A probabilistic small model theorem to assess confidentiality of dispersed cloud storage (extended abstract)

Marco Baldi[1], Ezio Bartocci[2], Franco Chiaraluce[1], Alessandro Cucchiarelli[1], Linda Senigagliesi[1], Luca Spalazzi[1], and Francesco Spegni[1]

[1] Università Politecnica delle Marche, Ancona, Italy
[2] TU Wien, Vienna, Austria

Abstract. Recent developments in cloud architectures and security concerns have originated new models of online storage clouds based on data dispersal algorithms. According to these algorithms the data is divided into several slices that are distributed among remote and independent storage nodes. Ensuring confidentiality in this context is crucial: only legitimate users should access any part of information they distribute among storage nodes.
We use *parameterized* Markov Decision Processes to model such a class of systems and Probabilistic Model Checking to assess the likelihood of breaking the confidentiality. We showed that a Small Model Theorem can be proven for a specific types of models, preserving $PCTL^\star$ formulae. Finally, we report the result of applying our methodology to feasibly assess the security of existing dispersed cloud storage solutions.

1 Introduction

New models of cloud storage protocols are emerging, that are based on data dispersal algorithms, where data is divided into several slices distributed among remote and independent storage node [4,5,6]. The main advantage of these techniques consists in their reliability since the dispersion is usually accompanied by redundancy. However, in this context, ensuring confidentiality is equally important: only legitimate users should have access to any part of the information that was dispersed among the independent storage nodes.

We developed an *assessment methodology for dispersed storage clouds* that can describe real-world implementations and the presence of a passive eavesdropper trying to spill the slices and reconstruct the users secret.

The proposed methodology relies on *parameterized Markov Decision Processes* to model such a class of systems and *probabilistic model checking* as a verification tool. Since many parameters contribute to the description of the system, we addressed it applying common bounded model checking techniques. For certain classes of models we were able to prove a small-model theorem, implying that certain security specifications hold irrespective of the actual number of storage providers.

2 The AONT-based dispersed storage clouds

In *dispersed storage clouds* any user has some amount of cloud storage space assigned on independent storage nodes. In order to assure reliability on the one hand and security on the other, several authors have proposed schemata based on fragmentation, erasure coding, and encryption [4,5].

From a purely abstract point of view, all these algorithms can be characterized by a set of parameters. Let x be the original file size (measured in bits) to be dispersed and let $l \cdot q$ be the size of each fragment called *slice* (when $q = 8$, l is the size of a slice in bytes). The parameters of interest are n and k, the former being the number of slices after the transformation of the original file, and the latter the minimum number of slices to recover the original file (i.e. $n - k$ is the maximum number of lost (erased) slices that still enables file recover).

In Fig. 1 we depicted AONT-RS [5], a popular schemata consisting in applying the Reed-Solomon erasure code (*Encoder*) to the All-or-Nothing-Transform (*AONT* and *Slicer*). The produced fragments are finally dispatched among the providers (*Dispatcher*).

Attacker models. To the aim of assessing the security of dispersed storage clouds, we introduced a probabilistic attacker model. Assuming the user is connected to the Internet from its (wired or wireless) local area network (LAN), a passive eavesdropper on the user LAN has some chances to intercept the exchanged slices. We have considered mainly two types of attackers: in one case the attacker can intercept each exchanged slice independently from the others (*slice attacker* [1,3]), while in the other he/she can choose to attack the storage providers trying to steal all the slices stored there (*provider attacker* [3]).

In the case of a slice attacker on a wireless LAN, one can use a *frequentistic approach* and rely on the physical properties of the channel to determine the likelihood of successfully intercepting a slice (as we have shown in [1]). In the case of a provider attacker, one can use a more *subjective approach* to determine the likelihood of breaking a storage provider. In both cases such likelihoods are required as input parameters of the system models.

3 Assessment methodology

Now we briefly describe how the user, links to storage nodes and attacker are modeled using MDPs.

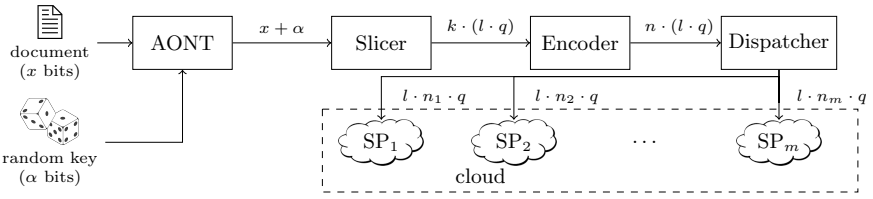

Fig. 1: Block diagram of AONT-RS (with data length expressed in bits)

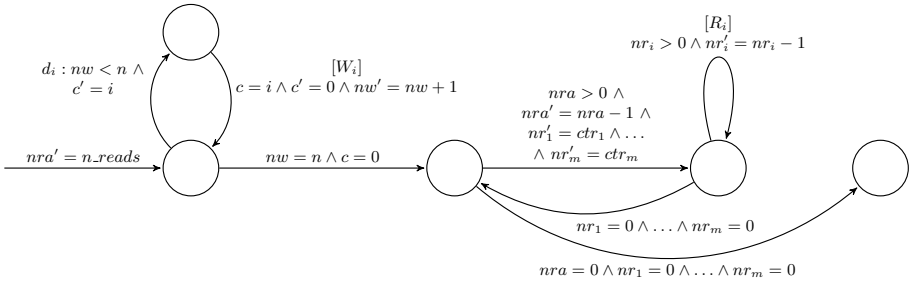

Fig. 2: The $\textsc{User}(d_1, \ldots, d_m)$

3.1 Modeling

In Figs. 2, 3 and 4 we represent the relevant MDPs for the case of a system with a slice attacker. We use straight variable names for edge pre-conditions, while primed variable names are considered edge post-conditions.

First, the user loops until all slices are dispatched to the storage providers (i.e. $nw = n$). Next, he/she loops until n_reads attempts have been consumed (i.e. $nra = 0$). The link to the storage provider i either receives a written slice $[W_i]$ or sends a slice to be read $[R_i]$. In both cases there is some chance a_i that the slice is leaked to the attacker. The latter increments a counter of stolen slices ctr_a for each leak event $[L_i]$.

Security assessment analysis. It is clear that assessing the security of such systems is a parametric problem. Indeed, by allowing an arbitrarily large number of read operations by the user, the attacker has probability 1 of intercepting more than k slices (every read the attacker has one more chance of intercepting the missing slices, until it intercepts all of them). Similarly, assuming the secret is split into an arbitrarily large number of slices gives the attacker a negligible probability of succeeding in his/her attack. Between these ends lie all the parameters values of the actual implementations of AONT-based algorithms. Very often such values are not bound to a clearly stated security metric. We employ bounded and probabilistic model checking to compute the likelihood of a successful attack for several parameter configurations. The obtinaed probabilities

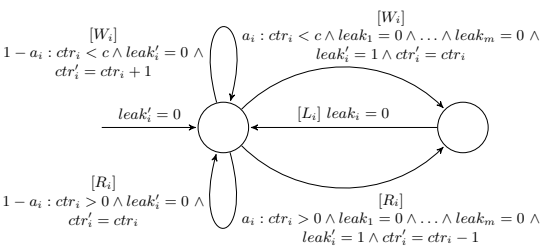

Fig. 3: The $\textsc{Link}_i^c(a_i)$

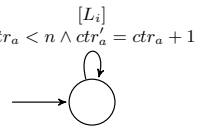

Fig. 4: The $\textsc{Attacker}$

are collected in a graph relating the probability of a successful attack with the parameter values. For m storage providers, the checked system is:

$$\mathcal{M}_m^{\text{AONT}} := \textsc{User}(d_1, \ldots, d_m) \parallel \textsc{Link}_1^n(a_1) \parallel \ldots \parallel \textsc{Link}_m^n(a_m) \parallel \textsc{Attacker}.$$

Finally, the probabilistic model checker is repeatedly invoked to solve the following problem varying the parameter values: $\mathcal{P}_{max}(F(ctr_a \geq k), \mathcal{M}_m^{\text{AONT}})$.

Small model theorem for node links. A small model theorem allows to verify a class of infinite state systems by only checking a finite size system. The key observation is that, in a system where slices are intercepted when traveling between user and storage nodes, two or more node links with the same attack probability are indistinguishable from a single node link having the same attack probability, modulo some technicalities. Since the attack probability is determined by the physical properties of the employed LAN, the number of required node links reduces to the number of used LANs.

We denote with $\textsc{Link}_i^c(a)$ a generic link to storage provider i, having *capacity* of c slices and probability a of leaking a slice to the slice attacker, during reads or writes. We write $\approx_\Gamma$ denoting the (equivalence) relation of *probabilistic bisimulation modulo action replacement* [1]. Then we can state the following.

Lemma 1 (Reduction [1]). *For any natural numbers $c, d, i, j, k > 0$ such that $i \neq j$, any probability a. Given the MDPs $\textsc{Link}_i^c(a)$, $\textsc{Link}_j^d(a)$, and $\textsc{Link}_k^{c+d}(a)$:*

$$\textsc{Link}_i^c(a) \parallel \textsc{Link}_j^d(a) \approx_\Gamma \textsc{Link}_k^{c+d}(a)$$

where Γ renames R_i, R_j (resp. W_i, W_j, resp. L_i, L_j) to R_k (resp. W_k, resp. L_k).

Given a sorted list of numbers $a_1, \ldots, a_m$ s.t. $a_1 \leq \ldots \leq a_m$, let us call its *distinction* the list of indices $i_1, \ldots, i_{q+1}$ satisfying the following:

- $i_1 = 1$, $i_{q+1} = m$, and $i_1 < \ldots < i_{q+1}$,
- $\forall j \in [1, q].\forall k \in [i_j, i_{j+1} - 1].\ a_{i_j} = a_k$, and
- $\forall j \in [1, q].\ a_{i_j} < a_{i_{j+1}}$.

Such constraints mean that the list $a_1, \ldots, a_m$ can be partitioned into q sublists, each containing identical values, and each pair of lists containing distinct values. For example, the distinction of the sorted list of probabilities $0.00, 0.00, 0.05, 0.10, 0.10, 0.10, 0.15$, is the list of indices $1, 3, 4, 7$.

The small model theorem states that there exists a *cutoff* to the number of Links to be considered in a system.

Theorem 1 (Small model theorem). *For any naturals $m, c_1, \ldots, c_m > 0$ and probabilities $a_1, \ldots, a_m$. Given the MDPs $\textsc{Link}_1^{c_1}(a_1), \ldots, \textsc{Link}_m^{c_m}(a_m)$. For any MDP $\mathcal{M}$ and formula $\Phi \in PCTL^\star$ the following holds:*

$$\mathcal{M} \parallel \textsc{Link}_1^{c_1}(a_1) \parallel \ldots \parallel \textsc{Link}_m^{c_m}(a_m) \models \Phi \Leftrightarrow \mathcal{M} \parallel \textsc{Link}_1^{c_{i_1}}(a_{i_1}) \parallel \ldots \parallel \textsc{Link}_q^{c_{i_q}}(a_{i_q}) \models \Phi$$

where, for some $0 < q \leq m$, the list of indices $i_1, \ldots, i_q$ is a distinction of the list $a_1, \ldots, a_m$ (assume w.l.o.g. that the latter is sorted), the dispatch probabilities are given by $d_{i_j} = \sum_{k=i_j}^{i_{j+1}-1} d_k$ while the capacities are defined as $c_{i_j} = \sum_{k=i_j}^{i_{j+1}-1} c_k$.

4 Conclusions

In our works we have introduced a novel formal probabilistic model to verify security properties of online storage clouds based on data dispersal algorithms. We have also considered different (probabilistic) models of attackers, namely some try to intercept the traveling slices while others try to attack the providers hosting the slices. In both cases their aim is to collect at least k slices, allowing to reconstruct the user's secret.

Based on this we designed a methodology for assessing security of dispersed cloud storage architectures. In the case of an attacker intercepting the exchanged slices, we were even able to prove a small model theorem for the number of storage providers to be model checked in the system. This in turn allowed us to measure the confidentiality of such systems for any number of storage providers in the network. Our methodology can be applied (1) *a posteriori* to measure the degree of security of an existing system w.r.t. a given specification, or (2) *a priori* in order to determine the best parameter values allowing the system to minimize the likelihood of an attack, from the considered intruder model.

In [1,2,3] we have shown how to apply our methodology on scenarios where a slice attacker is present, both for assessing security of an existing system and for determining best parameter values. In particular, in [2] we exploited our methodology to find parameter values ensuring that the dispersed cloud storage reaches the degree of security known as *perfect secrecy*. Intuitively, the latter means that the slice attacker has equal likelihoods of intercepting k slices and of guessing the entire message content out of nothing. In [3] the methodology has been used to model scenarios with a provider attacker.

References

1. Baldi, M., Bartocci, E., Chiaraluce, F., Cucchiarelli, A., Senigagliesi, L., Spalazzi, L., Spegni, F.: A probabilistic small model theorem to assess confidentiality of dispersed cloud storage. In: Quantitative Evaluation of Systems - 14th International Conference, QEST 2017, Proceedings. pp. 123–139 (2017)
2. Baldi, M., Chiaraluce, F., Senigagliesi, L., Spalazzi, L., Spegni, F.: Security in heterogeneous distributed storage systems: A practically achievable information-theoretic approach. In: Computers and Communications (ISCC), 2017 IEEE Symposium on. pp. 1021–1028. IEEE (2017)
3. Baldi, M., Cucchiarelli, A., Senigagliesi, L., Spalazzi, L., Spegni, F.: Parametric and probabilistic model checking of confidentiality in data dispersal algorithms. In: High Performance Computing & Simulation (HPCS), 2016 International Conference on. pp. 476–483 (2016)
4. Li, M., Qin, C., Li, J., Lee, P.P.: CDstore: Toward reliable, secure, and cost-efficient cloud storage via convergent dispersal. IEEE Internet Comp. 20(3), 45–53 (2016)
5. Resch, J.K., Plank, J.S.: Aont-rs: blending security and performance in dispersed storage systems. In: Proceedings of the 9th USENIX conference on File and stroage technologies. pp. 14–14 (2011)
6. Shen, L., Feng, S., Sun, J., Li, Z., Wang, G., Liu, X.: Clouds: A multi-cloud storage system with multi-level security. In: Proceedings of the International Conference on Algorithms and Architectures for Parallel Processing. pp. 703–716 (2015)