Updatable Tokenization: Formal Definitions and Provably Secure Constructions *

Christian Cachin, Jan Camenisch,
Eduarda Freire-Stögbuchner, and Anja Lehmann

IBM Research – Zurich
(cca|jca|efr|anj)@zurich.ibm.com

Abstract. Tokenization is the process of consistently replacing sensitive elements, such as credit cards numbers, with non-sensitive surrogate values. As tokenization is mandated for any organization storing credit card data, many practical solutions have been introduced and are in commercial operation today. However, all existing solutions are static yet, i.e., they do not allow for efficient updates of the cryptographic keys while maintaining the consistency of the tokens. This lack of updatability is a burden for most practical deployments, as cryptographic keys must also be re-keyed periodically for ensuring continued security. This paper introduces a model for updatable tokenization with key evolution, in which a key exposure does not disclose relations among tokenized data in the past, and where the updates to the tokenized data set can be made by an untrusted entity and preserve the consistency of the data. We formally define the desired security properties guaranteeing unlinkability of tokens among different time epochs and one-wayness of the tokenization process. Moreover, we construct two highly efficient updatable tokenization schemes and prove them to achieve our security notions.

1 Introduction

Increasingly, organizations outsource copies of their databases to third parties, such as cloud providers. Legal constraints or security concerns thereby often dictate the de-sensitization or anonymization of the data before moving it across borders or into untrusted environments. The most common approach is so-called *tokenization* which replaces any identifying, sensitive element, such as a social security or credit card number, by a surrogate random value.

Government bodies and advisory groups in Europe [7] and in the United States [11] have explicitly recommended such methods. Many domain-specific industry regulations require this as well, e.g., HIPAA [15] for protecting patient information or the Payment Card Industry Data Security Standard (PCI DSS) [12] for credit card data. PCI DSS is an industry-wide set of guidelines that must be met by any organization that handles credit card data and mandates

* An extended abstract of this work was published at Financial Crypto 2017. This is the full version.

that instead of the real credit card numbers only the non-sensitive tokens are stored.

For security, the tokenization process should be *one-way* in the sense that the token does not reveal information about the original data, even when the secret keys used for tokenization are disclosed. On the other hand, usability requires that a tokenized data set preserves *referential integrity*. That is, when the same value occurs multiple times in the input, it should be mapped consistently to the same token.

Many industrial white papers discuss solutions for tokenization [13, 14, 16], which rely on (keyed) hash functions, encryption schemes, and often also non-cryptographic methods such as random substitution tables. However, none of these methods guarantee the above requirements in a *provably secure* way, backed by a precise security model. Only recently an initial step towards formal security notions for tokenization has been made [6].

However, all tokenization schemes and models have been *static* so far, in the sense that the relation between a value and its tokenized form never changes and that the keys used for tokenization cannot be changed. Thus, *key updates* are a critical issue that has not yet been handled. In most practical deployments, all cryptographic keys must be re-keyed periodically for ensuring continued security. In fact, the aforementioned PCI DSS standard even mandates that keys (used for encryption) must be rotated at least annually. Similar to proactively secure cryptosystems [9], periodic updates reduce the risk of exposure when data leaks gradually over time. For tokenization, these key updates must be done in a consistent way so that already tokenized data maintains its referential integrity with fresh tokens that are generated under the updated key. None of the existing solutions allows for efficient key updates yet, as they would require to start from scratch and tokenize the complete data set with a fresh key. Given that the tokenized data sets are usually large, this is clearly not desirable for real-world applications. Instead the untrusted entity holding the tokenized data should be able to re-key an already tokenized representation of the data.

Our Contributions. As a solution for these problems, this paper introduces a model for *updatable tokenization (UTO)* with *key evolution*, distinguishes multiple security properties, and provides efficient cryptographic implementations. An updatable tokenization scheme considers a data *owner* producing data and tokenizing it, and an untrusted *host* storing tokenized data only. The scheme operates in *epochs*, where the owner generates a fresh tokenization key for every epoch and uses it to tokenize new values added to the data set. The owner also sends an *update tweak* to the host, which allows to "roll forward" the values tokenized for the previous epoch to the current epoch.

We present several formal security notions that refine the above security goals, by modeling the evolution of keys and taking into consideration adaptive corruptions of the owner, the host, or both, at different times. Due to the temporal dimension of UTO and the adaptive corruptions, the precise formal notions require careful modeling. We define the desired security properties in the form of *indistinguishability* games which require that the tokenized representa-

tions of two data values are indistinguishable to the adversary unless it trivially obtained them. An important property for achieving the desired strong indistinguishability notions is *unlinkability* and we clearly specify when (and when not) an untrusted entity may link two values tokenized in different epochs. A further notion, orthogonal to the indistinguishability-based ones, formalizes the desired one-wayness property in the case where the owner discloses its current key material. Here the adversary may guess an input by trying all possible values; the *one-wayness* notion ensures that this is also its best strategy to reverse the tokenization.

Finally, we present two efficient UTO constructions: the first solution ($\mathsf{UTO}_{\mathsf{SE}}$) is based on symmetric encryption and achieves one-wayness, and indistinguishability in the presence of a corrupt owner *or* a corrupt host. The second construction ($\mathsf{UTO}_{\mathsf{DL}}$) relies on a discrete-log assumption, and additionally satisfies our strongest indistinguishability notion that allows the adversary to (transiently) corrupt the owner *and* the host. Both constructions share the same core idea: First, the input value is hashed, and then the hash is encrypted under a key that changes every epoch.

We do not claim the cryptographic constructions are particularly novel. The focus of our work is to provide formal foundations for key-evolving and updatable tokenization, which is an important problem in real-world applications. Providing clear and sound security models for practitioners is imperative for the relevance of our field. Given the public demands for data privacy and the corresponding interest in tokenization methods by the industry, especially in regulated and sensitive environments such as the financial industry, this work helps to understand the guarantees and limitations of efficient tokenization.

Related Work. A number of cryptographic schemes are related to our notion of updatable tokenization: key-homomorphic pseudorandom functions (PRF), oblivious PRFs, updatable encryption, and proxy re-encryption, for which we give a detailed comparison below.

A key-homomorphic PRF [4] enjoys the property that given $\mathsf{PRF}_a(m)$ and $\mathsf{PRF}_b(m)$ one can compute $\mathsf{PRF}_{a+b}(m)$. This homomorphism does not immediately allow convenient data updates though: the data host would store values $\mathsf{PRF}_a(m)$, and when the data owner wants to update his key from a to b, he must compute $\Delta_m = \mathsf{PRF}_{b-a}(m)$ for each previously tokenized value m. Further, to allow the host to compute $\mathsf{PRF}_b(m) = \mathsf{PRF}_a(m) + \Delta_m$, the owner must provide some reference to which $\mathsf{PRF}_a(m)$ each Δ_m belongs. This approach has several drawbacks: 1) the owner must store all previously outsourced values m and 2) computing the update tweak(s) and its length would depend on the amount of tokenized data. Our solution aims to overcome exactly these limitations. In fact, tolerating 1)+2), the owner could simply use any standard PRF, re-compute all tokens and let the data host replace all data. This is clearly not efficient and undesirable in practice.

Boneh et al. [4] also briefly discuss how to use such a key-homomorphic PRF for updatable encryption or proxy re-encryption. Updatable encryption can be seen as an application of symmetric-key proxy re-encryption, where the

proxy re-encrypts ciphertexts from the previous into the current key epoch. Roughly, a ciphertext in [4] is computed as $C = m + \mathsf{PRF}_a(N)$ for a nonce N, which is stored along with the ciphertext C. To rotate the key from a to b, the data owner pushes $\Delta = b - a$ to the data host which can use Δ to update *all* ciphertexts. For each ciphertext, the host then uses the stored nonce N to compute $\mathsf{PRF}_\Delta(N)$ and updates the ciphertext to $C' = C + \mathsf{PRF}_\Delta(N) = m + \mathsf{PRF}_b(N)$. However, the presence of the static nonce prevents the solution to be secure in our tokenization context. The tokenized data should be *unlinkable* across epochs for any adversary not knowing the update tweaks, and we even guarantee unlinkability in a forward-secure manner, i.e., a security breach at epoch e does not affect any data exposed before that time.

In the full version of their paper [5], Boneh et al. present a different solution for updatable encryption that achieves such unlinkability, but which suffers from similar efficiency issues as mentioned above: the data owner must retrieve and partially decrypt all of his ciphertexts, and then produce a dedicated update tweak for each ciphertext, which renders the solution unpractical for our purpose. Further, no formal security definition that models adaptive key corruptions for such updatable encryption is given in the paper.

The Pythia service proposed by Everspaugh et al. [8] mentions PRFs with key rotation which is closer to our goal, as it allows efficient updates of the outsourced PRF values whenever the key gets refreshed. The core idea of the Pythia scheme is very similar to our second, discrete-logarithm based construction. Unfortunately, the paper does not give any formal security definition that covers the possibility to update PRF values nor describes the exact properties of such a key-rotating PRF. As the main goal of Pythia is an *oblivious* and *verifiable* PRF service for password hashing, the overall construction is also more complex and aims at properties that are not needed here, and vice-versa, our unlinkability property does not seem necessary for the goal of Pythia.

While the aforementioned works share some relation with updatable tokenization, they have conceptually quite different security requirements. Starting with such an existing concept and extending its security notions and constructions to additionally satisfy the requirements of updatable tokenization, would reduce efficiency and practicality, for no clear advantage. Thus, we consider the approach of directly targeting the concrete real-world problem more suitable.

An initial study of security notions for tokenization was recently presented by Diaz-Santiago et al. [6]; they formally define tokenization systems and give several security notions and provably secure constructions. In a nutshell, their definitions closely resemble the conventional definitions for deterministic encryption and one-way functions adopted to the tokenization notation. However, they do not consider adaptive corruptions and neither address updatable tokens, which are the crucial aspects of this work.

2 Preliminaries

In this section, we recall the definitions of the building blocks and security notions needed in our constructions.

Deterministic Symmetric Encryption. A deterministic symmetric encryption scheme SE consists of a key space $\mathcal{K}$ and three polynomial-time algorithms $\mathsf{SE.KeyGen}, \mathsf{SE.Enc}, \mathsf{SE.Dec}$ satisfying the following conditions:

SE.KeyGen: The probabilistic key generation algorithm $\mathsf{SE.KeyGen}$ takes as input a security parameter λ and produces an encryption key $s \xleftarrow{r} \mathsf{SE.KeyGen}(\lambda)$.
SE.Enc: The deterministic encryption algorithm takes a key $s \in \mathcal{K}$ and a message $m \in \mathcal{M}$ and returns a ciphertext $C \leftarrow \mathsf{SE.Enc}(s, m)$.
SE.Dec: The deterministic decryption algorithm $\mathsf{SE.Dec}$ takes a key $s \in \mathcal{K}$ and a ciphertext C to return a message $m \leftarrow \mathsf{SE.Dec}(s, C)$.

For correctness we require that for any key $s \in \mathcal{K}$, any message $m \in \mathcal{M}$ and any ciphertext $C \leftarrow \mathsf{SE.Enc}(s, m)$, we have $m \leftarrow \mathsf{SE.Dec}(s, C)$.

We now define a security notion of deterministic symmetric encryption schemes in the sense of indistinguishability against chosen-plaintext attacks, or IND-CPA security. This notion was informally presented by Bellare et al. in [1], and captures the scenario where an adversary that is given access to a left-or-right (LoR) encryption oracle is not able to distinguish between the encryption of two distinct messages of its choice with probability non-negligibly better than one half. Since the encryption scheme in question is deterministic, the adversary can only query the LoR oracle with *distinct* messages on the same side (left or right) to avoid trivial wins. That is, queries of the type $(m_0^i, m_1^i), (m_0^j, m_1^j)$ where $m_0^i = m_0^j$ or $m_1^i = m_1^j$ are forbidden. We do not grant the adversary an explicit encryption oracle, as it can obtain encryptions of messages of its choice by querying the oracle with a pair of identical messages.

Definition 1. *A deterministic symmetric encryption scheme* $\mathsf{SE} = (\mathsf{SE.KeyGen}, \mathsf{SE.Enc}, \mathsf{SE.Dec})$ *is called* IND-CPA secure *if for all polynomial-time adversaries* $\mathcal{A}$*, it holds that* $|\Pr[\mathsf{Exp}^{\mathsf{ind\text{-}cpa}}_{\mathcal{A},\mathsf{SE}}(\lambda) = 1] - 1/2| \leq \epsilon(\lambda)$ *for some negligible function* ϵ*.*

Experiment $\mathsf{Exp}^{\mathsf{ind\text{-}cpa}}_{\mathcal{A},\mathsf{SE}}(\lambda)$*:*
$s \xleftarrow{r} \mathsf{SE.KeyGen}(\lambda)$
$d \xleftarrow{r} \{0,1\}$
$d' \xleftarrow{r} \mathcal{A}^{\mathcal{O}_{\mathsf{enc}}(s,d,\cdot,\cdot)}(\lambda)$
where $\mathcal{O}_{\mathsf{enc}}$ on input two messages m_0, m_1 returns $C \leftarrow \mathsf{SE.Enc}(s, m_d)$.
return 1 if $d' = d$ and all values $m_0^1, \ldots, m_0^q$ and all values $m_1^1, \ldots, m_1^q$ are distinct, respectively, where q denotes the number of queries to $\mathcal{O}_{\mathsf{enc}}$.

Hash Functions. A hash function $\mathsf{H} : \mathcal{D} \to \mathcal{R}$ is a deterministic function that maps inputs from domain $\mathcal{D}$ to values in range $\mathcal{R}$. For our second and stronger construction we assume the hash function to behave like a random oracle.

In our first construction we use a *keyed* hash function, i.e., H gets a key $hk \xleftarrow{r} \mathsf{H.KeyGen}(\lambda)$ as additional input. We require the keyed hash function to be *pseudorandom* and *weakly collision-resistant* for any adversary not knowing the key hk. We also need H to be *one-way* when the adversary is privy of the key, i.e., H should remain hard to invert on random inputs.

Pseudorandomness: A hash function is called pseudorandom if no efficient adversary $\mathcal{A}$ can distinguish H from a uniformly random function $f : \mathcal{D} \to \mathcal{R}$ with non-negligible advantage. That is, $\left|\Pr[\mathcal{A}^{\mathsf{H}(hk,\cdot)}(\lambda)] - \Pr[\mathcal{A}^{f(\cdot)}(\lambda)]\right|$ is negligible in λ, where the probability in the first case is over $\mathcal{A}$'s coin tosses and the choice of $hk \xleftarrow{r} \mathsf{H.KeyGen}(\lambda)$, and in the second case over $\mathcal{A}$'s coin tosses and the choice of the random function f.

Weak collision resistance: A hash function H is called weakly collision-resistant if for any efficient algorithm $\mathcal{A}$ the probability that for $hk \xleftarrow{r} \mathsf{H.KeyGen}(\lambda)$ and $(m, m') \xleftarrow{r} \mathcal{A}^{\mathsf{H}(hk,\cdot)}(\lambda)$ the adversary returns $m \neq m'$, where $\mathsf{H}(hk, m) = \mathsf{H}(hk, m')$, is negligible (as a function of λ).

One-wayness: A hash function H is one-way if for any efficient algorithm $\mathcal{A}$ the probability that for $hk \xleftarrow{r} \mathsf{H.KeyGen}(\lambda)$, $m \xleftarrow{r} \mathcal{D}$ and $m' \xleftarrow{r} \mathcal{A}(hk, \mathsf{H}(hk, m))$ returns m', where $\mathsf{H}(hk, m) = \mathsf{H}(hk, m')$, is negligible (as a function of λ).

Decisional Diffie-Hellman Assumption. Our second construction requires a group $(\mathbb{G}, g, p)$ as input where $\mathbb{G}$ denotes a cyclic group $\mathbb{G} = \langle g \rangle$ of order p in which the Decisional Diffie-Hellman (DDH) problem is hard w.r.t. λ, i.e., p is a λ-bit prime. More precisely, a group $(\mathbb{G}, g, p)$ satisfies the DDH assumption if for any efficient adversary $\mathcal{A}$ the probability $\left|\Pr[\mathcal{A}(\mathbb{G}, p, g, g^a, g^b, g^{ab})] - \Pr[\mathcal{A}(\mathbb{G}, p, g, g^a, g^b, g^c)]\right|$ is negligible in λ, where the probability is over the random choice of p, g, the random choices of $a, b, c \in \mathbb{Z}_p$, and $\mathcal{A}$'s coin tosses.

3 Formalizing Updatable Tokenization

An updatable tokenization scheme contains algorithms for a data *owner* and a *host*. The owner de-sensitizes data through tokenization operations and dynamically outsources the tokenized data to the host. For this purpose, the data owner first runs an algorithm setup to create a tokenization key. The tokenization key evolves with *epochs*, and the data is tokenized with respect to a specific epoch e, starting with $e = 0$. For a given epoch, algorithm token takes a data value and tokenizes it with the current key k_e. When moving from epoch e to epoch $e + 1$, the owner invokes an algorithm next to generate the key material k_{e+1} for the new epoch and an update tweak Δ_{e+1}. The owner then sends Δ_{e+1} to the host, deletes k_e and Δ_{e+1} immediately, and uses k_{e+1} for tokenization from now on. After receiving Δ_{e+1}, the host first deletes Δ_e and then uses an algorithm upd to update all previously received tokenized values from epoch e to $e+1$, using Δ_{e+1}. Hence, during some epoch e the update tweak from $e-1$ to e is available at the host, but update tweaks from earlier epochs have been deleted.

Definition 2. *An updatable tokenization scheme* UTO *consists of a data space* $\mathcal{X}$*, a token space* $\mathcal{Y}$*, and a set of polynomial-time algorithms* UTO.setup*,* UTO.next*,* UTO.token*, and* UTO.upd *satisfying the following conditions:*

UTO.setup**:** The algorithm UTO.setup is a probabilistic algorithm run by the owner. On input a security parameter λ, this algorithm returns the tokenization key for the first epoch $k_0 \xleftarrow{r} \mathsf{UTO.setup}(\lambda)$.

UTO.next**:** This probabilistic algorithm is also run by the owner. On input a tokenization key k_e for some epoch e, it outputs a tokenization key k_{e+1} and an update tweak Δ_{e+1} for epoch $e+1$. That is, $(k_{e+1}, \Delta_{e+1}) \xleftarrow{r} \mathsf{UTO.next}(k_e)$.

UTO.token**:** This is a deterministic *injective* algorithm run by the owner. Given the secret key k_e and some input data $x \in \mathcal{X}$, the algorithm outputs a tokenized value $y_e \in \mathcal{Y}$. That is, $y_e \leftarrow \mathsf{UTO.token}(k_e, x)$.

UTO.upd**:** This deterministic algorithm is run by the host and uses the update tweak. On input the update tweak Δ_{e+1} and some tokenized value y_e, UTO.upd updates y_e to y_{e+1}, that is, $y_{e+1} \leftarrow \mathsf{UTO.upd}(\Delta_{e+1}, y_e)$.

The *correctness* condition of a UTO scheme ensures referential integrity inside the tokenized data set. A newly tokenized value from the owner in a particular epoch must be the same as the tokenized value produced by the host using update operations. More precisely, we require that for any $x \in \mathcal{X}$, for any $k_0 \xleftarrow{r} \mathsf{UTO.setup}(\lambda)$, for any sequence of tokenization key/update tweak pairs $(k_1, \Delta_1), \ldots, (k_e, \Delta_e)$ generated as $(k_{j+1}, \Delta_{j+1}) \xleftarrow{r} \mathsf{UTO.next}(k_j)$ for $j = 0, \ldots, e-1$ through repeated applications of the key-evolution algorithm, and for any $y_e \leftarrow \mathsf{UTO.token}(k_e, x)$, it holds that

$$\mathsf{UTO.token}(k_{e+1}, x) = \mathsf{UTO.upd}(\Delta_{e+1}, y_e).$$

3.1 Privacy of Updatable Tokenization Schemes

The main goal of UTO is to achieve *privacy* for data values, ensuring that an adversary cannot gain information about the tokenized values and cannot link them to input data tokenized in past epochs. We introduce three indistinguishability-based notions for the privacy of tokenized values, and one notion ruling out that an adversary may reverse the tokenization and recover the input value from a tokenized one. All security notions are defined through an experiment run between a challenger and an adversary $\mathcal{A}$. Depending on the notion, the adversary may issue queries to different oracles, defined in the next section.

At a high level, the four security notions for UTO are distinguished by the corruption capabilities of $\mathcal{A}$.

IND-HOCH: Indistinguishability with Honest Owner and Corrupted Host: This is the most basic security criterion, focusing on the updatable dynamic aspect of UTO. It considers the owner to be honest and permits corruption of the host during the interaction. The adversary gains access to the update tweaks for all epochs following the compromise and yet, it should (roughly speaking) not be able to distinguish values tokenized before the corruption.

IND-COHH: Indistinguishability with Corrupted Owner and Honest Host: Modeling a corruption of the owner at some point in time, the adversary learns the tokenization key of the compromised epoch and all secrets of the owner. Subsequently $\mathcal{A}$ may take control of the owner, but should not learn the correspondence between values tokenized before the corruption. The host is assumed to remain (mostly) honest.

IND-COTH: Indistinguishability with Corrupted Owner and Transiently Corrupted Host: As a refinement of the first two notions, $\mathcal{A}$ can transiently corrupt the host during multiple epochs according to its choice, and it may also permanently corrupt the owner. The adversary learns the update tweaks of the specific epochs where it corrupts the host, and learns the tokenization key of the epoch where it corrupts the owner. Data values tokenized prior to exposing the owner's secrets should remain unlinkable.

One-Wayness: This notion models the scenario where the owner is corrupted right at the first epoch and the adversary therefore learns all secrets. Yet, the tokenization operation should be one-way in the sense that observing a tokenized value does not give the adversary an advantage for guessing the corresponding input from $\mathcal{X}$.

3.2 Definition of Oracles

During the interaction with the challenger in the security definitions, the adversary may access oracles for *data tokenization*, for moving to the *next epoch*, for *corrupting the host*, and for *corrupting the owner*. In the following description, the oracles may access the state of the challenger during the experiment. The challenger initializes a UTO scheme with global state (k_0, Δ_0, e), where $k_0 \leftarrow \mathsf{UTO.setup}(\lambda)$, $\Delta_0 \leftarrow \perp$, and $e \leftarrow 0$. Two auxiliary variables e_h^* and e_o^* record the epochs where the host and the owner were first corrupted, respectively. Initially $e_h^* \leftarrow \perp$ and $e_o^* \leftarrow \perp$.

$\mathcal{O}_{\mathsf{token}}(x)$**:** On input a value $x \in \mathcal{X}$, return $y_e \leftarrow \mathsf{UTO.token}(k_e, x)$ to the adversary, where k_e is the tokenization key of the current epoch.

$\mathcal{O}_{\mathsf{next}}$**:** When triggered, compute the tokenization key and update tweak of the next epoch as $(k_{e+1}, \Delta_{e+1}) \leftarrow \mathsf{UTO.next}(k_e)$ and update the global state to $(k_{e+1}, \Delta_{e+1}, e+1)$.

$\mathcal{O}_{\mathsf{corrupt\text{-}h}}$**:** When invoked, return Δ_e to the adversary. If called for the first time ($e_h^* = \perp$), then set $e_h^* \leftarrow e$. This oracle models the corruption of the host and may be called multiple times.

$\mathcal{O}_{\mathsf{corrupt\text{-}o}}$**:** When invoked for the first time ($e_o^* = \perp$), then set $e_o^* \leftarrow e$ and return k_e to the adversary. This oracle models the corruption of the owner and can only be called once. After this call, the adversary no longer has access to $\mathcal{O}_{\mathsf{token}}$ and $\mathcal{O}_{\mathsf{next}}$.

Note that although corruption of the host at epoch e exposes the update tweak Δ_e, the adversary should not be able to compute update tweaks of future epochs from this value. To obtain those, $\mathcal{A}$ should call $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$ again in

the corresponding epochs; this is used for IND-HOCH security and IND-COTH security, with different side-conditions. A different case arises when the owner is corrupted, since this exposes all *relevant* secrets of the challenger. From that point the adversary can generate tokenization keys and update tweaks for all subsequent epochs on its own. This justifies why the oracle $\mathcal{O}_{\mathsf{corrupt\text{-}o}}$ can only be called once. For the same reason, it makes no sense for an adversary to query the $\mathcal{O}_{\mathsf{token}}$ and $\mathcal{O}_{\mathsf{next}}$ oracles after the corruption of the owner. Furthermore, observe that $\mathcal{O}_{\mathsf{corrupt\text{-}o}}$ does not return Δ_e according to the assumption that the owner deletes this atomically with executing the next algorithm.

We are now ready to formally define the security notions for UTO in the remainder of this section.

3.3 IND-HOCH: Honest Owner and Corrupted Host

The IND-HOCH notion ensures that tokenized data does not reveal information about the corresponding original data when $\mathcal{A}$ compromises the host and obtains the update tweaks of the current and all future epochs. Tokenized values are also unlinkable across epochs, as long as the adversary does not know at least one update tweak in that timeline.

Definition 3 (IND-HOCH). *An updatable tokenization scheme* UTO *is said to be* IND-HOCH secure *if for all polynomial-time adversaries* $\mathcal{A}$ *it holds that* $|\Pr[\mathsf{Exp}^{\mathsf{IND\text{-}HOCH}}_{\mathcal{A},\mathsf{UTO}}(\lambda) = 1] - 1/2| \leq \epsilon(\lambda)$ *for some negligible function* ϵ.

Experiment $\mathsf{Exp}^{\mathsf{IND\text{-}HOCH}}_{\mathcal{A},\mathsf{UTO}}(\lambda)$*:*
$k_0 \xleftarrow{r} \mathsf{UTO.setup}(\lambda)$
$e \leftarrow 0$; $e^*_h \leftarrow \perp$ *// these variables are updated by the oracles*
$(\tilde{x}_0, \tilde{x}_1, state) \xleftarrow{r} \mathcal{A}^{\mathcal{O}_{\mathsf{token}}, \mathcal{O}_{\mathsf{next}}, \mathcal{O}_{\mathsf{corrupt\text{-}h}}}(\lambda)$
$\tilde{e} \leftarrow e$; $d \xleftarrow{r} \{0,1\}$
$\tilde{y}_{d,\tilde{e}} \leftarrow \mathsf{UTO.token}(k_{\tilde{e}}, \tilde{x}_d)$
$d' \xleftarrow{r} \mathcal{A}^{\mathcal{O}_{\mathsf{token}}, \mathcal{O}_{\mathsf{next}}, \mathcal{O}_{\mathsf{corrupt\text{-}h}}}(\tilde{y}_{d,\tilde{e}}, state)$
return 1 if $d' = d$ and at least *one* of following conditions holds
a) $\big(e^*_h \leq \tilde{e} + 1\big) \wedge$ $\mathcal{A}$ has not queried $\mathcal{O}_{\mathsf{token}}(\tilde{x}_0)$ or $\mathcal{O}_{\mathsf{token}}(\tilde{x}_1)$ in epoch $e^*_h - 1$ or later
b) $\big(e^*_h > \tilde{e} + 1 \vee e^*_h = \perp\big) \wedge$ $\mathcal{A}$ has not queried $\mathcal{O}_{\mathsf{token}}(\tilde{x}_0)$ or $\mathcal{O}_{\mathsf{token}}(\tilde{x}_1)$ in epoch $\tilde{e}$

This experiment has two phases. In the first phase, $\mathcal{A}$ may query $\mathcal{O}_{\mathsf{token}}$, $\mathcal{O}_{\mathsf{next}}$ and $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$; it ends at an epoch $\tilde{e}$ when $\mathcal{A}$ outputs two challenge inputs $\tilde{x}_0$ and $\tilde{x}_1$. The challenger picks one at random (denoted by $\tilde{x}_d$), tokenizes it, obtains the challenge $\tilde{y}_{d,\tilde{e}}$ and starts the second phase by invoking $\mathcal{A}$ with $\tilde{y}_{d,\tilde{e}}$. The adversary may then further query $\mathcal{O}_{\mathsf{token}}$, $\mathcal{O}_{\mathsf{next}}$, and $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$ and eventually outputs its guess d' for which data value was tokenized. Note that only the first host corruption matters for our security notion, since we are assuming that once

corrupted, the host is always corrupted. For simplicity, we therefore assume that $\mathcal{A}$ calls $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$ once in every epoch after e_h^*.

The adversary wins the experiment if it correctly guesses d while respecting two conditions that differ depending on whether the adversary corrupted the host (roughly) before or after the challenge epoch:

a) If $e_h^* \leq \tilde{e} + 1$, then $\mathcal{A}$ first corrupts the host before, during, or immediately after the challenge epoch and may learn the update tweaks to epoch e_h^* and later ones. In this case, it must not query the tokenization oracle on the challenge inputs in epoch $e_h^* - 1$ or later.

   In particular, if this restriction was not satisfied, when $e_h^* \leq \tilde{e}$, the adversary could tokenize data of its choice, including $\tilde{x}_0$ and $\tilde{x}_1$, during any epoch from $e_h^* - 1$ to $\tilde{e}$, subsequently update the tokenized value to epoch $\tilde{e}$, and compare it to the challenge $\tilde{y}_{d,\tilde{e}}$. This would allow $\mathcal{A}$ to trivially win the security experiment.

   For the case $e_h^* = \tilde{e} + 1$, recall that according to the experiment, the update tweak Δ_e remains accessible until epoch $e+1$ starts. Therefore, $\mathcal{A}$ learns the update tweak from $\tilde{e}$ to $\tilde{e} + 1$ and may update $\tilde{y}_{d,\tilde{e}}$ into epoch $\tilde{e} + 1$. Hence, from this time on it must not query $\mathcal{O}_{\mathsf{token}}$ with the challenge inputs either.

b) If $e_h^* > \tilde{e} + 1 \vee e_h^* = \bot$, i.e., the host was first corrupted after epoch $\tilde{e} + 1$ or not at all, then the only restriction is that $\mathcal{A}$ must not query the tokenization oracle on the challenge inputs during epoch $\tilde{e}$. This is an obvious restriction to exclude trivial wins, as tokenization is deterministic.

   This condition is less restrictive than case a), but it suffices since the adversary cannot update tokenized values from earlier epochs to $\tilde{e}$, nor from $\tilde{e}$ to a later epoch. The reason is that $\mathcal{A}$ only gets the update tweaks from epoch $\tilde{e} + 2$ onwards.

3.4 IND-COHH: Corrupted Owner and Honest Host

The IND-COHH notion models a compromise of the owner in a certain epoch, such that the adversary learns the tokenization key and may generate tokenization keys and update tweaks of all subsequent epochs by itself. Given that the tokenization key allows to derive the update tweak of the host, this implicitly models some form of host corruption as well. The property ensures that data tokenized before the corruption remains hidden, that is, the adversary does not learn any information about the original data, nor can it link such data with data tokenized in other epochs.

Definition 4 (IND-COHH). *An updatable tokenization scheme* UTO *is said to be* IND-COHH secure *if for all polynomial-time adversaries* $\mathcal{A}$ *it holds that* $|\Pr[\mathsf{Exp}^{\mathsf{IND\text{-}COHH}}_{\mathcal{A},\mathsf{UTO}}(\lambda) = 1] - 1/2| \leq \epsilon(\lambda)$ *for some negligible function* ϵ.

Experiment $\mathsf{Exp}^{\mathsf{IND\text{-}COHH}}_{\mathcal{A},\mathsf{UTO}}(\lambda)$*:*
$k_0 \xleftarrow{r} \mathsf{UTO.setup}(\lambda)$
$e \leftarrow 0$*;* $e_o^* \leftarrow \perp$ *// these variables are updated by the oracles*
$(\tilde{x}_0, \tilde{x}_1, state) \xleftarrow{r} \mathcal{A}^{\mathcal{O}_{\mathsf{token}}, \mathcal{O}_{\mathsf{next}}}(\lambda)$
$\tilde{e} \leftarrow e$; $d \xleftarrow{r} \{0,1\}$
$\tilde{y}_{d,\tilde{e}} \leftarrow \mathsf{UTO.token}(k_{\tilde{e}}, \tilde{x}_d)$
$d' \xleftarrow{r} \mathcal{A}^{\mathcal{O}_{\mathsf{token}}, \mathcal{O}_{\mathsf{next}}, \mathcal{O}_{\mathsf{corrupt\text{-}o}}}(\tilde{y}_{d,\tilde{e}}, state)$
return 1 if $d' = d$ and all following conditions hold
a) $e_o^* > \tilde{e} \vee e_o^* = \perp$
b) $\mathcal{A}$ never queried $\mathcal{O}_{\mathsf{token}}(\tilde{x}_0)$ or $\mathcal{O}_{\mathsf{token}}(\tilde{x}_1)$ in epoch $\tilde{e}$

During the first phase of the IND-COHH experiment the adversary may query $\mathcal{O}_{\mathsf{token}}$ and $\mathcal{O}_{\mathsf{next}}$, but it may not corrupt the owner. At epoch $\tilde{e}$, the adversary produces two challenge inputs $\tilde{x}_0$ and $\tilde{x}_1$. Again, the challenger selects one at random and tokenizes it, resulting in the challenge $\tilde{y}_{d,\tilde{e}}$. Subsequently, $\mathcal{A}$ may further query $\mathcal{O}_{\mathsf{token}}$ and $\mathcal{O}_{\mathsf{next}}$, and now may also invoke $\mathcal{O}_{\mathsf{corrupt\text{-}o}}$. Once the owner is corrupted (during epoch e_o^*), $\mathcal{A}$ knows all key material of the owner and may generate tokenization keys and update tweaks of all subsequent epochs by itself. Thus, from this time on, we remove access to the $\mathcal{O}_{\mathsf{token}}$ or $\mathcal{O}_{\mathsf{next}}$ oracles for simplicity.

The adversary ends the experiment by guessing which input challenge was tokenized. It wins when the guess is correct and the following conditions are met:

a) $\mathcal{A}$ must have corrupted the owner only after the challenge epoch ($e_o^* > \tilde{e}$) or not at all ($e_o^* = \perp$). This is necessary since corruption during epoch $\tilde{e}$ would leak the tokenization key $k_{\tilde{e}}$ to the adversary. (Note that corruption before $\tilde{e}$ is ruled out syntactically.)
b) $\mathcal{A}$ must neither query the tokenization oracle with any challenge input ($\tilde{x}_0$ or $\tilde{x}_1$) during the challenge epoch $\tilde{e}$. This condition eliminates that $\mathcal{A}$ can trivially reveal the challenge input since the tokenization operation is deterministic.

On the (Im)possibility of Additional Host Corruption. As can be noted, the IND-COHH experiment does not consider the corruption of the host at all. The reason is that allowing host corruption in addition to owner corruption would either result in a non-achievable notion, or it would give the adversary no extra advantage. To see this, we first argue why additional host corruption capabilities at any epoch $e_h^* \leq \tilde{e} + 1$ is not allowed. Recall that such a corruption is possible in the IND-HOCH experiment if the adversary does not make any tokenization queries on the challenge values $\tilde{x}_0$ or $\tilde{x}_1$ at any epoch $e \geq e_h^* - 1$. This restriction is necessary in the IND-HOCH experiment to prevent the adversary from trivially linking the tokenized values of $\tilde{x}_0$ or $\tilde{x}_1$ to the challenge $\tilde{y}_{d,\tilde{e}}$. However, when the owner can also be corrupted, at epoch $e_o^* > \tilde{e}$, that restriction is useless. Note that upon calling $\mathcal{O}_{\mathsf{corrupt\text{-}o}}$ the adversary learns the owner's tokenization key and can simply tokenize $\tilde{x}_0$ and $\tilde{x}_1$ at epoch e_o^*. The results can be compared with an updated version of $\tilde{y}_{d,\tilde{e}}$ to trivially win the security experiment.

Now we discuss the additional corruption of the host at any epoch $e_h^* > \tilde{e}+1$. We note that corruption of the owner at epoch $e_o^* > \tilde{e}$ allows the adversary to obtain the tokenization key of epoch e_o^* and compute the tokenization keys and update tweaks of all epochs $e > e_o^*+1$. Thus, the adversary then trivially knows all tokenization keys from e_o^*+1 onward and modeling corruption of the host after the owner is not necessary. The only case left is to consider host corruption before owner corruption, at an epoch e_h^* with $\tilde{e}+1 < e_h^* < e_o^*$. However, corrupting the host first would not have any impact on the winning condition. Hence, without loss of generality, we assume that the adversary always corrupts the owner first, which allows us to fully omit the $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$ oracle in our IND-COHH experiment.

We stress that the impossibility of host corruption at any epoch $e_h^* \leq \tilde{e}+1$ only holds if we consider *permanent* corruptions, i.e., the adversary, upon invocation of $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$ is assumed to fully control the host and to learn all future update tweaks. In the following security notion, IND-COTH, we bypass this impossibility by modeling *transient* corruption of the host.

3.5 IND-COTH: Corrupted Owner and Transiently Corrupted Host

Extending both of the above security properties, the IND-COTH notion considers corruption of the owner and repeated but transient corruptions of the host. It addresses situations where some of the update tweaks received by the host leak to $\mathcal{A}$ and the keys of the owner are also exposed at a later stage.

Definition 5 (IND-COTH). *An updatable tokenization scheme* UTO *is said to be* IND-COTH secure *if for all polynomial-time adversaries $\mathcal{A}$ it holds that* $|\Pr[\mathsf{Exp}^{\mathsf{IND\text{-}COTH}}_{\mathcal{A},\mathsf{UTO}}(\lambda)=1] - 1/2| \leq \epsilon(\lambda)$ *for some negligible function ϵ.*

Experiment $\mathsf{Exp}^{\mathsf{IND\text{-}COTH}}_{\mathcal{A},\mathsf{UTO}}(\lambda)$*:*
$k_0 \xleftarrow{r} \mathsf{UTO.setup}(\lambda)$
$e \leftarrow 0$; $e_o^* \leftarrow \bot$ *// these variables are updated by the oracles*
$e_{\mathsf{last}} \leftarrow \bot$; $e_{\mathsf{first}} \leftarrow \bot$
$(\tilde{x}_0, \tilde{x}_1, state) \xleftarrow{r} \mathcal{A}^{\mathcal{O}_{\mathsf{token}}, \mathcal{O}_{\mathsf{next}}, \mathcal{O}_{\mathsf{corrupt\text{-}h}}}(\lambda)$
$\tilde{e} \leftarrow e$; $d \xleftarrow{r} \{0,1\}$
$\tilde{y}_{d,\tilde{e}} \leftarrow \mathsf{UTO.token}(k_{\tilde{e}}, \tilde{x}_d)$
$d' \xleftarrow{r} \mathcal{A}^{\mathcal{O}_{\mathsf{token}}, \mathcal{O}_{\mathsf{next}}, \mathcal{O}_{\mathsf{corrupt\text{-}h}}, \mathcal{O}_{\mathsf{corrupt\text{-}o}}}(\tilde{y}_{d,\tilde{e}}, state)$
$e_{\mathsf{last}} \leftarrow$ last epoch before $\tilde{e}$ in which $\mathcal{A}$ queried $\mathcal{O}_{\mathsf{token}}(\tilde{x}_0)$ or $\mathcal{O}_{\mathsf{token}}(\tilde{x}_1)$
$e_{\mathsf{first}} \leftarrow$ first epoch after $\tilde{e}$ in which $\mathcal{A}$ queried $\mathcal{O}_{\mathsf{token}}(\tilde{x}_0)$ or $\mathcal{O}_{\mathsf{token}}(\tilde{x}_1)$
return 1 if $d' = d$ and all following conditions hold
a) $e_o^* > \tilde{e} \vee e_o^* = \bot$
b) $\mathcal{A}$ never queried $\mathcal{O}_{\mathsf{token}}(\tilde{x}_0)$ or $\mathcal{O}_{\mathsf{token}}(\tilde{x}_1)$ in epoch $\tilde{e}$
c) either $e_h^* = \bot$ or *all* following conditions hold
i) $\big(e_{\mathsf{last}} = \bot\big) \vee \exists\ e'$ with $e_{\mathsf{last}} < e' \leq \tilde{e}$ where $\mathcal{A}$ has not queried $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$
ii) $\big(e_{\mathsf{first}} = \bot\big) \vee \exists\ e''$ with $\tilde{e} < e'' \leq e_{\mathsf{first}}$ where $\mathcal{A}$ has not queried $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$
iii) $\big(e_o^* = \bot\big) \vee \exists\ e'''$ with $\tilde{e} < e''' \leq e_o^*$ where $\mathcal{A}$ has not queried $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$

Observe that the owner can only be corrupted after the challenge epoch, just as in the IND-COHH experiment. As before, $\mathcal{A}$ then obtains all key material and, for simplicity, we remove access to the $\mathcal{O}_{\mathsf{token}}$ or $\mathcal{O}_{\mathsf{next}}$ oracles from this time on. The transient nature of the host corruption allows to grant $\mathcal{A}$ additional access to $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$ *before* the challenge, which would be impossible in the IND-COHH experiment if permanent host corruption was considered.

Compared to the IND-HOCH definition, here $\mathcal{A}$ may corrupt the host *and* ask for a challenge input to be tokenized after the corruption. Multiple host corruptions may occur before, during, and after the challenge epoch. But in order to win the experiment, $\mathcal{A}$ must leave out at least one epoch and miss an update tweak. Otherwise it could trivially guess the challenge by updating the challenge output or a challenge input tokenized in another epoch to the same stage. In the experiment this is captured through the conditions under c). In particular:

c-i) If $\mathcal{A}$ calls $\mathcal{O}_{\mathsf{token}}$ with one of the challenge inputs $\tilde{x}_0$ or $\tilde{x}_1$ *before* triggering the challenge, it must not corrupt the host and miss the update tweak in at least one epoch from this point up to the challenge epoch. Thus, the *latest* epoch before the challenge epoch where $\mathcal{A}$ queries $\mathcal{O}_{\mathsf{token}}(\tilde{x}_0)$ or $\mathcal{O}_{\mathsf{token}}(\tilde{x}_1)$, denoted e_{last}, must be smaller than the last epoch before $\tilde{e}$ where the host is not corrupted.

c-ii) Likewise if $\mathcal{A}$ queries $\mathcal{O}_{\mathsf{token}}$ with a challenge input $\tilde{x}_0$ or $\tilde{x}_1$ *after* the challenge epoch, then it must not corrupt the host and miss the update tweak in at least one epoch after $\tilde{e}$. Otherwise, it could update the challenge $\tilde{y}_{d,\tilde{e}}$ to the epoch where it calls $\mathcal{O}_{\mathsf{token}}$. The *first* epoch after the challenge epoch where $\mathcal{A}$ queries $\mathcal{O}_{\mathsf{token}}(\tilde{x}_0)$ or $\mathcal{O}_{\mathsf{token}}(\tilde{x}_1)$, denoted e_{first}, must be larger than or equal to the first epoch after $\tilde{e}$ where the host is not corrupted.

c-iii) If $\mathcal{A}$ calls $\mathcal{O}_{\mathsf{corrupt\text{-}o}}$, it must not obtain at least one update tweak after the challenge epoch and before, or during, the epoch of owner corruption e_o^*. Otherwise, $\mathcal{A}$ could tokenize $\tilde{x}_0$ and $\tilde{x}_1$ with the tokenization key of epoch e_o^*, exploit the exposed update tweaks to evolve the challenge value $\tilde{y}_{d,\tilde{e}}$ to that epoch, and compare the results.

PRF-style vs. IND-CPA-style definitions. We have opted for definitions based on indistinguishability in our model. Given that the goal of tokenization is to output random looking tokens, a security notion in the spirit of pseudorandomness might seem like a more natural choice at first glance. However, a definition in the PRF-style does not cope well with *adaptive* attacks: in our security experiments the adversary is allowed to adaptively corrupt the data host and corrupt the data owner, upon which it gets the update tweaks or the secret tokenization key. Modeling this in a PRF vs. random function experiment would require the random function to contain a key and to be compatible with an update function that can be run by the adversary. Extending the random function with these "features" would lead to a PRF vs. PRF definition. The IND-CPA inspired approach used in this paper allows to cover the adaptive attacks and consistency features in a more natural way.

Relation Among the Security Notions. Our notion of IND-COTH security is the strongest of the three indistinguishability notions above, as it implies both IND-COHH and IND-HOCH security, but not vice-versa. That is, IND-COTH security is not implied by IND-COHH and IND-HOCH security. A distinguishing example is our $\mathsf{UTO}_{\mathsf{SE}}$ scheme. As we will see in Section 4.1, $\mathsf{UTO}_{\mathsf{SE}}$ is both IND-COHH and IND-HOCH secure, but not IND-COTH secure.

The proof of Theorem 1 below is given in Appendix A.

Theorem 1 (IND-COTH ⇒ IND-COHH + IND-HOCH). *If an updatable tokenization scheme* UTO *is* IND-COTH secure*, then it is also* IND-COHH secure *and* IND-HOCH secure*.*

3.6 One-Wayness

The one-wayness notion models the fact that a tokenization scheme should not be reversible even if an adversary is given the tokenization keys. In other words, an adversary who sees tokenized values and gets hold of the tokenization keys cannot obtain the original data. Because the keys allow one to reproduce the tokenization operation and to test whether the output matches a tokenized value, the resulting security level depends on the size of the input space and the adversary's uncertainty about the input. Thus, in practice, the level of security depends on the prior knowledge of the adversary about $\mathcal{X}$.

Our definition is similar to the standard notion of one-wayness, with the difference that we ask the adversary to output the exact preimage of a tokenized challenge value, as our tokenization algorithm is an injective function.

Definition 6 (One-Wayness). *An updatable tokenization scheme* UTO *is said to be* one-way *if for all polynomial-time adversaries* $\mathcal{A}$ *it holds that*

$$\Pr[\ x = \tilde{x} : x \leftarrow \mathcal{A}(\lambda, k_0, \tilde{y}),\ \tilde{y} \leftarrow \mathsf{UTO.token}(k_0, \tilde{x}), \tilde{x} \xleftarrow{r} \mathcal{X}, k_0 \xleftarrow{r} \mathsf{UTO.setup}(\lambda)] \leq 1/|\mathcal{X}|.$$

4 UTO Constructions

In this section we present two efficient constructions of updatable tokenization schemes. The first solution ($\mathsf{UTO}_{\mathsf{SE}}$) is based on symmetric encryption and achieves one-wayness, IND-HOCH and IND-COHH security; the second construction ($\mathsf{UTO}_{\mathsf{DL}}$) relies on a discrete-log assumption, and additionally satisfies IND-COTH security. Both constructions share the same core idea: First, the input value is hashed, and then the hash is encrypted under a key that changes every epoch.

4.1 An UTO Scheme based on Symmetric Encryption

We build a first updatable tokenization scheme $\mathsf{UTO}_{\mathsf{SE}}$, that is based on a symmetric deterministic encryption scheme $\mathsf{SE} = (\mathsf{SE.KeyGen}, \mathsf{SE.Enc}, \mathsf{SE.Dec})$ with

message space $\mathcal{M}$ and a keyed hash function $\mathsf{H} : \mathcal{K} \times \mathcal{X} \to \mathcal{M}$. In order to tokenize an input $x \in \mathcal{X}$, our scheme simply encrypts the hashed value of x. At each epoch e, a distinct random symmetric key s_e is used for encryption, while a fixed random hash key hk is used to hash x. Both keys are chosen by the data owner. To update the tokens, the host receives the encryption keys of the previous and current epoch and re-encrypts all hashed values to update them into the current epoch. More precisely, our $\mathsf{UTO}_{\mathsf{SE}}$ scheme is defined as follows:

$\mathsf{UTO.setup}(\lambda)$**:** Generate keys $s_0 \xleftarrow{r} \mathsf{SE.KeyGen}(\lambda)$, $hk \xleftarrow{r} \mathsf{H.KeyGen}(\lambda)$ and output $k_0 \leftarrow (s_0, hk)$.

$\mathsf{UTO.next}(k_e)$**:** Parse k_e as (s_e, hk). Choose a new key $s_{e+1} \xleftarrow{r} \mathsf{SE.KeyGen}(\lambda)$ and set $k_{e+1} \leftarrow (s_{e+1}, hk)$ and $\Delta_{e+1} \leftarrow (s_e, s_{e+1})$. Output (k_{e+1}, Δ_{e+1}).

$\mathsf{UTO.token}(k_e, x)$**:** Parse k_e as (s_e, hk) and output $y_e \leftarrow \mathsf{SE.Enc}(s_e, \mathsf{H}(hk, x))$.

$\mathsf{UTO.upd}(\Delta_{e+1}, y_e)$**:** Parse Δ_{e+1} as (s_e, s_{e+1}) and output the updated value $y_{e+1} \leftarrow \mathsf{SE.Enc}(s_{e+1}, \mathsf{SE.Dec}(s_e, y_e))$.

This construction achieves IND-HOCH, IND-COHH, and one-wayness but not the stronger IND-COTH notion. The issue is that a transiently corrupted host can recover the static hash during the update procedure and thus can link tokenized values from different epochs, even without knowing all the update tweaks between them.

Theorem 2. *The* $\mathsf{UTO}_{\mathsf{SE}}$ *as defined above satisifies the IND-HOCH, IND-COHH and one-wayness properties based on the following assumptions on the underlying encryption scheme* SE *and hash function* H*:*

$\mathsf{UTO}_{\mathsf{SE}}$	SE	H
IND-COHH	IND-CPA	weak collision resistance
IND-HOCH	IND-CPA	pseudorandomness
one-wayness	-	one-wayness

The proof of Theorem 2 is given in Appendix B.

4.2 An UTO Scheme based on Discrete Logarithms

Our second construction $\mathsf{UTO}_{\mathsf{DL}}$ overcomes the limitation of the first scheme by performing the update in a proxy re-encryption manner using the re-encryption idea first proposed by Blaze et al. [3]. That is, the hashed value is raised to an exponent that the owner randomly chooses at every new epoch. To update tokens, the host is not given the keys itself but only the quotient of the current and previous exponent. While this allows the host to consistently update his data, it does not reveal the inner hash anymore and guarantees unlinkability across epochs, thus satisfying also our strongest notion of IND-COTH security.

More precisely, the scheme makes use of a cyclic group $(\mathbb{G}, g, p)$ and a hash function $\mathsf{H} : \mathcal{X} \to \mathbb{G}$. We assume the hash function and the group description to be publicly available. The algorithms of our $\mathsf{UTO}_{\mathsf{DL}}$ scheme are defined as follows:

$\mathsf{UTO.setup}(\lambda)$**:** Choose $k_0 \xleftarrow{r} \mathbb{Z}_p$ and output k_0.
$\mathsf{UTO.next}(k_e)$**:** Choose $k_{e+1} \xleftarrow{r} \mathbb{Z}_p$, set $\Delta_{e+1} \leftarrow k_{e+1}/k_e$, and output (k_{e+1}, Δ_{e+1}).
$\mathsf{UTO.token}(k_e, x)$**:** Compute $y_e \leftarrow \mathsf{H}(x)^{k_e}$, and output y_e.
$\mathsf{UTO.upd}(\Delta_{e+1}, y_e)$**:** Compute $y_{e+1} \leftarrow y_e^{\Delta_{e+1}}$, and output y_{e+1}.

Our $\mathsf{UTO}_{\mathsf{DL}}$ scheme is one-way and satisfies our strongest notion of IND-COTH security, from which IND-HOCH and IND-COHH security follows (see Theorem 1). The proof of Theorem 3 below is given in Appendix C.

Theorem 3. *The* $\mathsf{UTO}_{\mathsf{DL}}$ *scheme as defined above is IND-COTH secure under the DDH assumption in the random oracle model, and one-way if* H *is one-way.*

Acknowledgements. We would like to thank our colleagues Michael Osborne, Tamas Visegrady and Axel Tanner for helpful discussions on tokenization.

This work has been supported in part by the European Commission through the Horizon 2020 Framework Programme (H2020-ICT-2014-1) under grant agreement number 644371 WITDOM and through the Seventh Framework Programme under grant agreement number 321310 PERCY, and in part by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 15.0098.

References

1. Bellare, M., Boldyreva, A., O'Neill, A.: Deterministic and efficiently searchable encryption. In: Menezes, A. (ed.) Advances in Cryptology - CRYPTO 2007, 27th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2007, Proceedings. Lecture Notes in Computer Science, vol. 4622, pp. 535–552. Springer (2007), http://dx.doi.org/10.1007/978-3-540-74143-5_30
2. Bellare, M., Ristenpart, T., Rogaway, P., Stegers, T.: Format-preserving encryption. In: Jr., M.J.J., Rijmen, V., Safavi-Naini, R. (eds.) Selected Areas in Cryptography, 16th Annual International Workshop, SAC 2009, Calgary, Alberta, Canada, August 13-14, 2009, Revised Selected Papers. Lecture Notes in Computer Science, vol. 5867, pp. 295–312. Springer (2009), http://dx.doi.org/10.1007/978-3-642-05445-7_19
3. Blaze, M., Bleumer, G., Strauss, M.: Divertible protocols and atomic proxy cryptography. In: Nyberg, K. (ed.) Advances in Cryptology - EUROCRYPT '98, International Conference on the Theory and Application of Cryptographic Techniques, Espoo, Finland, May 31 - June 4, 1998, Proceeding. Lecture Notes in Computer Science, vol. 1403, pp. 127–144. Springer (1998), http://dx.doi.org/10.1007/BFb0054122
4. Boneh, D., Lewi, K., Montgomery, H.W., Raghunathan, A.: Key homomorphic prfs and their applications. In: Canetti, R., Garay, J.A. (eds.) Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I. Lecture Notes in Computer Science, vol. 8042, pp. 410–428. Springer (2013), http://dx.doi.org/10.1007/978-3-642-40041-4_23

5. Boneh, D., Lewi, K., Montgomery, H.W., Raghunathan, A.: Key homomorphic prfs and their applications. IACR Cryptology ePrint Archive 2015, 220 (2015), http://eprint.iacr.org/2015/220
6. Diaz-Santiago, S., Rodríguez-Henríquez, L.M., Chakraborty, D.: A cryptographic study of tokenization systems. In: Obaidat, M.S., Holzinger, A., Samarati, P. (eds.) SECRYPT 2014 - Proceedings of the 11th International Conference on Security and Cryptography, Vienna, Austria, 28-30 August, 2014. pp. 393–398. SciTePress (2014), http://dx.doi.org/10.5220/0005062803930398
7. European Commission, Article 29 Data Protection Working Party: Opinion 05/2014 on anonymisation techniques. Available online from http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/ (2014)
8. Everspaugh, A., Chatterjee, R., Scott, S., Juels, A., Ristenpart, T.: The pythia PRF service. In: Jung, J., Holz, T. (eds.) 24th USENIX Security Symposium, USENIX Security 15, Washington, D.C., USA, August 12-14, 2015. pp. 547–562. USENIX Association (2015), https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/everspaugh
9. Herzberg, A., Jakobsson, M., Jarecki, S., Krawczyk, H., Yung, M.: Proactive public key and signature systems. In: CCS '97, Proceedings of the 4th ACM Conference on Computer and Communications Security, Zurich, Switzerland, April 1-4, 1997. pp. 100–110 (1997), http://doi.acm.org/10.1145/266420.266442
10. Luchaup, D., Shrimpton, T., Ristenpart, T., Jha, S.: Formatted encryption beyond regular languages. In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014. pp. 1292–1303 (2014), http://doi.acm.org/10.1145/2660267.2660351
11. McCallister, E., Grance, T., Scarfone, K.: Guide to protecting the confidentiality of personally identifiable information (PII). NIST special publication 800-122, National Institute of Standards and Technology (NIST) (2010), available from http://csrc.nist.gov/publications/PubsSPs.html.
12. PCI Security Standards Council: PCI Data Security Standard (PCI DSS). https://www.pcisecuritystandards.org/document_library?document=pci_dss (2015)
13. Securosis: Tokenization guidance: How to reduce PCI compliance costs. https://securosis.com/assets/library/reports/TokenGuidance-Securosis-Final.pdf
14. Smart Card Alliance: Technologies for payment fraud prevention: EMV, encryption and tokenization. http://www.smartcardalliance.org/downloads/EMV-Tokenization-Encryption-WP-FINAL.pdf
15. United States Deapartment of Health and Human Services: Summary of the HIPAA Privacy Rule. http://www.hhs.gov/sites/default/files/privacysummary.pdf
16. Voltage Security: Voltage secure stateless tokenization. https://www.voltage.com/wp-content/uploads/Voltage_White_Paper_SecureData_SST_Data_Protection_and_PCI_Scope_Reduction_for_Todays_Businesses.pdf

A Proof of Theorem 1

We now prove that our strongest notion of IND-COTH indeed implies both IND-HOCH and IND-COHH.

Proof. Clearly, the oracles granted in the IND-COHH and IND-HOCH experiments are subsets of the oracles available to the adversary $\mathcal{A}$ in the IND-COTH experiment. Here we show that the winning conditions of both IND-COHH and IND-HOCH experiments are also covered by the IND-COTH experiment.

For the case of the IND-COHH experiment, we see that conditions a) and b) are equivalent to the winning conditions of the IND-COTH experiment. (Note that conditions c-i) to c-iii) of the IND-COTH experiment only apply if $\mathcal{A}$ makes a $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$ query, which is not allowed in the IND-COHH experiment.)

For the case of the IND-HOCH experiment, the winning conditions depend on whether the host was corrupted before or after the challenge epoch $\tilde{e}$. We analyze how conditions a) and b) of the IND-HOCH experiment are reflected in conditions b), c-i) and c-ii) of the IND-COTH experiment. (Note that conditions a) and c-iii) are based on the owner being corrupted, which does not apply to the IND-HOCH experiment.)

If an adversary $\mathcal{B}$ in the IND-HOCH experiment wins under condition b), *i.e.*, it corrupts the host at an epoch $e_h^* > \tilde{e} + 1$, or the host is not corrupted at all, $\mathcal{B}$ is only required to not make a tokenization query on $\tilde{x}_0$ or $\tilde{x}_1$ at the challenge epoch. We see that this condition is satisfied by condition b) of the IND-COTH experiment. Since $\mathcal{B}$ is allowed to make $\mathcal{O}_{\mathsf{token}}$ queries on $\tilde{x}_0$ or $\tilde{x}_1$ in all other epochs, we must argue that this is permitted by conditions c-i) and c-ii) of the IND-COTH experiment. Condition c-i) holds trivially as there is no epoch $e' \leq \tilde{e}$ in which a $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$ query was made, and condition c-ii) is always fulfilled with $e'' = \tilde{e} + 1$.

An adversary $\mathcal{B}$ winning the IND-HOCH experiment under condition a) is significantly more restricted in its $\mathcal{O}_{\mathsf{token}}$ queries. When the host is corrupted at an epoch $e_h^* \leq \tilde{e} + 1$, $\mathcal{B}$ is allowed to make $\mathcal{O}_{\mathsf{token}}$ queries on $\tilde{x}_0$ or $\tilde{x}_1$ latest at an epoch $e_{\mathsf{last}} \leq e_h^* - 2 \leq \tilde{e} - 1$. This immediately satisfies condition c-ii) as there can be no tokenization queries on $\tilde{x}_0$ or $\tilde{x}_1$ after the challenge epoch $\tilde{e}$ and thus $e_{\mathsf{first}} = \bot$. Condition c-i) is satisfied with $e' = e_h^* - 1$ and $e_{\mathsf{last}} \leq e_h^* - 2$, since $e_{\mathsf{last}} < e' \leq \tilde{e}$.

B Security of the $\mathsf{UTO}_{\mathsf{SE}}$ Scheme

In this section we show that our $\mathsf{UTO}_{\mathsf{SE}}$ construction is correct and achieves the notions of IND-HOCH, IND-COHH, and one-wayness as defined in Section 3. We also argue why this scheme does not satisfy the stronger IND-COTH notion.

Correctness. Let $\mathcal{X}$ be the data space of our updatable tokenization scheme. For any $x \in \mathcal{X}$, random encryption key s_e output by $\mathsf{SE.KeyGen}(\lambda)$ and random hash key $hk \xleftarrow{r} \mathsf{H.KeyGen}(\lambda)$, $\mathsf{UTO.token}(k_e, x)$ outputs $y_e = \mathsf{SE.Enc}(s_e, \mathsf{H}(hk, x))$. We see that $\mathsf{UTO.upd}(\Delta_{e+1}, y_e)$ (with $\Delta_{e+1} = (s_e, s_{e+1})$ and random s_{e+1} output by $\mathsf{SE.KeyGen}(\lambda)$) outputs $\mathsf{SE.Enc}(s_{e+1}, \mathsf{SE.Dec}(s_e, y_e)) = \mathsf{SE.Enc}(s_{e+1}, \mathsf{SE.Dec}(s_e, \mathsf{SE.Enc}(s_e, \mathsf{H}(hk, x)))) = \mathsf{SE.Enc}(s_{e+1}, \mathsf{H}(hk, x))$, which is also the output of $\mathsf{UTO.token}(k_{e+1}, x)$. Therefore, correctness is satisfied.

Theorem 4 (IND-COHH Security of the $\mathsf{UTO}_{\mathsf{SE}}$ Scheme). *Assume* $\mathsf{SE} = (\mathsf{SE.KeyGen}, \mathsf{SE.Enc}, \mathsf{SE.Dec})$ *is an* IND-CPA secure *deterministic symmetric encryption scheme and* H *is a weakly collision-resistant hash function. Then* $\mathsf{UTO}_{\mathsf{SE}}$ *is* IND-COHH secure *in the sense of Definition 4.*

Proof. Assume an IND-COHH adversary $\mathcal{A}_{\mathsf{UTO}}$ against the updatable tokenization scheme $\mathsf{UTO}_{\mathsf{SE}}$. We construct an adversary $\mathcal{A}_{\mathsf{SE}}$ that breaks the IND-CPA security of SE. Concretely, $\mathcal{A}_{\mathsf{SE}}$ simulates the IND-COHH experiment of Definition 4 for $\mathcal{A}_{\mathsf{UTO}}$, and concurrently plays the IND-CPA experiment of Definition 1. Let e_{max} be a polynomial upper bound on the total number of epochs used by the $\mathsf{UTO}_{\mathsf{SE}}$ scheme.

The idea is that in order to provide a perfect simulation, $\mathcal{A}_{\mathsf{SE}}$ will randomly select a value $\tilde{g} \leftarrow \{0, 1, \ldots, e_{\mathsf{max}} - 1\}$, guessing in which epoch $\mathcal{A}_{\mathsf{UTO}}$ will make the challenge query. For all epochs $e \neq \tilde{g}$, $\mathcal{A}_{\mathsf{SE}}$ will generate tokenization keys on its own and answer $\mathcal{O}_{\mathsf{token}}(x)$, $\mathcal{O}_{\mathsf{next}}$ and $\mathcal{O}_{\mathsf{corrupt\text{-}o}}$ queries made by $\mathcal{A}_{\mathsf{UTO}}$, whereas at epoch $\tilde{g}$, $\mathcal{A}_{\mathsf{SE}}$ will forward the challenge query and $\mathcal{O}_{\mathsf{token}}(x)$ queries to its own IND-CPA challenger, and respond to $\mathcal{A}_{\mathsf{UTO}}$ accordingly. More precisely, the simulation works as follows.

setup: The IND-CPA experiment selects $s \xleftarrow{\text{r}} \mathsf{SE.KeyGen}(\lambda)$ and $d \xleftarrow{\text{r}} \{0, 1\}$. $\mathcal{A}_{\mathsf{SE}}$ picks $\tilde{g} \xleftarrow{\text{r}} \{0, 1, \ldots, e_{\mathsf{max}} - 1\}$, $hk \xleftarrow{\text{r}} \mathsf{H.KeyGen}(\lambda)$, and $s_0 \xleftarrow{\text{r}} \mathsf{SE.KeyGen}(\lambda)$.

oracle queries: For $e < \tilde{g}$, $\mathcal{A}_{\mathsf{SE}}$ answers oracle queries to $\mathcal{A}_{\mathsf{UTO}}$ in the following way:

a) $\mathcal{O}_{\mathsf{next}}$: if $e \neq \tilde{g} - 1$, increment $e \leftarrow e + 1$ and run $s_{e+1} \xleftarrow{\text{r}} \mathsf{SE.KeyGen}(\lambda)$.
b) $\mathcal{O}_{\mathsf{token}}(x)$: compute $y_e \leftarrow \mathsf{SE.Enc}(s_e, \mathsf{H}(hk, x))$ and return y_e to $\mathcal{A}_{\mathsf{UTO}}$.
c) $\mathcal{O}_{\mathsf{corrupt\text{-}o}}$: abort simulation if oracle queried.
d) challenge $(\tilde{x}_0, \tilde{x}_1)$: abort simulation if challenge input received.

For $e = \tilde{g}$ the queries are answered as follows:

a) $\mathcal{O}_{\mathsf{next}}$: increment $e \leftarrow e + 1$ and run $s_{e+1} \xleftarrow{\text{r}} \mathsf{SE.KeyGen}(\lambda)$.
b) $\mathcal{O}_{\mathsf{token}}(x)$: abort simulation if x has already been given as one of the inputs to the challenge, *i.e.* if $x \in \{\tilde{x}_0, \tilde{x}_1\}$. Otherwise compute $\mathsf{H}(hk, x)$ and query the LoR oracle of the IND-CPA experiment, i.e., $\mathcal{O}_{\mathsf{enc}}$, with $(\mathsf{H}(hk, x), \mathsf{H}(hk, x))$, obtaining $y_e = \mathsf{SE.Enc}(s, \mathsf{H}(hk, x))$. Return y_e to $\mathcal{A}_{\mathsf{UTO}}$.
c) $\mathcal{O}_{\mathsf{corrupt\text{-}o}}$: abort simulation if oracle queried.
d) challenge $(\tilde{x}_0, \tilde{x}_1)$: abort simulation if challenge input *not* received, or if at least one of $\mathcal{O}_{\mathsf{token}}(\tilde{x}_0), \mathcal{O}_{\mathsf{token}}(\tilde{x}_1)$, was queried at epoch e. Otherwise compute $\mathsf{H}(hk, \tilde{x}_0)$, $\mathsf{H}(hk, \tilde{x}_1)$ and query $\mathcal{O}_{\mathsf{enc}}$ with $(\mathsf{H}(hk, \tilde{x}_0), \mathsf{H}(hk, \tilde{x}_1))$ obtaining $\tilde{y}_{d,\tilde{g}} = \mathsf{SE.Enc}(s, \mathsf{H}(hk, \tilde{x}_d))$. Forward $\tilde{y}_{d,\tilde{g}}$ to $\mathcal{A}_{\mathsf{UTO}}$.

For $e > \tilde{g}$ the queries are answered as follows:

a) $\mathcal{O}_{\mathsf{next}}$: increment $e \leftarrow e + 1$ and run $s_{e+1} \xleftarrow{\text{r}} \mathsf{SE.KeyGen}(\lambda)$.
b) $\mathcal{O}_{\mathsf{token}}(x)$: compute $y_e \leftarrow \mathsf{SE.Enc}(s_e, \mathsf{H}(hk, x))$ and return y_e to $\mathcal{A}_{\mathsf{UTO}}$.
c) $\mathcal{O}_{\mathsf{corrupt\text{-}o}}$: return $k_e = (s_e, hk)$ to $\mathcal{A}_{\mathsf{UTO}}$.
d) challenge $(\tilde{x}_0, \tilde{x}_1)$: at this point the challenge was already queried or the simulation was aborted.

output: $\mathcal{A}_{\mathsf{UTO}}$ outputs a bit d', which $\mathcal{A}_{\mathsf{SE}}$ forwards to its IND-CPA experiment.

In the simulation above $\mathcal{A}_{\mathsf{UTO}}$ could have queried $\mathcal{O}_{\mathsf{token}}(x)$ at epoch $\tilde{g}$ such that $\mathsf{H}(hk, x) = \mathsf{H}(hk, \tilde{x}_0)$ or $\mathsf{H}(hk, x) = \mathsf{H}(hk, \tilde{x}_1)$. This would allow $\mathcal{A}_{\mathsf{UTO}}$ to obtain $\tilde{y}_{0,\tilde{g}} = \mathsf{SE.Enc}(s, \mathsf{H}(hk, \tilde{x}_0))$ or $\tilde{y}_{1,\tilde{g}} = \mathsf{SE.Enc}(s, \mathsf{H}(hk, \tilde{x}_1))$, respectively, and compare with $\tilde{y}_{d,\tilde{g}}$. We denote this probability of collision by coll.

We see that if no hash collision was found, and if an adversary $\mathcal{A}_{\mathsf{UTO}}$ against our $\mathsf{UTO}_{\mathsf{SE}}$ wins its (simulated) IND-COHH security experiment, then $\mathcal{A}_{\mathsf{SE}}$ also wins its own IND-CPA experiment. Thus, we have that

$$\Pr\left[\mathcal{A}_{\mathsf{SE}} \text{ wins}\right] \geq \left|\Pr\left[\mathcal{A}_{\mathsf{UTO}} \text{ wins}\right] - \mathsf{coll}\right|,$$

which can be written as $\Pr\left[\mathcal{A}_{\mathsf{UTO}} \text{ wins}\right] \leq \left|\Pr\left[\mathcal{A}_{\mathsf{SE}} \text{ wins}\right] + \mathsf{coll}\right|$.

This means that if SE is IND-CPA secure (see Definition 1) and H is a weakly collision-resistant hash function, then our $\mathsf{UTO}_{\mathsf{SE}}$ scheme is IND-COHH secure in the sense of Definition 4, which concludes our proof. □

Theorem 5 (IND-HOCH Security of the $\mathsf{UTO}_{\mathsf{SE}}$ Scheme). *Assume* $\mathsf{SE} = (\mathsf{SE.KeyGen}, \mathsf{SE.Enc}, \mathsf{SE.Dec})$ *is an* IND-CPA secure *deterministic symmetric encryption scheme and* H *is a pseudorandom function. Then* $\mathsf{UTO}_{\mathsf{SE}}$ *is* IND-HOCH secure *in the sense of Definition 3.*

Proof. Assume an IND-HOCH adversary $\mathcal{A}_{\mathsf{UTO}}$ against the updatable tokenization scheme $\mathsf{UTO}_{\mathsf{SE}}$. We construct an adversary $\mathcal{A}$ that breaks the IND-CPA security of SE or the pseudorandomness of H. Concretely, $\mathcal{A}$ simulates the IND-HOCH experiment of Definition 3 for $\mathcal{A}_{\mathsf{UTO}}$, and concurrently plays the IND-CPA experiment of Definition 1 or the pseudorandomness experiment specified in Section 2. Let e_{max} be a polynomial upper bound on the total number of epochs used by our $\mathsf{UTO}_{\mathsf{SE}}$ scheme.

The idea is that in order to provide the simulation, $\mathcal{A}$ will first flip a coin to guess whether or not $\mathcal{A}_{\mathsf{UTO}}$ will corrupt the host at an epoch $e_h^* \leq \tilde{e} + 1$. The way $\mathcal{A}$ will be behave depends on its guess. If the guess is yes, $\mathcal{A}_{\mathsf{UTO}}$ is assumed not to be allowed to make tokenizing queries on challenge values $\tilde{x}_0$, $\tilde{x}_1$ at any epoch $e \geq e_h^* - 1$, and also able to get all update tweaks from epoch e_h^* onwards, which includes the encryption key $s_{\tilde{e}}$ of the challenge epoch $\tilde{e}$. Since $\mathcal{A}_{\mathsf{UTO}}$ will be able to decrypt its challenge using the key $s_{\tilde{e}}$, in this case the security of $\mathsf{UTO}_{\mathsf{SE}}$ depends *at a first glance* only on the pseudorandomness of H. Given that the latter statement assumes that $\mathcal{A}_{\mathsf{UTO}}$ obtains no information whatsoever about $\mathsf{H}(hk, \tilde{x}_0)$ or $\mathsf{H}(hk, \tilde{x}_1)$, we must argue that this is indeed the case. For this, note that although $\mathcal{A}_{\mathsf{UTO}}$ is allowed to make tokenization queries on $\tilde{x}_0$ and $\tilde{x}_1$ at any epoch $e \leq e_h^* - 2$, $\mathcal{A}_{\mathsf{UTO}}$ *cannot* obtain the encryption key s_e of those epochs, and by the IND-CPA security property of SE, the values $\mathsf{SE.Enc}(s_e, \mathsf{H}(hk, x))$ reveal no information about $\mathsf{H}(hk, x)$ even if the adversary knows some plaintext/ciphertext pairs, according to Definition 1. Therefore, we see that the security of $\mathsf{UTO}_{\mathsf{SE}}$ also depends on the IND-CPA security of SE. For this proof, *i.e.*, for the case where the guess is that $e_h^* \leq \tilde{e}+1$, we assume that SE is IND-CPA secure and build an adversary $\mathcal{A}$ that breaks the pseudorandomness of H. During the simulation $\mathcal{A}$ will generate the encryption keys of all epochs, but will not generate the hash key of H. Instead of computing the hash values

of x on its own, $\mathcal{A}$ will forward x to the PRF experiment, which will always use either a random function f or the hash function H with a random key hk to compute x. Now, the guess that $\mathcal{A}_{\mathsf{UTO}}$ will not corrupt the host at an epoch $e_h^* \leq \tilde{e}+1$, assumes that $\mathcal{A}_{\mathsf{UTO}}$ is restricted to not make tokenization queries on the challenge values *only* at the challenge epoch $\tilde{e}$, and that the update tweaks obtained by $\mathcal{A}_{\mathsf{UTO}}$ do not contain the encryption key $s_{\tilde{e}}$. The security of our $\mathsf{UTO}_{\mathsf{SE}}$ scheme here depends solely on the IND-CPA security of SE. So here the simulator $\mathcal{A}$ will act as an adversary against SE. For this, $\mathcal{A}$ will randomly select a value $\tilde{g} \leftarrow \{0, 1, \ldots, e_{\mathsf{max}} - 1\}$, guessing in which epoch $\mathcal{A}_{\mathsf{UTO}}$ will make the challenge query on $(\tilde{x}_0, \tilde{x}_1)$, and will use its IND-CPA oracle to respond to the challenge query and to all tokenization queries made at that epoch. Notice that the fact that $\mathcal{A}$ will not know the encryption key of the challenge epoch $\tilde{e}$ is not a problem for $\mathcal{A}$'s simulation as $\mathcal{A}_{\mathsf{UTO}}$ cannot query an update tweak containing $s_{\tilde{e}}$. For all other epochs $e \neq \tilde{e}$, $\mathcal{A}$ will randomly generate the encryption keys s_e.

Theorem 6 (One-Wayness of the $\mathsf{UTO}_{\mathsf{SE}}$ Scheme). *If* H *is one-way, then* $\mathsf{UTO}_{\mathsf{SE}}$ *is* one-way *in the sense of Definition 6.*

Proof. In the one-wayness experiment of Definition 6, an adversary $\mathcal{A}_{\mathsf{UTO}}$ against our $\mathsf{UTO}_{\mathsf{SE}}$ scheme having access to the tokenization key of epoch 0, $k_0 = (s_0, hk)$, receives as a challenge a token $\tilde{y} \leftarrow \mathsf{SE.Enc}(s_0, \mathsf{H}(hk, \tilde{x}))$ for random $s_0 \xleftarrow{\text{r}} \mathsf{SE.KeyGen}(\lambda)$, $hk \xleftarrow{\text{r}} \mathsf{H.KeyGen}(\lambda)$, and $\tilde{x} \xleftarrow{\text{r}} \mathcal{X}$. Since $\mathcal{A}_{\mathsf{UTO}}$ obtains s_0, it can decrypt $\tilde{y}$, obtaining $\mathsf{H}(hk, \tilde{x})$. We see that $\mathcal{A}_{\mathsf{UTO}}$ can only win the one-wayness experiment if it can break the one-wayness of H. As this is infeasible according to our stated assumption, then $\mathsf{UTO}_{\mathsf{SE}}$ is one-way.

$\mathsf{UTO}_{\mathsf{SE}}$ is not IND-COTH secure. We stress that although our updatable tokenization scheme $\mathsf{UTO}_{\mathsf{SE}}$ is IND-COHH and IND-HOCH secure, it does not achieve our stronger security notion IND-COTH. To see this, assume for instance that an adversary $\mathcal{A}_{\mathsf{UTO}}$ against our $\mathsf{UTO}_{\mathsf{SE}}$ scheme queries $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$ at epoch $\tilde{e}$, receiving $\Delta_{\tilde{e}} = (s_{\tilde{e}-1}, s_{\tilde{e}})$, and does not corrupt the host at epoch $\tilde{e}+1$. Assume also that $\mathcal{A}_{\mathsf{UTO}}$ queries $\mathcal{O}_{\mathsf{corrupt\text{-}o}}$ at epoch $e_0^* = \tilde{e}+1$, receiving $k_{\tilde{e}+1} = (s_{\tilde{e}+1}, hk)$. Note that with these corruptions of host and owner, $\mathcal{A}_{\mathsf{UTO}}$ gets $k_{\tilde{e}} = (s_{\tilde{e}}, hk)$, which allows it to trivially win the IND-COTH experiment by computing $\mathsf{SE.Enc}(s_{\tilde{e}}, \mathsf{H}(hk, \tilde{x_0}))$ or $\mathsf{SE.Enc}(s_{\tilde{e}}, \mathsf{H}(hk, \tilde{x_1}))$ on its own and comparing the result with the received challenge $\tilde{y}_{d,\tilde{e}} = \mathsf{SE.Enc}(s_{\tilde{e}}, \mathsf{H}(hk, \tilde{x_d}))$.

C Security of the $\mathsf{UTO}_{\mathsf{DL}}$ Scheme

We now show that our $\mathsf{UTO}_{\mathsf{DL}}$ construction is correct, one-way, and achieves the notion of IND-COTH security as defined in Section 3.

Correctness. Correctness is easy to verify: for any tokenized value $y_e \leftarrow \mathsf{H}(x)^{k_e}$ and update tweak Δ_{e+1}, output by $\mathsf{UTO.next}(k_e)$, $\mathsf{UTO.upd}(\Delta_{e+1}, y_e)$, produces $y_e^{\Delta_{e+1}} = y_e^{k_{e+1}/k_e} = (\mathsf{H}(x)^{k_e})^{k_{e+1}/k_e} = \mathsf{H}(x)^{k_{e+1}} = y_{e+1}$.

Theorem 7 (IND-COTH Security of the $\mathsf{UTO}_{\mathsf{DL}}$ Scheme). *Assume* H *behaves as a random oracle. Then, under the DDH assumption, the* $\mathsf{UTO}_{\mathsf{DL}}$ *scheme is* IND-COTH secure *in the sense of Definition 5.*

Proof. Assume an adversary $\mathcal{A}_{\mathsf{UTO}}$ against the updatable tokenization scheme $\mathsf{UTO}_{\mathsf{DL}}$. We construct an adversary $\mathcal{A}_{\mathsf{UTO}}$ that breaks the DDH assumption relative to a cyclic group $\mathbb{G} = \langle g \rangle$ of order p. Concretely, $\mathcal{A}_{\mathsf{UTO}}$ simulates the IND-COTH experiment of Definition 5 for $\mathcal{A}_{\mathsf{UTO}}$, and concurrently plays a DDH experiment as specified in Section 2. Since our proof is in the random oracle model, $\mathcal{A}_{\mathsf{UTO}}$ can only obtain hash values by querying a random oracle, which is also simulated by $\mathcal{A}_{\mathsf{UTO}}$.

Assume $\mathcal{A}_{\mathsf{UTO}}$ receives (g, g^a, g^b, T) from the DDH experiment and needs to decide whether or not $T = g^{ab}$. The simulator, will use g, g^a and g^b to answer $\mathcal{A}_{\mathsf{UTO}}$'s queries, and will embed the DDH challenge T in its response to the challenge query made by the adversary. In the simulation $\mathcal{A}_{\mathsf{UTO}}$ will choose a uniformly chosen bit d in a way that when T is the Diffie-Hellman value g^{ab}, then all the values returned to $\mathcal{A}_{\mathsf{UTO}}$ in the simulation are according to the IND-COTH experiment, and the response to the challenge query corresponds to the tokenized value of $\tilde{x}_d$. When T is g^c, for a uniformly chosen $c \in \mathbb{Z}_p^*$, the response to the challenge query is uniformly distributed in the token space, and $\mathcal{A}_{\mathsf{UTO}}$ has no information about the chosen bit d. Therefore, in that case $\mathcal{A}_{\mathsf{UTO}}$'s chance of winning the simulation is $1/2$.

At the end of the simulation, $\mathcal{A}_{\mathsf{UTO}}$ will guess that $T = g^{ab}$ if and only if $\mathcal{A}_{\mathsf{UTO}}$ outputs d. We argue that if $\mathcal{A}_{\mathsf{UTO}}$ can break the IND-COTH security of $\mathsf{UTO}_{\mathsf{DL}}$, then $\mathcal{A}_{\mathsf{UTO}}$ can break the DDH assumption. Let $\tilde{e}$ denote the challenge epoch and $(\tilde{x}_0, \tilde{x}_1)$ the pair of challenge values that will be given by $\mathcal{A}_{\mathsf{UTO}}$ as input to the challenge query. Roughly, the simulator proceeds as follows.

Flip a coin, $\mathsf{coin}_1 \xleftarrow{r} \{0,1\}$, to guess whether the adversary will, during the whole simulation, ever make a tokenization query, or a random oracle query, on a challenge value.

1. $\mathsf{coin}_1 = 0$: The guess is yes. Here the idea is that the simulator will set g^a as the hash output of one of the challenge values, and will set b as a tokenization key not obtained by $\mathcal{A}_{\mathsf{UTO}}$. All other hash outputs will be consistently set to g^r for a random $r \in \mathbb{Z}_p^*$ per input x. The response to the challenge query will be T^{Δ}, where Δ is the product of the update tweaks of the epoch immediately after the epoch where b was embedded, up to epoch $\tilde{e}$.
   (a) Flip a coin $d \xleftarrow{r} \{0,1\}$, guessing that the adversary will make at least one tokenization query, or random oracle query, on $\tilde{x}_d$.
   (b) Flip a coin $\mathsf{coin}_{\mathsf{h1}} \xleftarrow{r} \{0,1\}$ to guess whether the adversary will corrupt the host in all epochs $e \leq \tilde{e}$.
      i. $\mathsf{coin}_{\mathsf{h1}} = 0$: the host is honest in at least one epoch $e \leq \tilde{e}$. This means that tokenization queries on challenge values at $e \leq \tilde{e}$ is allowed.
         A. Guess which will be the *last* epoch, before or at the challenge epoch, where the host is *not* corrupted by the adversary. Denote

this epoch by $e_h^{\mathsf{last\text{-}nc}}$. When the time arrives, *implicitly* set the tokenization key of epoch $e_h^{\mathsf{last\text{-}nc}}$ to b. Before this point, $\mathcal{A}_{\mathsf{UTO}}$ will randomly choose a tokenization key k_e for each epoch e. After $e_h^{\mathsf{last\text{-}nc}}$, $\mathcal{A}_{\mathsf{UTO}}$ will randomly choose update tweaks Δ_e for the next epochs. The tokenization key of those subsequent epochs will be the multiplication of b and a product of update tweaks. The fact that $\mathcal{A}_{\mathsf{UTO}}$ does not know b is not an issue in its simulation since there the owner cannot be corrupted before the challenge epoch, and $\mathcal{A}_{\mathsf{UTO}}$ can use g^b to answer the tokenization queries made by $\mathcal{A}_{\mathsf{UTO}}$. Moreover, in this setup the update tweaks of all, but of epoch $e_h^{\mathsf{last\text{-}nc}}$ are known to $\mathcal{A}_{\mathsf{UTO}}$. Since the host is assumed not to be corrupted at $e_h^{\mathsf{last\text{-}nc}}$, then the simulator can answer all $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$ queries.

B. Flip a coin $\mathsf{coin}_2 \xleftarrow{r} \{0,1\}$ to guess whether the adversary will make a tokenization query, or random oracle query, on $\tilde{x}_d$ before the challenge epoch.

- $\mathsf{coin}_2 = 0$: The guess is yes.
  - Guess when the adversary will make its first tokenization query, or random oracle query, on $\tilde{x}_d$. This guess will consider the epoch and the query number of the event. (Note that although in the simulation the adversary will not be allowed to make tokenization queries on challenge values at any epoch e in the range $e_h^{\mathsf{last\text{-}nc}} \leq e \leq \tilde{e}$, this is not the case for random oracle queries. This means that the guessed epoch can be any epoch smaller than, or equal to, the challenge epoch.)
    When the time arrives, set the hash output of $\tilde{x}_d$ to g^a. Notice that $\mathcal{A}_{\mathsf{UTO}}$ will not know the actual value of $\tilde{x}_d$ beforehand. When the adversary makes tokenization queries on $\tilde{x}_d$, the simulator will use g^a to compute the tokenized value. Note also that by the way b was set up, the simulator will never have to use the unknown value g^{ab} in its simulation.
- $\mathsf{coin}_2 = 1$: The guess is no, but there will be a tokenization query, or random oracle query, on $\tilde{x}_d$ after the challenge epoch $\tilde{e}$. By then the simulator will already know $\tilde{x}_d$, from the challenge query. Here there is no special set up before $\tilde{e}$.

ii. $\mathsf{coin}_{\mathsf{h1}} = 1$: The host is corrupted in *all* epochs $e \leq \tilde{e}$. This means that tokenization queries on challenge values is not allowed at $e \leq \tilde{e}$. However, this is not the case for random oracle queries. Moreover, the simulator still needs to be able to answer all $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$ queries, and to embed its DDH challenge T in the challenge query made by $\mathcal{A}_{\mathsf{UTO}}$. For these reasons, $\mathcal{A}_{\mathsf{UTO}}$ will:

A. Implicitly set the tokenization key of epoch 0 to b, and randomly choose update tweaks Δ_e for the next epochs. From epoch 0

until epoch $\tilde{e}-1$, the simulator will use g^b in the computation of tokenized values.

B. Flip a coin $\mathsf{coin_3} \xleftarrow{r} \{0,1\}$ to guess whether the adversary will make a random oracle query on $\tilde{x}_d$ before the challenge epoch.
   - $\mathsf{coin_3} = 0$: The guess is yes.
     - Guess when the adversary will make its first random oracle query on $\tilde{x}_d$ before the challenge epoch. This guess will consider the epoch and the query number of the event.
       When the time arrives, set the hash output of $\tilde{x}_d$ to g^a. As before, note that $\mathcal{A}_{\mathsf{UTO}}$ will not know the actual value of $\tilde{x}_d$ beforehand.
   - $\mathsf{coin_3} = 1$: The guess is no, but there will be a tokenization query, or random oracle query, on $\tilde{x}_d$ after the challenge epoch $\tilde{e}$. By then the simulator will already know $\tilde{x}_d$, from the challenge query. There is no special set up before $\tilde{e}$ here.

(c) Flip a coin $\mathsf{coin_{h2}} \xleftarrow{r} \{0,1\}$ to guess whether the adversary will corrupt the host in all epochs $e > \tilde{e}$.

i. $\mathsf{coin_{h2}} = 0$: The host is honest in at least one epoch $e > \tilde{e}$. This means that there might be an $\mathcal{O}_{\mathsf{corrupt\text{-}o}}$ query, or tokenization queries on challenge values.

A. Guess which will be the *first* epoch, after the challenge epoch, where the host is *not* corrupted by the adversary. Denote this epoch by $e_h^{\mathsf{first\text{-}nc}}$. When the time arrives, create a fresh and uniformly chosen tokenization key for epoch $e_h^{\mathsf{first\text{-}nc}}$. The set up of a fresh tokenization key at epoch $e_h^{\mathsf{first\text{-}nc}}$ is transparent to the adversary since it will not corrupt the host at that epoch and consequently cannot update any tokenized value previously received to check for consistency. The tokenization keys of all subsequent epochs will also be randomly generated by $\mathcal{A}_{\mathsf{UTO}}$.
   When $\tilde{x}_d$ appears in tokenization queries or random oracle queries, set the hash output of $\tilde{x}_d$ to g^a. At this point $\mathcal{A}_{\mathsf{UTO}}$ will know the value of $\tilde{x}_d$, and thus will be expecting it.
   According to the IND-COTH experiment, at epochs $e > \tilde{e}$, the adversary is only allowed to make tokenization queries on challenge values, or to corrupt the owner from epoch $e_h^{\mathsf{first\text{-}nc}}$ onwards. So by setting a fresh tokenization key for epoch $e_h^{\mathsf{first\text{-}nc}}$, the exponent b will not be part of the tokenization key anymore, and $\mathcal{A}_{\mathsf{UTO}}$ can appropriately reply to tokenization queries on the challenge values by using g^a in the computation of the tokenized values. For all epochs e in the range $\tilde{e} < e < e_h^{\mathsf{first\text{-}nc}}$, the simulator will use g^b for the computation of the tokenized values. Now, for $\mathcal{O}_{\mathsf{corrupt\text{-}o}}$ and $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$ queries, first note that the simulator has all the tokenization keys for the epochs where the adversary can corrupt the owner, i.e., the epochs $e \geq e_h^{\mathsf{first\text{-}nc}}$. Second, notice that the update tweaks of all epochs $e > \tilde{e}$, but of epoch $e_h^{\mathsf{first\text{-}nc}}$, are known to $\mathcal{A}_{\mathsf{UTO}}$.

ii. $\mathsf{coin}_{\mathsf{h2}} = 1$: the host is corrupted in *all* epochs $e > \tilde{e}$. This means that there is no $\mathcal{O}_{\mathsf{corrupt\text{-}o}}$ query and no tokenization queries on challenge values. However, the host still needs to be able to respond to all $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$ queries.

A. The simulator will randomly choose update tweaks for all epochs $e > \tilde{e}$ and will use g^b to answer tokenization queries. The hash output of $\tilde{x}_d$ will be set to g^a. The simulator will use this value whenever $\mathcal{A}_{\mathsf{UTO}}$ makes a random oracle query on $\tilde{x}_d$.

2. $\mathsf{coin}_1 = 1$: The guess is no, the adversary will not make any tokenization query or random oracle queries, on challenge values during the whole simulation. Here the idea is that the simulator will choose two values $r_0, r_1 \xleftarrow{r} \mathbb{Z}_p^*$, set $g^{a \cdot r_0}$ as the hash output of $\tilde{x}_0$ and $g^{a \cdot r_1}$ as the hash output of $\tilde{x}_1$, and implicitly set b as the tokenization key of epoch 0. The hash outputs of all other values will be consistently set to g^r for a random $r \in \mathbb{Z}_p^*$ per input x. The simulator will randomly and uniformly generate the update tweaks of all epochs e in the range $0 < e < e_h^{\mathsf{first\text{-}nc}}$, where as in item 1, $e_h^{\mathsf{first\text{-}nc}}$ is a guess for the *first* epoch after the challenge epoch where the host will *not* be corrupted by the adversary. From $e_h^{\mathsf{first\text{-}nc}}$ on, the simulator will randomly and uniformly generate all tokenization keys. This set up will enable the simulator to not only answer an $\mathcal{O}_{\mathsf{corrupt\text{-}o}}$ query, but also $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$ queries. Notice that if the simulator started self generating tokenization keys at an epoch e prior to $e_h^{\mathsf{first\text{-}nc}}$, then it would not be able to answer an $\mathcal{O}_{\mathsf{corrupt\text{-}h}}$ query at epoch e since b, which is unknown to the simulator, would be one of the factors of the update tweak of that epoch. It is easy to see that although $\mathcal{A}_{\mathsf{UTO}}$ does not have the tokenization keys of the epochs e in the range $0 < e < e_h^{\mathsf{first\text{-}nc}}$, it can answer all tokenization queries by using g^b in the computation of a tokenized value; since the simulator is assuming that there will be no tokenization queries on a challenge value, g^{ab} will never be needed in those computations. By set up, the simulator has, or can compute, all update tweaks, except for the one of epoch $e_h^{\mathsf{first\text{-}nc}}$, where the adversary is assumed to not corrupt the host anyway. For the challenge query, $\mathcal{A}_{\mathsf{UTO}}$ will flip a coin $d \xleftarrow{r} \{0,1\}$ and answer with $T^{r_d \cdot \Delta}$, where Δ is the product of the update tweaks of epoch 1 up to epoch $\tilde{e}$.

Notice that in $\mathcal{A}_{\mathsf{UTO}}$'s simulation, the response to the challenge query will correspond to the tokenized value of $\tilde{x}_d$ whenever $T = g^{ab}$. Furthermore, in that case the simulation will be indistinguishable from the real experiment, and thus if $\mathcal{A}_{\mathsf{UTO}}$ outputs a bit $d' = d$, this is equivalent to winning the IND-COTH experiment. When $T = g^c$ the response to the challenge query will be a uniformly distributed value in the token space of $\mathsf{UTO}_{\mathsf{DL}}$ that has no relation to any other value received by $\mathcal{A}_{\mathsf{UTO}}$, and therefore the adversary's probability of succeeding will equal 1/2.

At the end of the simulation $\mathcal{A}_{\mathsf{UTO}}$ will output 0 if and only if $d' = d$. Considering that $\mathcal{A}_{\mathsf{UTO}}$ did not abort the simulation, we have

$$\begin{aligned}\mathsf{Adv}^{\mathsf{DDH}}_{\mathcal{A}_{\mathsf{UTO}},\mathsf{DDHgen}}(\lambda) &= \left|\Pr[0 \leftarrow \mathcal{A}_{\mathsf{UTO}} \,|\, g^{ab}] - \Pr[0 \leftarrow \mathcal{A}_{\mathsf{UTO}} \,|\, g^{c}]\right|, \\ &= |\Pr[\mathcal{A}_{\mathsf{UTO}} \text{ wins}] - 1/2| \\ &= \mathsf{Adv}^{\mathsf{IND\text{-}COTH}}_{\mathcal{A}_{\mathsf{UTO}},\mathsf{UTO}}(\lambda).\end{aligned}$$

We stress that even if the simulation runs until $\mathcal{A}_{\mathsf{UTO}}$ outputs a bit d', i.e., $\mathcal{A}_{\mathsf{UTO}}$ does not abort because of a wrong guess, $\mathcal{A}_{\mathsf{UTO}}$ still has to check if the simulation was perfect and if it can make use of d' in its output to the DDH experiment. For example, if $\mathcal{A}_{\mathsf{UTO}}$ guessed that the adversary would make a tokenization query on $\tilde{x}_d$ at some point in the simulation but it did not, then $T = g^{ab}$ did *not* result in the answer to the challenge query corresponding to the tokenized value of $\tilde{x}_d$, and thus, if $\mathcal{A}_{\mathsf{UTO}}$ outputs $d' = d$, it does not mean that $\mathcal{A}_{\mathsf{UTO}}$ wins the IND-COTH experiment.

Theorem 8 (One-Wayness of the $\mathsf{UTO}_{\mathsf{DL}}$ Scheme). *Assume* H *is one-way. Then* $\mathsf{UTO}_{\mathsf{DL}}$ *is* one-way *in the sense of Definition 6.*

Proof. In the one-wayness experiment of Definition 6, an adversary $\mathcal{A}_{\mathsf{UTO}}$ against our $\mathsf{UTO}_{\mathsf{DL}}$ scheme having access to the tokenization key of epoch 0, k_0, receives as a challenge some tokenized value $\tilde{y} \leftarrow \mathsf{H}(\tilde{x})^{k_0}$ for random $k_0 \xleftarrow{\text{r}} \mathbb{Z}_p$, and $\tilde{x} \xleftarrow{\text{r}} \mathcal{X}$, with $\mathcal{X}$ being the data space of the updatable tokenization scheme. Since $\mathcal{A}_{\mathsf{UTO}}$ obtains k_0, it can retrieve $\mathsf{H}(\tilde{x})$. We see that $\mathcal{A}_{\mathsf{UTO}}$ can only win the one-wayness experiment if it can break the one-wayness of H with advantage better than negligible, considering $|\mathcal{X}|$ as the security parameter.

As this is infeasible according to our stated assumption, $\mathsf{UTO}_{\mathsf{DL}}$ is one-way.