Zaidan, O.S. Albahri, A.S. Albahri, M.A. Alsalem, and K.I. Mohammed. Blockchain authentication of network applications: Taxonomy, classification, capabilities, open challenges, motivations, recommendations and future directions. 2019. URL <https://www.sciencedirect.com/science/article/abs/pii/S0920548918303477>. (Date last accessed 28-February-2020).
- [28] Dominic Deuber, Bernardo Magri, and Sri Aravinda Krishnan Thyagarajan. Redactable Blockchain in the Permissionless Setting. pages 645–659.
- [29] Shourya Shirsha Nandi. A walk through the evolution of dlts: From blockchain to hashgraph and tangle. 2018. URL <https://medium.com/@micobo/april-2018-3f575ef0cb31>. (Date last accessed 28-February-2020).
- [30] A. Banerjee. Infosys: Product tracking and tracing with iot and blockchain. 2018. URL <https://www.infosys.com/Oracle/insights/Documents/product-tracking-tracing.pdf>. (Date last accessed 28-February-2020).
- [31] Jeff John Roberts and Nicolas Rapp. Exclusive: Nearly 4 million bitcoins lost forever, new study says. 2017. URL <https://fortune.com/2017/11/25/lost-bitcoins/>. (Date last accessed 28-February-2020).
- [32] P. H. Madore. Binance: Monero Household Mining Unprofitable, Vulnerable to 51Attack Post-Hard Fork. *CCN*, 2019. URL <https://www.ccn.com/binance-monero-mining-unprofitable-51-attack/>. (Date last accessed 28-February-2020).
- [33] Ethan Heilman, Alison Kendler, Aviv Zohar, and Sharon Goldberg. Eclipse Attacks on Bitcoin's Peer-to-Peer Network. 2015. URL <https://eprint.iacr.org/2015/263.pdf>. (Date last accessed 28-February-2020).
- [34] John R. Douceur. The Sybil Attack. 2002. URL <https://www.freehaven.net/anonbib/cache/sybil.pdf>. (Date last accessed 28-February-2020).
- [35] Dale Skeen. A Quorum-Based commit protocol. Technical report, Cornell, TR 82-483.