A preliminary version of this paper appears in the proceedings of ESORICS 2014. This is the full version.

Even more practical secure logging: Tree-based Seekable Sequential Key Generators

Giorgia Azzurra Marson[1] and Bertram Poettering[2]

[1] CASED & TU Darmstadt
[2] Information Security Group at Royal Holloway, University of London

Abstract. Computer log files constitute a precious resource for system administrators for discovering and comprehending security breaches. A prerequisite of any meaningful log analysis is that attempts of intruders to cover their traces by modifying log entries are thwarted by storing them in a tamper-resistant manner. Some solutions employ cryptographic authentication when storing log entries locally, and let the authentication scheme's property of *forward security* ensure that the cryptographic keys in place at the time of intrusion cannot be used to manipulate past log entries without detection. This strong notion of security is typically achieved through frequent updates of the authentication keys via hash chains. However, as security demands that key updates take place rather often (ideally, at a resolution of milliseconds), in many settings this method quickly reaches the limits of practicality. Indeed, a log auditor aiming at verifying a specific log record might have to compute millions of hash iterations before recovering the correct verification key. This problem was addressed only recently by the introduction of *seekable sequential key generators* (SSKG). Every instance of this cryptographic primitive produces a forward-secure sequence of symmetric (authentication) keys, but also offers an explicit fast-forward functionality. The only currently known SSKG construction replaces traditional hash chains by the iterated evaluation of a *shortcut one-way permutation*, a factoring-based and hence in practice not too efficient building block. In this paper we revisit the challenge of marrying forward-secure key generation with seekability and show that symmetric primitives like PRGs, block ciphers, and hash functions suffice for obtaining secure SSKGs. Our scheme is not only considerably more efficient than the prior number-theoretic construction, but also extends the seeking functionality in a way that we believe is important in practice. Our construction is provably (forward-)secure in the standard model.

Keywords: secured logging, forward security, seekability, pseudorandom generators

1 Introduction

Computer log files can be configured to record a large variety of system events that occur on network hosts and communication systems, including users logging on or off, memory resources reaching their capacity, malfunctioning of disk drives, etc. Therefore, log files represent one of the most essential sources of information that support system administrators in understanding the activity of systems and keeping them fully functional. Not less important is the role that log files play in computer forensics: events like login failures and software crashes serve as standard indicators for (attempted) intrusions. Unfortunately, as log files are often recorded locally (i.e., on the monitored machine itself), in many practical cases intruders can a posteriori manipulate the log entries related to their attacks.

Online logging and its disadvantages. In a network environment, one obvious strategy to prevent adversarial tampering of audit logs is to forward log messages immediately after their creation to a remote log sink—in the hope that the attacker cannot also corrupt the latter. Necessary in such a setting is that the log sink is continuously available, as every otherwise required local buffering of log records would increase the risk that their delivery is suppressed by the adversary. However, in many cases it has to be assumed that the reachability of the log sink can be artificially restrained by the intruder, e.g., by confusing routing protocols with false ARP messages, by sabotaging TCP connections with injected reset packets, by jamming wireless connections, or by directing application-level denial-of-service attacks against the log sink. Independently of these issues, it is inherently difficult to choose an appropriate logging granularity: while the creation of individual records for each established TCP connection, file deletion, or subprocess invocation might be desirable from the point of view of computer forensics, network links and log sinks might quickly reach their capacities if events are routinely reported with such a high resolution. This holds in particular if log sinks serve multiple monitored hosts simultaneously.

Forward-secure cryptography for log file protection. A solution for tamper-resistant log-entry storage that does not require a remote log sink but offers integrity protection via cryptographic means is *secured local logging*. Here, each log entry is stored together with a specific authentication tag that is generated and verified using a secret key. Note that regular message authentication codes (MACs) by themselves seem not to constitute a secure solution, as corresponding tags will be forgeable by intruders that succeed in extracting the secret key from the attacked device. Rather, the forward-secure variant of a MAC is required, as elaborated next.

In a nutshell, a cryptosystem provides *forward security* (FS) if it continues to give meaningful security guarantees after the adversary got a copy of the used keys. A standard example is key exchange: here, all recent security models require established session keys to remain safe when the adversary obtains access to the involved long-term private keys [17,6]. Likely less known is that the notion of forward security also extends to non-interactive primitives. For instance, in forward-secure public key encryption [5] messages are encrypted in respect to a combination (pk, t), where pk is a public key and $t \in \mathbb{N}$ identifies one out of a set of consecutive time epochs; for each such epoch t, knowledge of a specific decryption key sk_t is necessary for decrypting corresponding ciphertexts. In addition, while by design it is efficiently possible to perform updates $sk_t \mapsto sk_{t+1}$, forward security requires that the reverse mapping be inefficient, i.e., it shall be infeasible to 'go backwards in time'. More precisely, forward security guarantees that plaintexts encrypted for 'expired' epochs remain confidential even if the decryption keys of all later epochs are revealed.

Analogously to the described setting, signatures and authentication tags of the forward-secure variants of signature schemes and MACs, respectively, remain unforgeable for past epochs if only current and future keys are disclosed to the adversary [2,4]. One possible way to obtain such a MAC is to combine a (forward-secure) *sequential key generator* (SKG) with a regular MAC [11,4], where the former can be seen as a stateful *pseudorandom generator* (PRG) that, once initialized with a random seed, deterministically outputs a pseudorandom sequence of fixed-length keys. These keys are then used together with a MAC to ensure unforgeability of messages within the epochs.

The challenge of seekability. Forward-secure SKGs are typically constructed by deterministically evolving an initially random state using a hash chain, i.e., by regularly replacing a 'current' key K_t by $K_{t+1} = H(K_t)$, where H is a cryptographic hash function [11,4]. Although hash chains, in principle, lead to (forward-)secure local logging, they also come with an efficiency penalty on the side of the log auditor: the latter, in order to verify a log record of a certain epoch t, first needs to recover the corresponding key K_t; however, as a high level of security requires a high key update rate, this might involve millions of hash function evaluations. This problem was addressed only recently: in [15], efficient forward-secure local logging is achieved via a *seekable sequential key generator* (SSKG).

We give a rough overview over the ideas in [15]. Essentially, the authors propose a generic construction of an SSKG from a *shortcut one-way permutation* (SCP), a primitive that implements a one-way permutation $\pi\colon D \to D$, for a domain D, with a dedicated shortcut algorithm allowing the computation of the k-fold composition π^k in sublinear time. The concrete SCP considered in [15] is given by the squaring operation modulo a Blum integer N, where applying the shortcut corresponds to reducing a certain exponent modulo $\varphi(N)$. Given an SCP, an SSKG can be obtained by letting its state consist of a single element in D, performing state updates by applying π to this element, and deriving keys by hashing it (more precisely, by applying a random oracle). While it is instructive to observe how the forward security of the SSKG corresponds with the one-wayness of the SCP, and how its seekability is based on the SCP's shortcut property, a notable technical artifact of the squaring-based SCP is that seekability requires knowledge of $\varphi(N)$ while forward security requires this value to be unknown. This dilemma is side-stepped in [15] by giving only the owners of a *seeking key* the ability to fast-forward through the SSKG output sequence.

1.1 Contributions and organization

The central contribution of this paper is the design of a new seekable sequential key generator. In contrast to the prior SSKG from [15], our scheme relies on just symmetric building blocks; in particular we propose instantiations that exclusively use either PRGs, block ciphers, or hash functions. By consequence, our implementation beats the one from [15] by 1–3 orders of magnitude, on current CPUs. In addition to this efficiency gain, we also identify new and appealing functionality features of our SSKG. In particular,

getting rid of the discussed seeking limitations of [15], our scheme allows *every* user to efficiently advance any state by an arbitrary number of epochs. Our SSKG is supported by a security proof in the standard model.

This paper is organized as follows. After starting with preliminaries in Section 2, we formally specify the functionality, syntax, and security requirements of SSKGs in Section 3; this includes both a comparison with the (different) formalizations in [15] and a concrete proposal on how SSKGs can be used to protect audit logs in practice. In Section 4 we describe our new PRG-based SSKG, including its generalized seekability notion and some possible time-memory trade-offs. Finally, in Section 5, we discuss implementational aspects and efficiency results from our implementation.

1.2 Related work

The first published work that highlights the importance of *seekability* as a desirable property of sequential key generators in the context of secured local logging is [15]. An extensive comparison of the corresponding results with the ones of the current paper can be found in the preceding paragraphs and in Section 3. In the following we discuss further publications and proposed protocols targeting the topics of sequential key generation and cryptographic audit log protection. We observe that all considered protocols either are forward-secure or offer seekability, but not both simultaneously.

An early approach towards secured local logging originates from Bellare and Yee [3]; they study the role of forward security in authentication, develop the security notion of *forward integrity*, and realize a corresponding primitive via a PRF chain. Later, the same authors provide the first systematic analysis of forward security in the symmetric setting [4], covering forward-secure variants of pseudorandom generators, symmetric encryption, and message authentication codes, and also providing constructions and formal proofs of security for these primitives.

Shortly after [3], an independent cryptographic scheme specifically targeted at protecting log files was described by Kelsey and Schneier [11,12,16]. Their scheme draws its (forward) security from frequent key updates via iterated hashing, but is unfortunately not supported by a formal security analysis. A couple of implementations exist, notably the one by Chong, Peng, and Hartel in tamper-resistant hardware [7], the construction of Stathopoulos, Kotzanikolaou, and Magkos [18], who investigate the role of secure logging in public communication networks, and the *logcrypt* system by Holt [9]. The latter improves on [11] by paving the way towards provable security, but also adds new functionality and concepts. Most notable is the suggestion to embed regular metronome entries into log files to thwart *truncation attacks* where the adversary cuts off the most recent set of log entries. Similar work is due to Accorsi [1] who presents *BBox*, a hash-chain-based framework for protecting the integrity and confidentiality of log files in distributed systems.

Ma and Tsudik consider the concept of *forward-secure sequential aggregate authentication* for protecting the integrity of system logs [13,14]. Their constructions build on compact constant-size authenticators with all-or-nothing security (i.e., adversarial deletion of *any* single log message is detected), naturally defend against truncation attacks, and enjoy provable security.

The proposals by Yavuz and Ning [19], and Yavuz, Ning, and Reiter [20], specifically aim at secured logging on constraint devices and support a shift of computation workload from the monitored host to the log auditor. Notably, their key update procedure and the computation of authentication tags takes only a few hash function evaluations and finite field multiplications. In common with the schemes discussed above, their authentication systems are not seekable.

Kelsey, Callas, and Clemm [10] introduced secured logging into the standardization process at IETF. However, their proposal of *signed syslog messages* focuses on remote logging instead of on local logging. Precisely, their extension to the standard UNIX *syslog* facility authenticates log entries via signatures before sending them to a log sink over the network. While this proposal naturally offers seekability, it is bound to the full-time availability of an online log sink. Indeed, periods where the latter is not reachable are not securely covered, as the scheme is not forward-secure.

2 Preliminaries

We recall basic notions and facts from cryptography, graph theory, and data structures that we require in the course of this paper. Notably, in the section on trees, we define what we understand by the 'co-path' of a node. If not explicitly specified differently, all logarithms are understood to be taken to base 2.

2.1 Pseudorandom generators

A *pseudorandom generator* (PRG) is a function that maps a random string ('*seed*') to a longer 'random-looking' string. The security property of *pseudorandomness* requires that it be infeasible to distinguish the output of a PRG from random:

Definition 1 (Pseudorandom generator). *For security parameter λ and a polynomial $c\colon \mathbb{N} \to \mathbb{N}^{\geq 1}$, an efficiently computable function $G\colon \{0,1\}^\lambda \to \{0,1\}^{\lambda+c(\lambda)}$ is a* pseudorandom generator *if for all efficient distinguishers $\mathcal{D}$ the following advantage function is negligible, where the probabilities are taken over the random choices of s and y, and over $\mathcal{D}$'s randomness:*

$$\mathrm{Adv}^{\mathsf{PRG}}_{G,\mathcal{D}}(\lambda) = \left|\Pr\left[\mathcal{D}(G(s)) = 1 \ : \ s \leftarrow_R \{0,1\}^\lambda\right] - \Pr\left[\mathcal{D}(y) = 1 \ : \ y \leftarrow_R \{0,1\}^{\lambda+c(\lambda)}\right]\right| \ .$$

As we discuss in Section 5, in practice PRGs can be obtained from appropriate block ciphers, stream ciphers, or hash functions.

2.2 Binary and *d*-ary trees

A *tree* is a simple, undirected, connected graph without cycles. We particularly consider rooted trees, i.e., trees with a distinguished *root* node. The nodes adjacent to the root node are called its *children*; each child can be considered, in turn, the root of a subtree. The *level* L of a node indicates its distance to the root, where we assign level $L = 1$ to the latter. Children of the same node are *siblings* of each other. We will assume that the children of each node are ordered, i.e., can be identified by a number $1 \leq i \leq d$, where d is the number of children. For two siblings with indices i and j, respectively, in case $i < j$ we say that node i is *left* of node j and that node j is *right* of node i. In binary trees we may also refer to the children as left and right directly. Nodes that have no children are called *leaves*, all other nodes are called *internal*. A tree is *d-regular* (or *d-ary*, or *binary* in case $d = 2$) if every internal node has exactly d children.

In this paper we focus on d-ary trees of constant height H, i.e., where all leaves have the same level $L = H$. If $\mu_d(L)$ denotes the number of nodes at level L, then for such trees we have $\mu_d(1) = 1$ and $\mu_d(L) = d \cdot \mu_d(L-1) = d^{L-1}$ for all $L > 1$. For the total number of nodes $\nu_d(H)$ we hence obtain

$$\nu_d(H) = \sum_{L=1}^{H} \mu_d(L) = \sum_{L=1}^{H} d^{L-1} = (d^H - 1)/(d-1) \ ,$$

by the geometric summation formula. As a special case, for binary trees the total number of nodes is $\nu_2(H) = 2^H - 1$.

We finally define the notion of *co-path* of a node. Let v denote an arbitrary node of a tree. Intuitively speaking, the (right) co-path of v is the list of the right siblings of the nodes on the (unique) path connecting the root node with v; if for individual nodes on this path there are multiple right siblings, all of them appear in the co-path. For a formal definition, let L denote the level of $v = v_L$ and let $(v_1, \ldots, v_L)$ denote the path that connects the root (denoted here with v_1) with v_L. For each $1 \leq i \leq L$ let $V_i^{\rightarrow}$ be the list of right siblings of node v_i, in left-to-right order (some of these lists might be empty, and particularly $V_1^{\rightarrow}$ always is). We define the co-path of v_L to be the list $V_L^{\rightarrow} \,\|\, \ldots \,\|\, V_1^{\rightarrow}$ obtained by combining these lists into a single one using concatenation.

2.3 Stacks and their operations

A *stack* is a standard data structure for the storage of objects. Stacks follow the last-in first-out principle: the last element stored in a stack is the first element to be read back (and removed). The following procedures can be used to operate on stacks for storing, reading, and deleting elements. By $\mathbf{Init}(\mathcal{S})$ we denote the initialization of a fresh and empty stack $\mathcal{S}$. To add an element x 'on top of' stack $\mathcal{S}$, operation $\mathbf{Push}(\mathcal{S}, x)$ is used. We write $x \leftarrow \mathbf{Pop}(\mathcal{S})$ for reading and removing the top element of stack $\mathcal{S}$. Finally, with $x \leftarrow \mathbf{Peek}_k(\mathcal{S})$ the k-th element of stack $\mathcal{S}$ can be read without deleting it; here, elements are counted from the top, i.e., $\mathbf{Peek}_1(\mathcal{S})$ reads the top-most element. When using these notations, operations **Init**, **Push**, and **Pop** are understood to modify their argument $\mathcal{S}$ in place, while $\mathbf{Peek}_k$ leaves it unchanged.

3 Seekable sequential key generators

The main contribution of this paper is a new construction of a *seekable sequential key generator* (SSKG). This cryptographic primitive can be seen as a stateful PRG that outputs a sequence of fixed-length keys—one per invocation. The specific property of *seekability* ensures that it is possible to jump directly to any position in the output sequence. At the same time, the security goal of forward security ensures that keys remain indistinguishable from random even upon corruption of the primitive's state. We next recall the syntactical definition and security properties, (mainly) following the notation from [15]. We defer the exposition of our new scheme to Section 4.

3.1 Functionality and syntax

Generally speaking, a seekable sequential key generator consists of four algorithms: $\mathsf{GenSSKG}$ generates an initial state st_0, the update procedure Evolve maps each state st_i to its successor state st_{i+1}, GetKey derives from any state st_i a corresponding (symmetric) key K_i, and Seek permits to compute any state st_i directly from initial state st_0 and index i. We consider each state associated with a specific period of time, called *epoch*, where the switch from epoch to epoch is carried out precisely with the Evolve algorithm. This setting is illustrated in Figure 1 and formalized in Definition 2.

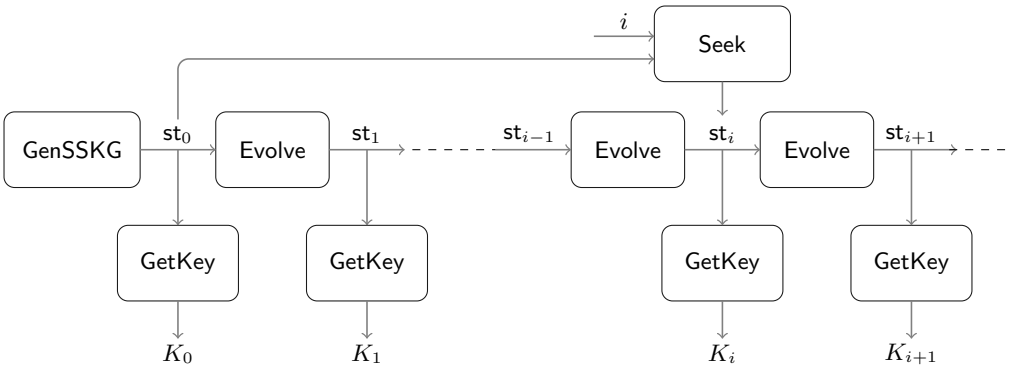

Fig. 1: Illustration of the interplay of the different SSKG algorithms.

Definition 2 (Syntax of SSKG). *Let $\ell\colon \mathbb{N} \to \mathbb{N}$ be a polynomial. A* seekable sequential key generator *with key length ℓ is a tuple $\mathsf{SSKG} = \{\mathsf{GenSSKG}, \mathsf{Evolve}, \mathsf{GetKey}, \mathsf{Seek}\}$ of efficient algorithms as follows:*

- $\mathsf{GenSSKG}$. *On input of security parameter 1^λ and total number $N \in \mathbb{N}$ of supported epochs, this probabilistic algorithm outputs an initial state st_0.*
- Evolve. *On input of a state st_i, this deterministic algorithm outputs the 'next' state st_{i+1}. For convenience, for $k \in \mathbb{N}$, by Evolve^k we denote the k-fold composition of Evolve, i.e., $\mathsf{Evolve}^k(\mathsf{st}_i) = \mathsf{st}_{i+k}$.*
- GetKey. *On input of state st_i, this deterministic algorithm outputs a key $K_i \in \{0,1\}^{\ell(\lambda)}$. For $k \in \mathbb{N}$, we write $\mathsf{GetKey}^k(\mathsf{st}_i)$ for $\mathsf{GetKey}(\mathsf{Evolve}^k(\mathsf{st}_i))$.*
- Seek. *On input of initial state st_0 and $k \in \mathbb{N}$, this deterministic algorithm returns state st_k.*

Implicit in Definition 2 is the following natural consistency requirement on the interplay of Evolve and Seek algorithms:

Definition 3 (Correctness of SSKG). *A seekable sequential key generator* SSKG *is* correct *if, for all security parameters λ, all $N \in \mathbb{N}$, all $\mathsf{st}_0 \leftarrow_R \mathsf{GenSSKG}(1^\lambda, N)$, and all $k \in \mathbb{N}$ we have*

$$0 \le k < N \qquad \Longrightarrow \qquad \mathsf{Evolve}^k(\mathsf{st}_0) = \mathsf{Seek}(\mathsf{st}_0, k) \ .$$

Remark 1 (Epochs outside of supported range). Note that the correctness requirement leaves unspecified the effect of Evolve and Seek algorithms when invoked in the context of unsupported epoch numbers $k \geq N$. Clearly, if the application demands it, an SSKG can always be implemented such that Evolve, Seek, and GetKey output an error symbol for such epochs, simply by including an epoch counter in the state. We abstain from formally requiring such a behavior as our construction in Section 4 in many cases does guarantee security for a larger number of epochs than requested and there seems to be no reason to generally disregard systems that offer this extra service.

Remark 2 (Comparison with the definition from [15]). The syntax specified in Definition 2 does slightly deviate from the one in [15, Definition 3]: firstly, the SSKG setup routine of [15] has a secret 'seeking key' as additional output; it is required as auxiliary input for the Seek algorithm. The necessity of this extra key should be considered an artifact of the number-theory-based construction from [15] (see Section 3.4 for details): the seeking key contains the factorization of the RSA modulus underlying the scheme. As the proposed Evolve algorithm is one-way only if this factorization is not known, the Seek algorithm is available exclusively to those who know the seeking key as a 'trapdoor'. In contrast to that, our syntax for Seek is not only more natural, we also allow *everybody* to use the Seek algorithm to fast-forward efficiently to future epochs. Secondly, in [15] the number of supported epochs does not have to be specified at the time of SSKG initialization; instead, an infinite number of epochs is supported by every instance. We had to introduce this restriction for technical reasons that become clear in Section 4; however, we believe that the requirement of specifying the number of epochs in advance does not constrain the practical usability of our scheme too much: indeed, regarding our scheme from Section 4, instantiations with, say, $N = 2^{30}$ supported epochs are perfectly practical.

3.2 Security requirements

As the security property of SSKGs we demand indistinguishability of generated keys from random strings of the same length. This is modeled in [15] via an experiment involving an adversary $\mathcal{A}$ who first gets adaptive access to a set of (real) keys K_i of her choosing, and is then challenged with a string K_n^b that is either the real key K_n or a random string of the same length; the adversary has to distinguish these two cases. This shall model the intuition that keys K_n 'look random' even if the adversary is given (all) other keys K_i, for $i \neq n$. Below we formalize a stronger security notion that also incorporates forward security, i.e., additionally lets the adversary corrupt any state that comes after the challenged epoch.

Definition 4 (IND-FS security of SSKG [15]). *A seekable sequential key generator* SSKG *is* indistinguishable with forward security against adaptive adversaries (IND-FS) *if, for all efficient adversaries* $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ *that interact in experiments* $\mathrm{Expt}^{\mathsf{IND\text{-}FS},b}$ *from Figure 2 and all* $N \in \mathbb{N}$ *bounded by a polynomial in* λ*, the following advantage function is negligible, where the probabilities are taken over the random coins of the experiment (including over* $\mathcal{A}$*'s randomness):*

$$\mathrm{Adv}^{\mathsf{IND\text{-}FS}}_{\mathsf{SSKG},N,\mathcal{A}}(\lambda) = \left|\Pr\left[\mathrm{Expt}^{\mathsf{IND\text{-}FS},1}_{\mathsf{SSKG},N,\mathcal{A}}(1^\lambda) = 1\right] - \Pr\left[\mathrm{Expt}^{\mathsf{IND\text{-}FS},0}_{\mathsf{SSKG},N,\mathcal{A}}(1^\lambda) = 1\right]\right| \quad .$$

3.3 Application of SSKGs: protecting locally stored log files

Given the definitions from Sections 3.1 and 3.2, the role of SSKGs in the context of secure logging is now immediate: in every epoch i, corresponding key K_i is used to instantiate a message authentication code (MAC) that equips all occurring log messages with an authentication tag. In addition, the Evolve algorithm is regularly invoked to advance from one epoch to the next, burying for all times the previously used keys. In such a setting, an auxiliary copy of initial state st_0 is made available to the log auditor who can use the Seek algorithm to check the integrity of log entries in any order. Clearly, the goal of forward security can be achieved only if the secure erasure of old states is an inherent part of the transition between epochs—for instance using the methods developed in [8].

3.4 Prior SSKG constructions

While general sequential key generators have been considered in a variety of publications [11,16,4,9], the importance of seekability to obtain practical secure logging was only identified very recently [15]. By consequence, we are aware of only a single SSKG that precedes our current work.

$\mathrm{Expt}^{\mathsf{IND\text{-}FS},b}_{\mathsf{SSKG},N,\mathcal{A}}(1^\lambda)$:

1 $\mathsf{KList} \leftarrow \emptyset$
2 $\mathsf{st}_0 \leftarrow_R \mathsf{GenSSKG}(1^\lambda, N)$
3 $(state, n, m) \leftarrow_R \mathcal{A}_1^{\mathcal{O}_{\mathsf{Key}}}(1^\lambda, N)$
4 Abort if not $0 \le n < m < N$
5 $K_n^0 \leftarrow_R \{0,1\}^{\ell(\lambda)}$
6 $K_n^1 \leftarrow \mathsf{GetKey}^n(\mathsf{st}_0)$
7 $\mathsf{st}_m \leftarrow \mathsf{Evolve}^m(\mathsf{st}_0)$
8 $b' \leftarrow_R \mathcal{A}_2^{\mathcal{O}_{\mathsf{Key}}}(state, \mathsf{st}_m, K_n^b)$
9 Abort if $n \in \mathsf{KList}$
10 Return b'

If $\mathcal{A}$ queries $\mathcal{O}_{\mathsf{Key}}(i)$:

1 Abort if not $0 \le i < N$
2 $\mathsf{KList} \leftarrow \mathsf{KList} \cup \{i\}$
3 $K_i \leftarrow \mathsf{GetKey}^i(\mathsf{st}_0)$
4 Answer $\mathcal{A}$ with K_i

Fig. 2: Security experiments for indistinguishability with forward security. The abort operation lets the experiment return 0, disregarding any output of the adversary.

Intuitively speaking, the SSKG construction from [15] follows the permute-then-hash paradigm. In more detail, the authors consider so-called shortcut one-way permutations $\pi\colon \mathcal{D} \to \mathcal{D}$ that allow the evaluation of the k-fold composition π^k in less than $\mathcal{O}(k)$ time. Given such a primitive, state st_0 consists of a random element $x_0 \in \mathcal{D}$, and keys K_i are computed as $K_i = H(\pi^i(x_0))$, where H is a hash function modeled as a random oracle. The authors propose a number-theory-based shortcut permutation where π implements precisely the squaring operation modulo a Blum integer N; in this case, $\pi^i(x) = x^{2^i} = x^{2^i \bmod \varphi(N)}$ can be evaluated quite efficiently if the factorization of N is known.

4 SSKGs from pseudorandom generators

We propose a novel construction of a seekable sequential key generator that assumes only symmetric building blocks. Unlike the scheme in [15] which draws security from shortcut one-way permutations in the random oracle model, our new SSKG assumes just the existence of PRGs, i.e., it relies on a minimal cryptographic assumption. In a nutshell, similarly to the works in [2] and [5] that achieve forward-secure signing and forward-secure public key encryption, respectively, we identify time epochs with the nodes of specially formed trees and let the progression of time correspond to a pre-order visit of these nodes. We will start with the exposition of our scheme in the case of binary trees, and then extend the results to the more general case.

4.1 Sequential key generators from binary trees

From Section 2.2 we know that for any fixed $H \in \mathbb{N}^{\ge 1}$ the binary tree of constant height H has exactly $N = \nu_2(H) = 2^H - 1$ nodes. In our SSKG we identify time epochs with the nodes of such a tree. More precisely, given the pre-order depth-first enumeration $w_0, \ldots, w_{N-1}$ of the nodes (first visit the root, then recursively the left subtree, then recursively the right subtree; cf. Figure 3), we let time epoch i and node w_i correspond.

The idea is to assign to each node w_i a (secret) seed $s_i \in \{0,1\}^\lambda$ from which the corresponding epoch's key K_i and the seeds of all subordinate nodes can be deterministically derived via PRG invocations. Here, exclusively the secret of the root node is assigned at random. Intuitively, the pseudorandomness of the PRG ensures that all keys and seeds look random to the adversary.

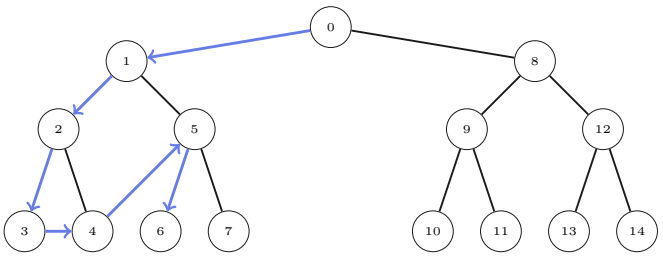

Fig. 3: A binary tree with height $H = 4$ and $N = 2^4 - 1 = 15$ nodes. The latter are numbered according to a pre-order depth-first search, as partially indicated by the arrow from the root node w_0 to node w_6.

We proceed with specifying which information the states associated with the epochs shall record. Recall that from each state st_i, $0 \le i < N$, two pieces of information have to be derivable: the epoch-specific key K_i and the successor state st_{i+1} (and, by induction, also all following states and keys). Clearly, in our construction, the notions of seed and state do not coincide; for instance, in the tree of Figure 3 key K_9 cannot be computed from just seed s_4. However, if state st_4 contained (s_4, s_5, s_8), then for all $4 \le i < N$ the keys K_i could be computed from this state. Inspired by this observation, our SSKG stores in each state st_i a collection of seeds, namely the seeds of the roots of the 'remaining subtrees'. The latter set of nodes is precisely what we called in Section 2.2 the *co-path* of node w_i. Intuitively speaking, this construction is forward-secure as each state stores only the minimal information required to compute all succeeding states. In particular, as each node precedes all vertices on its co-path (in the sense of a pre-order visit of the tree), the corresponding key remains secure even if any subsequent epoch's seed is leaked to the adversary.

4.2 Our SSKG construction

We present next the algorithms of our SSKG construction. Particularly interesting, we believe, are the details on how the required pre-order depth-first search is implicitly performed by help of a stack data structure.

Construction 1 (TreeSSKG) *Fix a polynomial* $\ell\colon \mathbb{N} \to \mathbb{N}$ *and a PRG* $G\colon \{0,1\}^\lambda \to \{0,1\}^{2\lambda+\ell(\lambda)}$. *For all* $s \in \{0,1\}^\lambda$ *write* $G(s)$ *as*

$$G(s) = G_L(s) \,\|\, G_R(s) \,\|\, G_K(s) \quad \textit{where} \quad G_L(s), G_R(s) \in \{0,1\}^\lambda \textit{ and } G_K(s) \in \{0,1\}^{\ell(\lambda)} \ .$$

Assuming the notation for stacks from Section 2.3, the algorithms $\mathsf{TreeSSKG} = \{\mathsf{GenSSKG}, \mathsf{Evolve}, \mathsf{GetKey}, \mathsf{Seek}\}$ *of our SSKG are defined by Algorithms 1–4 in Figure 4.*

Algorithm 1: GenSSKG

Input: 1^λ, integer N
Output: initial state st_0

1 **Init**$(\mathcal{S})$
2 $s \leftarrow_R \{0,1\}^\lambda$
3 $h \leftarrow \lceil \log_2(N+1) \rceil$
4 **Push**$(\mathcal{S}, (s,h))$
5 **return** $\mathcal{S}$ *as* st_0

Algorithm 2: GetKey

Input: state st_i as $\mathcal{S}$
Output: key K_i

1 $(s,h) \leftarrow$ **Peek**$_1(\mathcal{S})$
2 $K \leftarrow G_K(s)$
3 **return** K *as* K_i

Algorithm 3: Evolve

Input: state st_i as $\mathcal{S}$
Output: next state st_{i+1}

1 $(s,h) \leftarrow$ **Pop**$(\mathcal{S})$
2 **if** $h > 1$ **then**
3 **Push**$(\mathcal{S}, (G_R(s), h-1))$
4 **Push**$(\mathcal{S}, (G_L(s), h-1))$
5 **return** $\mathcal{S}$ *as* st_{i+1}

Algorithm 4: Seek

Input: state st_0 as $\mathcal{S}$, integer k
Output: state st_k

1 $\delta \leftarrow k$
2 $(s,h) \leftarrow$ **Pop**$(\mathcal{S})$
3 **while** $\delta > 0$ **do**
4 $h \leftarrow h - 1$
5 **if** $\delta < 2^h$ **then**
6 **Push**$(\mathcal{S}, (G_R(s), h))$
7 $s \leftarrow G_L(s)$
8 $\delta \leftarrow \delta - 1$
9 **else**
10 $s \leftarrow G_R(s)$
11 $\delta \leftarrow \delta - 2^h$
12 **Push**$(\mathcal{S}, (s,h))$
13 **return** $\mathcal{S}$ *as* st_k

Fig. 4: Algorithms of TreeSSKG for binary trees. As building blocks we assume a PRG and a stack (cf. Section 2). For the meaning of symbols G_L, G_R, G_K see Construction 1. Observe that the number of supported epochs is potentially greater than N due to the rounding operation in line 3 of GenSSKG.

Let us discuss the algorithms of TreeSSKG in greater detail.

GenSSKG. Besides picking a random seed $s = s_0$ for the root node, Algorithm 1 computes the minimum number $h \in \mathbb{N}$ such that the binary tree of constant height h consists of at least N nodes (cf. Section 2.2). Observe that this tree might have more than N nodes, i.e., more epochs are supported than required. The algorithm stores in state st_0 a stack $\mathcal{S}$ that contains only a single element: the pair (s,h). Here and in the following such pairs should be understood as 'seed s shall generate a subtree of height h'.

Evolve. The stack $\mathcal{S}$ stored in state st_i generally contains two types of information: the top element is a pair (s, h) associated with the current node w_i, and the remaining elements are associated with the corresponding pairs of the nodes on w_i's co-path. After taking the current entry (s, h) off the stack, in order to implement the depth-first search idea from Section 4.1, Algorithm 3 distinguishes two cases: if node w_i is an internal node (i.e., $h > 1$), the update step computes the seeds of its two child nodes using PRG G, starting with the right seed as it needs to be prepended to the current co-path. The new seeds $G_L(s)$ and $G_R(s)$ can be considered roots of subtrees of one level less than w_i; they are hence pushed onto the stack with decreased h-value. In the second case, if the current node w_i is a leaf (i.e., $h = 1$), no further action has to be taken: the next required seed is the 'left-most' node on w_i's co-path, which resides on the stack's top position already.

GetKey. Algorithm 2 is particularly simple as it requires only a single evaluation of PRG G. Observe that the $\mathbf{Peek}_1$ operation leaves its argument unchanged.

Seek. Deriving state st_k from the initial state st_0 via iteratively evoking k times the Evolve procedure is equivalent to visiting all nodes of the tree according to a pre-order traversal until reaching node w_k. However, there is an appealing way to obtain seed s_k more directly, without passing through all the intermediate vertices. The idea is to just walk down the path connecting the root node with w_k. Taking this shortcut decreases the seeking cost to only $\mathcal{O}(\log N)$, as opposed to $\mathcal{O}(N)$. This is the intuition behind the design of our algorithm Seek from Figure 4.

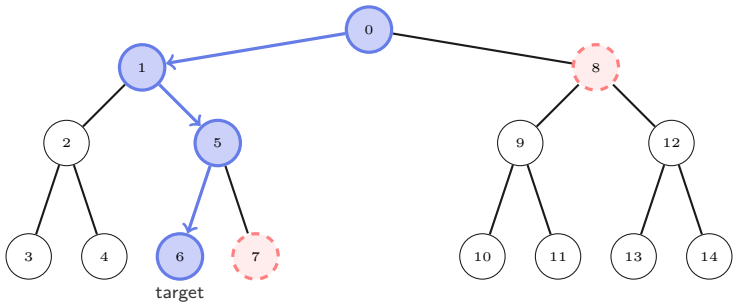

Fig. 5: A visualization of the procedure Seek when computing state st_6. As indicated by the arrows, the algorithm walks down the path from the root node w_0 to the target node w_6 (blue, thick nodes); simultaneously, it records the nodes of w_6's co-path, i.e., (w_7, w_8) (red, dashed nodes).

Recall that Seek is required to output the whole state st_k, and not just seed s_k. In other words, the execution of the algorithm needs to comprehend the construction of the co-path of node w_k. We provide details on how Algorithm 4 fulfills this task. Our strategy, illustrated in Figure 5, is to walk down the path from the root to node w_k, recording the right siblings of the visited nodes on a stack. During this process, with a variable δ we keep track of the remaining number of epochs that needs to be skipped. This counter is particularly helpful for deciding whether, in the path towards w_k, the left or the right child node have to be taken. Indeed, the number of nodes covered by the left and right subtrees is $2^h - 1$ each; if $\delta \leq 2^h - 1$ then the left child is the next to consider, but the right child has to be recorded for the co-path. On the other hand, if $\delta \geq 2^h$, then the left child can be ignored, the co-path doesn't have to be extended, and the walk towards w_k is continued via the right child. The procedure terminates when for the number of remaining epochs we have $\delta = 0$, which means that we arrived at target node w_k.

Security of tree-based SSKGs. We next formally assess the security of Construction 1. For better legibility, in the following theorem we restrict attention to the setting $N = 2^H - 1$, i.e., where $\log(N + 1)$ is an integer; the extension to the general case is straightforward. We will also shorten the notation for some of the concepts from Definitions 2 and 4 (e.g., we denote $\ell(\lambda)$ simply by ℓ, etc.).

Theorem 1 (Security of TreeSSKG). *Assuming a secure PRG is used, our tree-based SSKG from Construction 1 provides* indistinguishability with forward security (IND-FS). *More precisely, for any*

efficient adversary $\mathcal{A}$ against the TreeSSKG *scheme there exist efficient distinguishers $\mathcal{D}_i$ against the underlying PRG such that*

$$\mathrm{Adv}^{\mathsf{IND\text{-}FS}}_{N,\mathcal{A}} \leq 2(N-1) \sum_{i=1}^{\log(N+1)} \mathrm{Adv}^{\mathsf{PRG}}_{\mathcal{D}_i} \ .$$

Proof. The security argument for our scheme reflects the intuition that every SSKG key K_i, for being (part of) the output of a PRG invocation, looks like a random string to any efficient adversary as long as the seed used to compute it remains hidden. In the IND-FS experiment, in addition to state st_m, the adversary gets a challenge K_n^b, which is the real key K_n in case $b = 1$, or is a random string of length ℓ otherwise; here, n and m are adversarially chosen conditioned on $n < m$. The state st_m reveals seed s_m and possibly some subsequent seeds. However, by construction, from these seeds none of the preceding states can be computed. Thus, corrupting state st_m should be of no help to the adversary in distinguishing keys prior to epoch m. In particular, key K_n can be expected to stay secure. We formalize this intuition in the following.

We make use of game hops which progressively transform the IND-FS experiment, denoted here by $\mathcal{G}^{0,b}$, into one for which every adversary has advantage exactly zero. The first hop changes the experiment in that we let the challenger guess the epoch n corresponding to the challenge key K_n^b and abort the simulation if the guess is wrong. More precisely, experiment $\mathcal{G}^{1,b}$ is like $\mathcal{G}^{0,b}$ but, before running $\mathcal{A}_1$, the challenger also chooses a random integer $n^\star$ such that $0 \leq n^\star < N-1$. Then, when $\mathcal{A}_1$ discloses $(state, n, m)$, the challenger checks whether its guess was correct: if it was, i.e., if $n^\star = n$, it proceeds as in $\mathcal{G}^{0,b}$; otherwise, it returns 0 and halts. We obtain

$$\Pr\left[\mathcal{G}^{1,b}_{N,\mathcal{A}}(1^\lambda) = 1\right] = \Pr\left[n^\star = n \ \wedge \ \mathcal{G}^{0,b}_{N,\mathcal{A}}(1^\lambda) = 1\right] = \frac{1}{N-1} \cdot \Pr\left[\mathrm{Expt}^{\mathsf{IND\text{-}FS},b}_{N,\mathcal{A}}(1^\lambda) = 1\right] \ ,$$

or, equivalently,

$$\mathrm{Adv}^{\mathsf{IND\text{-}FS}}_{N,\mathcal{A}}(\lambda) = (N-1) \cdot \left|\Pr\left[\mathcal{G}^{1,1}_{N,\mathcal{A}}(1^\lambda) = 1\right] - \Pr\left[\mathcal{G}^{1,0}_{N,\mathcal{A}}(1^\lambda) = 1\right]\right| \ . \tag{1}$$

Regarding the analysis of game $\mathcal{G}^{1,b}$, observe that key K_n^1 is not just computed as the output of a PRG on input a random seed; rather, it is computed by iterating a PRG up to $\log(N+1)$ times, where only the first input (seed s_0) is truly random. We hence proceed via a hybrid argument, by considering intermediate experiments $\mathcal{H}^{0,b}, \mathcal{H}^{1,b}, \ldots, \mathcal{H}^{L,b}$, where we set $\mathcal{H}^{0,b} = \mathcal{G}^{1,b}$ and let $L \leq \log(N+1)$ denote the level of challenge node w_n in the tree. The idea is to let each transition $\mathcal{H}^{i-1,b} \to \mathcal{H}^{i,b}$ replace the output of the PRG associated with the i-th node on the path from the root node w_0 to node w_n by a random value. Then, in $\mathcal{H}^{L,b}$ the challenge key K_n^b will be random independently of bit b: in case $b = 1$ due to the argument just given, and in case $b = 0$ by the definition of $\mathrm{Expt}^{\mathsf{IND\text{-}FS},0}$. In particular, every adversary $\mathcal{A}$ playing in $\mathcal{H}^{L,b}$ will have zero advantage.

More precisely, let L be the level of node w_n and let $(v_1, \ldots, v_L)$ denote the path from the root $v_1 = w_0$ to node $v_L = w_n$. For every $i = 1, \ldots, L$, derive $\mathcal{H}^{i,b}$ from $\mathcal{H}^{i-1,b}$ by replacing the output of $G(s_k)$ by a random string in $\{0,1\}^{2\lambda+\ell}$, where k is the epoch number corresponding to node v_i.

Observe that, except for $\mathcal{H}^{0,b}$, as we follow the path top-to-bottom, seed s_k was replaced by a random value in the hybrid before, i.e., in $\mathcal{H}^{i-1,b}$. By consequence, for every $\mathcal{A}$ there exists a distinguisher $\mathcal{D}_i$ that allows bounding the difference between $\mathcal{H}^{i-1,b}$ and $\mathcal{H}^{i,b}$ as follows:

$$\left|\Pr\left[\mathcal{H}^{i-1,b}_{N,\mathcal{A}}(1^\lambda) = 1\right] - \Pr\left[\mathcal{H}^{i,b}_{N,\mathcal{A}}(1^\lambda) = 1\right]\right| \leq \mathrm{Adv}^{\mathsf{PRG}}_{\mathcal{D}_i}(\lambda) \ . \tag{2}$$

As already stated, the challenge key K_n^b in hybrid $\mathcal{H}^{L,b}$ is uniformly random, independently of bit b. In other words, $\mathcal{H}^{L,0}$ and $\mathcal{H}^{L,1}$ are the very same experiment, and we have, even for unbounded distinguishers,

$$\left|\Pr\left[\mathcal{H}^{L,1}_{N,\mathcal{A}}(1^\lambda) = 1\right] - \Pr\left[\mathcal{H}^{L,0}_{N,\mathcal{A}}(1^\lambda) = 1\right]\right| = 0 \ . \tag{3}$$

Using an induction argument and the triangle inequality, we can combine (2) and (3) into

$$\left|\Pr\left[\mathcal{G}^{1,1}_{N,\mathcal{A}}(1^\lambda) = 1\right] - \Pr\left[\mathcal{G}^{1,0}_{N,\mathcal{A}}(1^\lambda) = 1\right]\right| \leq 2 \sum_{i=1}^{L} \mathrm{Adv}^{\mathsf{PRG}}_{\mathcal{D}_i}(\lambda) \ . \tag{4}$$

By combining equations (1) and (4) we obtain the bound indicated in the theorem statement. □

4.3 Extending our scheme towards d-ary trees

We show how to extend our binary-tree-based construction towards the general case of d-ary trees, for arbitrary $d \geq 2$. Recall that a d-ary tree of constant height H has $\nu_d(H) = (d^H - 1)/(d-1)$ nodes. The intuition behind our design is mainly the same as in the binary case: we consider an enumeration $w_0, \ldots, w_{N-1}$ of the tree's nodes according to a pre-order depth-first search (first visit the root, then, from left to right, recursively the subtrees) and associate with every node w_i a seed s_i from which epoch's key K_i and, where applicable, the seeds of its children are derived using a PRG. It is clear that this PRG must have a larger expansion than in Section 4.2: every PRG invocation has to yield one (sub)string of length $\ell(\lambda)$ for the key, and d-many (sub)strings of length λ for the subordinate seeds. We construct the corresponding algorithms as follows.

Construction 2 (TreeSSKG for d-ary trees) *Fix a polynomial $\ell\colon \mathbb{N} \to \mathbb{N}$, an integer $d \in \mathbb{N}^{\geq 2}$, and a PRG $G\colon \{0,1\}^\lambda \to \{0,1\}^{d\lambda+\ell(\lambda)}$. For all $s \in \{0,1\}^\lambda$ write $G(s)$ as*

$$G(s) = G_1(s) \,\|\, \ldots \,\|\, G_d(s) \,\|\, G_K(s) \quad \textit{where} \quad \forall j\colon G_j(s) \in \{0,1\}^\lambda \textit{ and } G_K(s) \in \{0,1\}^{\ell(\lambda)} \ .$$

Assuming the notation for stacks from Section 2.3, our SSKG based on d-ary trees $\mathsf{TreeSSKG}_d = \{\mathsf{GenSSKG}_d, \mathsf{Evolve}_d, \mathsf{GetKey}, \mathsf{Seek}_d\}$ is defined by Algorithm 2 in Figure 4, and Algorithms 5, 6, and 7 in Figure 6.

Algorithm 5: $\mathsf{GenSSKG}_d$

Input: 1^λ, integer N
Output: initial state st_0

1 **Init**$(\mathcal{S})$
2 $s \leftarrow_R \{0,1\}^\lambda$
3 $h \leftarrow \lceil \log_d(N+1) \rceil$
4 **Push**$(\mathcal{S}, (s,h))$
5 **return** $\mathcal{S}$ *as* st_0

Algorithm 6: Evolve_d

Input: state st_i as $\mathcal{S}$
Output: next state st_{i+1}

1 $(s,h) \leftarrow$ **Pop**$(\mathcal{S})$
2 **if** $h > 1$ **then**
3 **for** $j = d$ **downto** 1 **do**
4 **Push**$(\mathcal{S}, (G_j(s), h-1))$
5 **return** $\mathcal{S}$ *as* st_{i+1}

Algorithm 7: Seek_d

Input: state st_0 as $\mathcal{S}$, integer k
Output: state st_k

1 $\delta \leftarrow k$
2 $(s,h) \leftarrow$ **Pop**$(\mathcal{S})$
3 **while** $\delta > 0$ **do**
4 $h \leftarrow h - 1$
5 $\nu \leftarrow \nu_d(h)$
6 $c \leftarrow \lfloor (\delta - 1)/\nu \rfloor + 1$
7 **for** $j = d$ **downto** $c+1$ **do**
8 **Push**$(\mathcal{S}, (G_j(s), h))$
9 $s \leftarrow G_c(s)$
10 $\delta \leftarrow \delta - (1 + (c-1)\nu)$
11 **Push**$(\mathcal{S}, (s,h))$
12 **return** $\mathcal{S}$ *as* st_k

Fig. 6: Algorithms of $\mathsf{TreeSSKG}_d$ (for d-ary trees). As building blocks we assume a PRG and a stack (cf. Section 2). For the meaning of symbol G_j see Construction 2. Observe that the number of supported epochs is potentially greater than N due to the rounding operation in line 3 of $\mathsf{GenSSKG}_d$.

Observe that the proposed GetKey algorithm is identical with the one from the binary case, and that the only modification in $\mathsf{GenSSKG}_d$ is the basis to which the logarithm is taken for computing h. More interesting is the new Evolve_d algorithm. Here, whenever an internal node needs to be expanded, all d direct successor seeds are computed; the left-most seed will be associated with the next state (i.e., st_{i+1}), and its right siblings become part of the new co-path, i.e., are pushed in the correct order onto the stack. We point out two differences in the Seek_d algorithm. Firstly, in line 6, with c we calculate the number $1 \leq c \leq d$ of current node's child that is on the path from the root to target node w_k (consequently, only the siblings with numbers $c+1, \ldots, d$ need to be recorded for the co-path). Secondly, in line 10, the number δ of epochs that still need to be skipped is decreased by $(c-1)\nu$ in one shot, where ν indicates the number of nodes of the subtree generated by the respectively considered seed s.

We give the following security statement for our generalized construction. The proof is essentially identical with the one for Theorem 1.

Theorem 2 (Security of $\mathsf{TreeSSKG}_d$). *Assuming a secure PRG is used, our generalized tree-based SSKG from Construction 2 provides* IND-FS *security. More precisely, for any efficient adversary $\mathcal{A}$ against the $\mathsf{TreeSSKG}_d$ scheme there exist efficient distinguishers $\mathcal{D}_i$ against the underlying PRG such that*

$$\mathrm{Adv}^{\mathsf{IND\text{-}FS}}_{N,\mathcal{A}} \leq 2(N-1) \sum_{i=1}^{\log_d(N+1)} \mathrm{Adv}^{\mathsf{PRG}}_{\mathcal{D}_i} \ .$$

Remark 3 (Degenerate trees). In the above definitions we insisted on fixing d such that $d \geq 2$. This choice guarantees that the time it takes to seek to an arbitrary target node is $\mathcal{O}(\log N)$. Observe however, that also for $d = 1$ our scheme is correct and secure, but falls back to linear seeking time. It is interesting to observe that this degenerate case corresponds exactly with the hash chain approach of early SKGs (e.g., from [11,16,4]).

4.4 An enhanced seeking procedure

As required by Definition 2, our Seek algorithm allows computing any state st_k given the initial state st_0. Observe, however, that in many applications this initial state might not be accessible; indeed, forward security can be attained only if states of expired epochs are securely erased. From a practical perspective it is hence appealing to generalize the functionality of Seek to allow efficient computation of st_{i+k} from any state st_i, and not just from st_0.[3] We correspondingly extend the notion of SSKG by introducing a new algorithm, SuperSeek, which realizes the Evolve^k functionality for arbitrary starting points; when invoked on input st_0, the new procedure behaves exactly as Seek.

Definition 5 (SSKG with SuperSeek). *A seekable sequential key generator* SSKG *supports* SuperSeek *if it has an auxiliary algorithm as follows:*

– SuperSeek. *On input of a state st_i and $k \in \mathbb{N}$, this deterministic algorithm returns state st_{i+k}.*

For correctness we require that for all $N \in \mathbb{N}$, all $\mathsf{st}_0 \leftarrow_R \mathsf{GenSSKG}(1^\lambda, N)$, all $i, k \in \mathbb{N}$, and $\mathsf{st}_i = \mathsf{Evolve}^i(\mathsf{st}_0)$ we have

$$0 \leq i \leq i + k < N \qquad \Longrightarrow \qquad \mathsf{Evolve}^k(\mathsf{st}_i) = \mathsf{SuperSeek}(\mathsf{st}_i, k) \ .$$

Assume a TreeSSKG instance is in state st_i and an application requests it to seek to position st_{i+k}, for arbitrary $0 \leq i \leq i + k < N$. Recall from the discussions in Sections 4.1 and 4.2 that state st_i encodes both the seed s_i and the co-path of node w_i. Recall also that, as a property of the employed pre-order visit of the tree, for each state st_j, $j > i$, the co-path of node w_i contains an ancestor w of w_j. Following these observations, our SuperSeek construction consists of two consecutive phases. For seeking to state st_{i+k}, in the first phase the algorithm considers all nodes on the co-path of w_i until it finds the ancestor w of w_{i+k}. The second phase is then a descent from that node to node w_{i+k}, similarly to what we had in the regular Seek algorithms. In both phases care has to be taken that the co-path of target node w_{i+k} is correctly assembled as part of st_{i+k}. The working principle of our new seeking method is also illustrated in Figure 7.

We present explicit instructions for implementing SuperSeek in Figure 8, separately for the binary and d-ary case. In both cases the first while loop identifies the ancestor w of target node w_{i+k} on w_i's co-path by comparing δ (i.e., the remaining number of epochs to be skipped) with the number of nodes $\nu(h)$ in the subtree where w is the root. The second loop is equivalent to the one from Algorithms 4 and 7.

[3] Our reason to require in Definition 2 that st_0 be input to the Seek algorithm was to maintain consistency with prior work [15], where generalized seekability was not possible for technical reasons (cf. Remark 2).

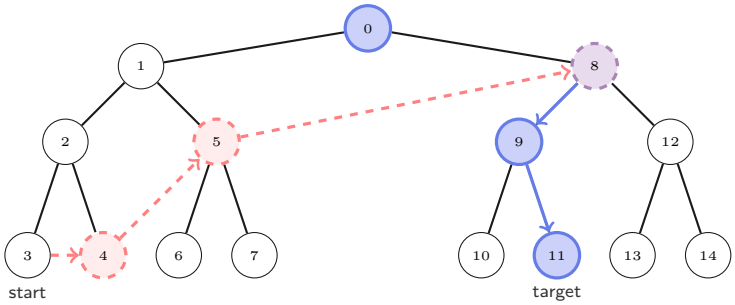

Fig. 7: A visualization of the procedure SuperSeek jumping from epoch 3 to 11. As indicated by the arrows, the algorithm first finds the intersection, here w_8, between the co-path of node w_3 (red, dashed nodes) and the path that connects the root with the target node w_{11} (blue, thick nodes); from there it proceeds downwards until it reaches target node w_{11}.

Algorithm 8: SuperSeek

Input: state st_i as $\mathcal{S}$, integer k
Output: state st_{i+k}

$\delta \leftarrow k$
$(s, h) \leftarrow \mathbf{Pop}(\mathcal{S})$
while $\delta \geq 2^h - 1$ **do**
  $\delta \leftarrow \delta - (2^h - 1)$
  $(s, h) \leftarrow \mathbf{Pop}(\mathcal{S})$
while $\delta > 0$ **do**
  $h \leftarrow h - 1$
  if $\delta < 2^h$ **then**
    $\mathbf{Push}(\mathcal{S}, (G_R(s), h))$
    $s \leftarrow G_L(s)$
    $\delta \leftarrow \delta - 1$
  else
    $s \leftarrow G_R(s)$
    $\delta \leftarrow \delta - 2^h$
$\mathbf{Push}(\mathcal{S}, (s, h))$
return $\mathcal{S}$ *as* st_{i+k}

Algorithm 9: $\mathsf{SuperSeek}_d$

Input: state st_i as $\mathcal{S}$, integer k
Output: state st_{i+k}

$\delta \leftarrow k$
$(s, h) \leftarrow \mathbf{Pop}(\mathcal{S})$
$\nu \leftarrow \nu_d(h)$
while $\delta \geq \nu$ **do**
  $\delta \leftarrow \delta - \nu$
  $(s, h) \leftarrow \mathbf{Pop}(\mathcal{S})$
  $\nu \leftarrow \nu_d(h)$
while $\delta > 0$ **do**
  $h \leftarrow h - 1$
  $\nu \leftarrow \nu_d(h)$
  $c \leftarrow \lfloor (\delta - 1)/\nu \rfloor + 1$
  for $j = d$ **downto** $c + 1$ **do**
    $\mathbf{Push}(\mathcal{S}, (G_j(s), h))$
  $s \leftarrow G_c(s)$
  $\delta \leftarrow \delta - (1 + (c - 1)\nu)$
$\mathbf{Push}(\mathcal{S}, (s, h))$
return $\mathcal{S}$ *as* st_{i+k}

Fig. 8: Algorithm SuperSeek for binary and d-ary trees.

4.5 Evolving beyond the supported number of epochs

We showed that our TreeSSKG scheme (with or without the modifications from Sections 4.3 and 4.4) is secure and correct. Note, however, that the latter property only holds for up to N epochs—a number that currently needs to be fixed once and forever when initializing the SSKG. As deployed systems often run much longer than anticipated, practical implementations might require built-in precautions for the case that, after all, more than N epochs are required, i.e., that the SSKG state needs to be evolved beyond the well-defined range.

To achieve a certain level of graceful degradation, we propose, possibly as an amendment to Remark 1, two ways to deal with the a posteriori increase of the number of required epochs. Firstly, the algorithms from Figures 4, 6, and 8 could be adapted such that when the bound N is reached they switch to computing and outputting keys according to a simple hash chain (or, even better, a PRG chain; see also Section 1). This modification would yield a sound SSKG with an unlimited number of supported epochs—however with only linear seeking time for the retroactively added epochs. Alternatively, in fact as a generalization of the first approach, the key K_{N-1} computed for the last 'regular' epoch could be used as initial seed for an otherwise independent TreeSSKG instance. Note that this would imply a

somewhat non-uniform average seeking time, but the latter would nevertheless remain in the logarithmic domain.

5 Practical aspects

In the preceding sections we left open how PRGs can be instantiated in practice; indeed, the well-known recommendations and standards related to symmetric key cryptography exclusively consider block ciphers, stream ciphers, and hash functions. Fortunately, secure PRG instantiations can be boot-strapped from all three named primitives. For instance, a block cipher operated in counter mode can be seen as a PRG where the block cipher's key acts as the PRG's seed[4]. Similar counter-based constructions derived from hash functions or PRFs (e.g., HMAC) are possible.

A specific property of PRGs that are constructed by combining a symmetric primitive with a counter is particularly advantageous for efficiently implementing our TreeSSKG schemes. Recall that the PRGs used in Constructions 1 and 2 are effectively evaluated in a blockwise fashion. More precisely, while the PRGs are formally defined to output strings of length $d\lambda + \ell(\lambda)$, in our TreeSSKG algorithms it is sufficient to compute only a considerably shorter substring per invocation. This property is perfectly matched by the 'iterated PRGs' proposed above, as the latter allow exactly this kind of evaluation very efficiently.

Implementational details. We implemented our TreeSSKG scheme both in the setting of binary trees and of d-ary trees. In particular we have C code for Algorithms 1–9, and we claim that the level of optimization is sufficient for practical deployment. Our code relies on the OpenSSL library [21] for random number generation and the required cryptographic primitives. We consider a total of four PRG instantiations, using the AES128 and AES256 block ciphers and the MD5[5] and SHA256 hash functions as described. That is, we have two instantiations at the $\lambda = 128$ security level, and two at the $\lambda = 256$ level.

We experimentally evaluated the performance of our implementation, using the following setup. We generated SSKG instances with parameters $(d, H) = (2, 20)$, i.e., instances that support $N = 2^{20} - 1 \approx 10^6$ epochs. We iterated through all epochs in linear order, determining both the average and the worst-case time consumed by the Evolve algorithm. Similarly we measured the average and worst-case time it takes for the Seek algorithm to recover states st_k, ranging over all values $k \in [0, N-1]$. Concerning SuperSeek, we picked random pairs $i, j \in [0, N-1]$, $i < j$, and measured the time required by the algorithm to jump from st_i to st_j. Finally, we performed analogous measurements for GenSSKG and GetKey (here, average and worst-case coincide). The results of our analysis are summarized in Table 1.

For comparison we also include the corresponding timing values of our competitor, the (factoring-based) SSKG from [15][6], for security levels roughly equivalent to ours. We point out that the analogue of GenSSKG from [15] in fact consists of two separate algorithms: one that produces public parameters and an associated 'seeking key', and one that generates the actual initial SSKG state. As any fixed combination of public parameters and corresponding seeking key can be used for many SSKG instances without security compromises, for fairness we decided not to count the generation costs of the former when indicating the GenSSKG performance in Table 1. Instead, we report the results of our timing analysis here as follows: for the costs of parameters and seeking key generation with 2048 bit and 3072 bit RSA moduli we measured $400ms$ and $2300ms$, respectively.

It might be instructive to also study the required state sizes for both our TreeSSKG scheme and the scheme from [15]. In our implementation, for fixed parameter d, the (maximum) state size scales roughly linearly in both H and the seed length of the used PRG. Concretely, for $(d, H) = (2, 20)$ and 128 bit keys (e.g., for AES128- and MD5-based PRGs) the state requires 350 bytes, while for 256 bit security

[4] Observe that in this construction only very few enciphering operations need to be performed per key. By consequence, block ciphers with high key agility seem to be the preferable choice over ciphers with expensive key setup, as the latter cannot be amortized over time.

[5] We considered MD5 in our tests because it is fast and ubiquitously available; the well-known attacks against its collision resistance do no seem to affect the hash function's strength in the PRG context. Nevertheless, our choice was made for reasons of comparability and should not be misunderstood as a suggestion to use an MD5-based PRG in practice.

[6] The reference implementation from [15] can be found at `http://cgit.freedesktop.org/systemd/systemd/tree/src/journal/fsprg.c`.

a total of 670 bytes of storage are necessary. In the scheme from [15] the space in the state variable is taken by an RSA modulus N, a value $x \in \mathbb{Z}_N^\times$, a 64 bit epoch counter, and a small header. Precisely, for 2048 and 3072 bit RSA moduli this results in 522 and 778 bytes of state, respectively.

	AES128		**MD5**		**[15]/2048 bit**	**AES256**		**SHA256**		**[15]/3072 bit**
	[average]	[max]	[average]	[max]		[average]	[max]	[average]	[max]	
GenSSKG	22μs		22μs		27μs	22μs		22μs		38μs
Evolve	0.2μs	0.5μs	0.2μs	0.4μs	8μs	0.5μs	1μs	0.4μs	0.8μs	13μs
Seek	7μs	9μs	6μs	7μs	4.9ms	14μs	18μs	11μs	15μs	12.6ms
SuperSeek	6μs	9μs	5μs	7μs	–	13μs	18μs	8μs	15μs	–
GetKey	0.2μs		0.2μs		12μs	0.4μs		0.4μs		13μs

Table 1: Results of efficiency measurements of our TreeSSKG algorithms when instantiated with different PRGs, and a comparison with the algorithms from [15]. All experiments were performed on an Intel Core i7-3517U CPU clocked at 1.90GHz. We used OpenSSL version 0.9.8 for the implementation of our TreeSSKG routines, while for the compilation of the reference code from [15] we used the gcrypt library in version 1.5.0.

Results and discussion. We discuss the results from Table 1 as follows, beginning with those of our tree-based SSKG (i.e., columns AES128, MD5, AES256, and SHA256). Our first observation is that the GenSSKG time is independent of the respectively used PRG. This is not surprising as the former algorithm never invokes the latter, but spends its time with memory allocation and requesting the random starting seed from OpenSSL's core routines. The timings for Evolve indicate that, as expected, 128-bit cryptographic primitives are faster than 256-bit primitives, and that for a fixed security level the hash-function-based constructions are (slightly) preferable. The hypothesis that the time spent by the individual algorithms is dominated by the internal PRG executions is supported by the observation that the running time of Evolve (on average) and GetKey coincide, and that the worst-case running time of Evolve is twice that value; to see this, recall that Evolve executions perform either two internal PRG invocations or none, and that the average number of invocations is one. We understand that the SuperSeek timings are generally better than the Seek values as the first **while** loop in Algorithm 8 does not comprise a PRG invocation, whereas the second **while** loop requires less iterations on average than the corresponding loop in Algorithm 4.

The routines from [15] are clearly outperformed by the ones from our SSKG. Firstly, for the Evolve algorithm our timing values are about 30 times better than those for [15] (recall that the latter's state update involves a modular squaring operation). Similar results show our tree-based GetKey algorithm to be faster, by a factor between 30 and 60, depending on the considered security level. This might be surprising at first sight, as the algorithm from [15] consists of just hashing the corresponding state variable, but presumably the explication for this difference is that [15] operates with considerably larger state sizes than we do. Finally, the superiority of our tree-based construction in terms of efficiency is made even more evident by studying the performance of the seek Seek algorithms, where we can report our routines to be 700–1000 times faster than those from [15], again depending on the security level.

Conclusion

The recently introduced concept of seekable sequential key generator (SSKG) combines the forward-secure generation of sequences of cryptographic keys with an explicit fast-forward functionality. While prior constructions of this primitive require specific number-theoretic building blocks, we show that symmetric tools like block ciphers or hash functions suffice for obtaining secure SSKGs; this leads to impressive performance improvements in practice, by factors of 30–1000, depending on the considered algorithms. In addition to the performance gain, our scheme enhances the functionality of SSKGs by generalizing the notion of seekability, making it more natural and concise, an improvement that we believe is very relevant for applications. Our scheme enjoys provable security in the standard model.

Acknowledgments

The authors thank all anonymous reviewers for their valuable comments. Giorgia Azzurra Marson was supported by CASED and Bertram Poettering by EPSRC Leadership Fellowship EP/H005455/1.

References

1. R. Accorsi. BBox: A distributed secure log architecture. In J. Camenisch and C. Lambrinoudakis, editors, *EuroPKI*, volume 6711 of *Lecture Notes in Computer Science*, pages 109–124. Springer, 2010.
2. M. Bellare and S. K. Miner. A forward-secure digital signature scheme. In M. J. Wiener, editor, *CRYPTO'99*, volume 1666 of *LNCS*, pages 431–448, Santa Barbara, CA, USA, Aug. 15–19, 1999. Springer, Berlin, Germany.
3. M. Bellare and B. S. Yee. Forward integrity for secure audit logs. Technical report, 1997.
4. M. Bellare and B. S. Yee. Forward-security in private-key cryptography. In M. Joye, editor, *CT-RSA 2003*, volume 2612 of *LNCS*, pages 1–18, San Francisco, CA, USA, Apr. 13–17, 2003. Springer, Berlin, Germany.
5. R. Canetti, S. Halevi, and J. Katz. A forward-secure public-key encryption scheme. *Journal of Cryptology*, 20(3):265–294, July 2007.
6. R. Canetti and H. Krawczyk. Analysis of key-exchange protocols and their use for building secure channels. In B. Pfitzmann, editor, *EUROCRYPT 2001*, volume 2045 of *LNCS*, pages 453–474, Innsbruck, Austria, May 6–10, 2001. Springer, Berlin, Germany.
7. C. N. Chong, Z. Peng, and P. H. Hartel. Secure audit logging with tamper-resistant hardware. In D. Gritzalis, S. D. C. di Vimercati, P. Samarati, and S. K. Katsikas, editors, *SEC*, volume 250 of *IFIP Conference Proceedings*, pages 73–84. Kluwer, 2003.
8. P. Gutmann. Secure deletion of data from magnetic and solid-state memory. In *Proceedings of the Sixth USENIX Security Symposium, San Jose, CA*, volume 14, 1996.
9. J. E. Holt. Logcrypt: forward security and public verification for secure audit logs. In R. Buyya, T. Ma, R. Safavi-Naini, C. Steketee, and W. Susilo, editors, *ACSW Frontiers*, volume 54 of *CRPIT*, pages 203–211. Australian Computer Society, 2006.
10. J. Kelsey, J. Callas, and A. Clemm. Signed Syslog Messages. RFC 5848 (Proposed Standard), May 2010.
11. J. Kelsey and B. Schneier. Cryptographic support for secure logs on untrusted machines. In *Proceedings of the 7th USENIX Security Symposium*, 1998.
12. J. Kelsey and B. Schneier. Minimizing bandwidth for remote access to cryptographically protected audit logs. In *Recent Advances in Intrusion Detection*, 1999.
13. D. Ma and G. Tsudik. Extended abstract: Forward-secure sequential aggregate authentication. In *2007 IEEE Symposium on Security and Privacy*, pages 86–91, Oakland, California, USA, May 20–23, 2007. IEEE Computer Society Press.
14. D. Ma and G. Tsudik. A new approach to secure logging. *Trans. Storage*, 5(1):2:1–2:21, Mar. 2009.
15. G. A. Marson and B. Poettering. Practical secure logging: Seekable sequential key generators. In J. Crampton, S. Jajodia, and K. Mayes, editors, *ESORICS 2013*, volume 8134 of *LNCS*, pages 111–128, Egham, UK, Sept. 9–13, 2013. Springer, Berlin, Germany.
16. B. Schneier and J. Kelsey. Secure audit logs to support computer forensics. *ACM Trans. Inf. Syst. Secur.*, 2(2):159–176, 1999.
17. V. Shoup. On formal models for secure key exchange. Technical Report RZ 3120, IBM, 1999.
18. V. Stathopoulos, P. Kotzanikolaou, and E. Magkos. A framework for secure and verifiable logging in public communication networks. In J. López, editor, *CRITIS*, volume 4347 of *Lecture Notes in Computer Science*, pages 273–284. Springer, 2006.
19. A. A. Yavuz and P. Ning. BAF: An efficient publicly verifiable secure audit logging scheme for distributed systems. In *ACSAC*, pages 219–228. IEEE Computer Society, 2009.
20. A. A. Yavuz, P. Ning, and M. K. Reiter. BAF and FI-BAF: Efficient and publicly verifiable cryptographic schemes for secure logging in resource-constrained systems. *ACM Trans. Inf. Syst. Secur.*, 15(2):9, 2012.
21. E. Young and T. Hudson. OpenSSL: The Open Source Toolkit for SSL/TLS. http://www.openssl.org.