Graph-Theoretic Algorithms for the "Isomorphism of Polynomials" Problem

Charles Bouillaguet, Pierre-Alain Fouque, Amandine Véber

Graph-Theoretic Algorithms for the "Isomorphism of Polynomials" Problem

Charles Bouillaguet[1], Pierre-Alain Fouque[2] and Amandine Véber[3]

[1] University of Lille-1
charles.bouillaguet@univ-lille1.fr
[2] University of Rennes-1
pierre-alain.fouque@univ-rennes1.fr
[3] CMAP Lab, CNRS and Ecole Polytechnique
amandine.veber@cmap.polytechnique.fr

Abstract. We give three new algorithms to solve the "isomorphism of polynomial" problem, which was underlying the hardness of recovering the secret-key in some multivariate trapdoor one-way functions. In this problem, the adversary is given two quadratic functions, with the promise that they are equal up to linear changes of coordinates. Her objective is to compute these changes of coordinates, a task which is known to be harder than Graph-Isomorphism. Our new algorithm build on previous work in a novel way. Exploiting the birthday paradox, we break instances of the problem in time $q^{2n/3}$ (rigorously) and $q^{n/2}$ (heuristically), where q^n is the time needed to invert the quadratic trapdoor function by exhaustive search. These results are obtained by turning the algebraic problem into a combinatorial one, namely that of recovering partial information on an isomorphism between two exponentially large graphs. These graphs, derived from the quadratic functions, are new tools in multivariate cryptanalysis.

1 Introduction

The notion of *equivalent linear maps* is a basic concept in linear algebra; two linear functions f and g over vector spaces are equivalent if and only if there exist two other linear bijective functions S and T such that $f = T \circ g \circ S$. Geometrically speaking, this means that f and g are essentially the same function, but with coordinates expressed in different bases. The computational problem consisting in testing the equivalence of two linear functions (given by matrices) is easy, because it is well-known that two linear maps are equivalent if and only if they have the same rank.

This notion of equivalent linear maps lends itself to an obvious generalization, by dropping the requirement that the functions shall be linear. Then, given two vector spaces U and V, of respective dimension n and m, two functions $f, g : U \to V$ are said to be equivalent if there exist an invertible $n \times n$ matrix S and an invertible $m \times m$ matrix T such that $g = T \circ f \circ S$. Again, the geometric interpretation of this notion is that g and f are "the same function", up to linear

changes of coordinates. However, deciding the equivalence of two such functions is no longer easy in general.

The case where f and g are polynomial maps is particularly relevant, not only because it is a natural generalization of the linear case, but also because f and g admit a compact representation. It is understood that a polynomial map f is such that each coordinate of the vector $f(x)$ is a polynomial expression of the coordinates of the vector x. Testing the equivalence of two polynomial maps has been called the "Isomorphism of Polynomials" (IP) problem by Patarin in 1996 [43], and later the "Polynomial Linear Equivalence" (PLE) problem by Faugère et al. in 2006 [25].

One aspect of PLE that makes it a bit difficult to study is that depending on the parameters (dimensions and base field of the vector spaces, degree of the polynomials, special restrictions, etc.), the problem can take very different forms. We will thus focus on the case where the base field of the vector space is finite (of size q), where polynomials are quadratic, and where their domain and codomain are the same, *i.e.*, where $f, g : (\mathbb{F}_q)^n \to (\mathbb{F}_q)^n$ are quadratic maps. This is the setting that appears in most cryptographic constructions. In the sequel we will call this particular restriction the Quadratic Maps Linear Equivalence (QMLE) problem. In order to make our exposition simpler, we will furthermore assume that q, the size of the finite field, is a power of two. The theory of quadratic forms presents itself very differently for odd characteristic and for characteristic two, and in order not to expose two variants of each of our results, we chose the most computer-oriented setting.

The first "multivariate" cryptographic schemes relied on a somewhat heuristic construction to build Trapdoor One-Way Functions, whose security was based on the hardness of QMLE. Starting with an easy-to-invert quadratic map f, one builds an apparently random-looking one by setting $g = T \circ f \circ S$. The idea is that the changes of coordinate would hide the structure of f that makes it easy to invert, so that g would look random. Inverting random quadratic maps is extremely hard, and the best options in general are exhaustive search (if q is small), or the computation of a Groebner basis (when q is large), both techniques being exponential in n. This construction backed one of the advertized goals of multivariate cryptography, namely the ability to encrypt or sign n-bit blocks while offering n bits of security, as opposed to, *e.g.* RSA.

In this setting, g (and eventually f) is the public key, while S and T are the secret key. When f is public, then recovering the secret-key precisely means solving an instance of QMLE. Several cryptosystems have been built on this idea [10, 55, 18, 32, 15, 7], but they have *all* been broken [29, 24, 20, 19, 37, 29, 35, 9, 28, 40, 11]. The main reason behind this fiasco is that the specific instances of QMLE exposed by these schemes were weak because f was too special, so that polynomial-time and/or efficient algorithms to crack them have eventually been designed.

In a different direction, Patarin also proposed to use the hardness of *arbitrarily chosen* instances of the PLE problem to design a public-key identification scheme, thus potentially avoiding the aforementioned disaster. A *prover*, who

has generated a pair of private/public keys (PK, SK), wants to prove her identity to a *verifier* who knows PK. In fact the prover aims to convince that she knows SK, but without revealing any information about SK to the verifier, or to anybody else. In 1986, Goldreich, Micali and Wigderson [33] built an elegant zero-knowledge proof system for Graph Isomorphism (GI) and used it to build an identification scheme. There, PK is a pair of isomorphic graphs, and SK is the isomorphism (a permutation of the vertices). In order for this system to be secure, it must be hard to solve the instance of GI formed by the public-key. Despite a large research effort, until now no algorithm has been able to solve instances of GI in worst-case polynomial, which is certainly encouraging. However, most instances of GI, and in particular random instances, are *extremely easy* to solve. Thus, the identification scheme of [33] relied on a presumably hard problem for which we do not know how to generate non-trivial instances...

Patarin's suggestion was that Graph Isomorphism could be replaced by QMLE, with the hope that *random instances* of the problem would then be hard, and that key-generation would then be straightforward. There was apparently nothing to lose with the new problem, because it was shown to be harder than GI [44]. Using random instances would in principle avoid the weak instances that had been broken. The resulting QMLE-based identification scheme is not particularly efficient, and does not enjoy very attractive key-sizes, but it is quite simple. It also has a few interesting features compared to other identification schemes based on NP-hard combinatorial problems such as [47–52]: most notably, it does not require hash functions nor commitment schemes, and it does not require the parties to share a (usually large) public common string describing an instance of the NP-complete problem.

1.1 Related Work

The QMLE problem is reminiscent of the Even-Mansour cipher [23], which turns a fixed n-bit permutation P into an n-bit block-cipher with $2n$-bit key by setting $E_{k_1,k_2}(x) = P(x + k_1) + k_2$. Attacks against this construction aim to recover the keys while only having black-box access to E and P. One of its distinctive features is that the performance of a successful adversary running in time t and sending q queries is limited by $t \cdot q \geq 2^n$, under the assumption that P is a random permutation. The known attacks match this bound [17, 22]. As mentioned above, the hardness of QMLE would allow a similar construction where a fixed and public quadratic permutation P is turned into a public-key encryption primitive $E_{S,T} = T \circ P \circ S$. In this context, adversaries not only have oracle to E and P, but know their full description.

Essentially two non-trivial algorithms have been proposed so far for QMLE: the "To-and-Fro" approach [44] on the one hand, and the "Groebner Basis" approach [25] on the other hand. There are also several, more efficient algorithms for the special case where the secret T matrix is known to be the identity matrix [31, 46, 14, 36]. This sub-problem is also GI-hard, even in very restricted settings [1]. The article [3] considers the particular case of testing whether two boolean func-

tions are equal modulo a permutation of their inputs. It shows that $2^{n/2}$ queries are necessary if one only has black-box access to the boolean functions.

Back to the full QMLE problem, the "To-and-Fro" algorithm, while being simple, was exposed on a toy example, without pseudo-code nor detailed analyzis. We are convinced that the algorithm works when the polynomial maps f and g are *bijective*, but it cannot work as-is when they are not (the authors of [25] made the same observation). Note that a random polynomial map is not bijective with overwhelming probability. As is it given in [44], the "to-and-fro" algorithm is thus not applicable to random instances of QMLE. We found out that it *is* nevertheless possible to adapt the algorithm to work in the non-bijective case, but there are several ways to do so, and some are more efficient than others. Figuring that out required some work, and exposing it requires some space, so we will not go deeper into this issue in this paper. In any case, the authors of [44] claim that the complexity of their algorithm is of order $\mathcal{O}\left(q^{2n}\right)$ when $q > 2$ and $\mathcal{O}\left(2^{3n}\right)$ when $q = 2$, and we agree with them. The algorithm was later independently rediscovered under the form of a procedure to test the linear equivalence of S-boxes [12].

The "Groebner basis" algorithm, on the other hand is not heuristic, and is well-specified. It consists in identifying coefficient-wise the equation $T^{-1} \circ g = f \circ S$, which relates two vectors of n quadratic forms. It is therefore equivalent to about n^3 quadratic equations in the $2n^2$ coefficients of the unknown changes of coordinates. These equations are then solved through the computation of a Groebner basis. The complexity of Groebner basis algorithms is notoriously tricky to study, and the authors of [25] did not give any definitive results. However, they empirically observed an important fact, namely that when f and g are *inhomogeneous* quadratic maps, *i.e.*, when f and g contains non-zero linear and constant terms, then their algorithm terminated in polynomial time $\mathcal{O}\left(n^9\right)$. In the homogeneous case, the authors of [25] conjectured that their algorithm is subexponential, without providing any argument nor any evidence that it is the case. This assertion is impossible to verify in practice because the complexities are too high, but our own reasoning makes us more inclined to believe that the algorithm is plainly exponential. Assuming that the equations form a semi-regular sequence would allow to estimate the complexity of the Groebner basis computation [8]; doing so results in a total complexity of $\mathcal{O}\left(2^{18n}\right)$, yet assuming that the equations are semi-regular is probably a bit of a stretch. Establishing the complexity of this algorithm is thus essentially an open problem.

In the sequel, we will nevertheless take for granted that inhomogeneous instances of QMLE are tractable and can be solved in polynomial time, using the "Groebner-based" algorithm for instance.

It must be noted that in [44], the existence of an algorithm based on the birthday paradox and running in time $\mathcal{O}\left(q^{n/2}\right)$ is asserted, and that this algorithm is itself partially described in [45], where it is called the "combined powers attack". This algorithm is sometimes acknowledged for in the literature (*e.g.* in [25]). However, it is underspecified to the point that it is impossible to implement it, and some of the bits that are specified have major problems. Some of

them deterministically fail to meet their goal, and the whole construction relies on heuristic assumptions that are empirically false (sometimes provably). This "algorithm" should thus be disregarded.

1.2 Our Results

We give three algorithms to solve QMLE in the homogeneous case. All these algorithms work by reducing the solution of a homogeneous (hard) instance into that of one or several inhomogeneous (easy) instances after some preprocessing. We will thus assume that we are given a (black-box) *Inhomogeneous solver* that presumably works in polynomial time, and we will count the number of *inhomogeneous queries* sent to this oracle. We are well-aware that this assumption is quite strong. The empirical success of the algorithm of [25] convinced us that it works in polynomial-time on average, yet moving from there to "worst-case polynomial time" seems like a leap of faith. However, this assumption eases our exposition considerably, and in practice there does not seem to be any problem (probably because the queries sent to the inhomogeneous oracle are random enough).

Our three algorithms differ by the number of queries they send to the oracle, by the amount of computation they perform themselves, and by their success probability.

Algo.	Preprocessing	Inhom. queries	success prob.	
1		q^n	1	
2	$\mathcal{O}\left(n^3 \cdot q^{2n/3}\right)$	$q^{2n/3}$	62%	
3	$\mathcal{O}\left(n^5 \cdot q^{n/2}\right)$	1	62 %	only when $q = 2$

Algorithm 1 is deterministic, and essentially performs an exhaustive search in $(\mathbb{F}_q)^n$, sending one inhomogeneous query per vector. Using the algorithm of [25] to deal with the inhomogeneous instances, the resulting complexity is $\mathcal{O}\left(n^9 \cdot q^n\right)$, which already improves on the "to-and-fro" algorithm of [44].

Algorithms 2 and 3 rely on the birthday paradox to improve on exhaustive search and break the q^n barrier. To this end, two exponentially large isomorphic graphs are derived from the two quadratic maps. Recovering a bit of information on an isomorphism allows to make the problem inhomogeneous, and thus easy to solve. The trick is that this partial information must be extracted without knowing the full graphs, because they are too large. The construction of these graphs borrows from the differential techniques that have broken SFLASH, amongst others.

Algorithm 2 is relatively easy to analyze and we rigorously establish its complexity and success probability when dealing with random instances of the problem. Algorithm 3 is more efficient but more sophisticated and harder to analyze (as well as somewhat heuristic). We provide an as-rigorous-as-possible complexity analysis under a conjecture on random quadratic maps, and we verify experimentally that we are not off by too much.

Because our algorithms are exponential in n, we do not fully break Patarin's identification scheme (it is of no practical value anyway), even though its key-sizes should in principle be doubled. The construction of a Trapdoor One-Way Function from QMLE outlined above has already been bludgeoned to death by cryptanalysts, and it now lies on the autopsy table. We take the role of the medical examiner that appears in every good police drama, only to discover that the corpse had a fatal disease even before being brutally assaulted. We indeed believe that our algorithms condemn this generic construction of a Trapdoor One-Way Function *post-mortem*, and give a theoretical reason not to try again, besides the obvious "they have all been broken" argument. Our algorithms indeed break the QMLE instance and retrieve the secret-key (asymptotically) much faster than inverting the quadratic map by exhaustive search. This shows in passing that this construction can only offer $n/2$ bits of security, instead of the n that was its original objective.

2 A First Algorithm Based on Dehomogenization

Confronted with a *homogeneous* instance of QMLE, our strategy throughout this paper is to build an *inhomogeneous instance* admitting the exact same solutions. This inhomogeneous instance can in turn be solved in polynomial time, and reveals the solution(s) of the original problem. The downside of this approach is that the image of S must be known at one arbitrary point of the vector space. Indeed, if $\beta = S \cdot \alpha$, then:

$$\forall x.\ g(x) = T \cdot f(S \cdot x) \qquad \Longleftrightarrow \qquad \forall x.\ g(x+\alpha) = T \cdot f(S \cdot x + \beta).$$

Thus defining $g'(x) = g(x+\alpha)$ and $f' = f(x+\beta)$ yields an equivalent problem, *i.e.*, an instance that has the same solutions as the original one. In addition, the new instance is inhomogeneous. This follows from the simple observation that although x^2 is a homogeneous polynomial, $(x+\alpha)^2 = x^2 + 2\alpha x + \alpha^2$ is not since it has a non-trivial linear term αx and a non-trivial constant term α^2.

It follows that solving (homogeneous) instances of QMLE essentially boils down to finding $S\alpha$, for some known and non-zero vector α. Exhaustive search is the first option that comes to mind, leading to Algorithm 1. This algorithm sends q^n queries to the inhomogeneous solver in the worst case, and finds the solutions when they exist. This algorithm terminates with probability one in time $\mathcal{O}\left(n^9 \cdot q^n\right)$ if the Groebner-based algorithm of [25] is used to solve the inhomogeneous instances. Despite being extremely simple, Algorithm 1 is asymptotically q^n times faster than to the "to-and-fro" algorithm of [44].

This *dehomogenization* technique exposes a crucial asymmetry in the problem: it is apparently much more critical to obtain knowledge on S than on T. This is not new: the "To-and-Fro" algorithm relies on the ability to transfer knowledge of a relation $\beta = S \cdot \alpha$ to a relation $g(\alpha) = T \cdot f(\beta)$.

Algorithm 1 Simple algorithm based on dehomogenization.

function EXHAUSTIVE-DEHOMOGENIZATION(f, g)
  $x \leftarrow$ random non-zero vector in $(\mathbb{F}_q)^n$
  for all $0 \neq y \in (\mathbb{F}_q)^n$ **do**
    $f'(z) \leftarrow f(z + y)$
    $g'(z) \leftarrow g(z + x)$
    query IQMLE-SOLVER with (f', g'')
    if solution (S, T) found **then return** (S, T)
  return "Not Equivalent"

3 Moving the Problem Into a Graphic World

Using the birthday paradox is a natural idea to improve on exhaustive search algorithms in many scenarii, with the hope to halve the exponent in the complexity. Here, we wish to use the birthday paradox to obtain the image of S at one point, and build a dehomogenized instance, just as we did in the previous section. One difficulty is that we want to focus only on S, and leave T alone. To this end, we introduce a tool which is, to the best of our knowledge, new. We associate a *graph* G_h to any quadratic map $h : (\mathbb{F}_q)^n \mapsto (\mathbb{F}_q)^n$. Its vertices are the elements of $(\mathbb{F}_q)^n$, and there is an edge between $x, y \in (\mathbb{F}_q)^n$ if and only if $h(x + y) = h(x) + h(y)$. To some extent, G_h expresses the "linear behavior" of h (even though h is not linear) and thus we call these graphs the "linearity graphs" of the associated quadratic maps.

These graphs are natural objects associated to quadratic maps. For instance, the distinguisher of [21] to determine whether a given quadratic map f is an HFE public key can be rephrased as follows: pick a random node in G_f, and count its neighbors. If their number exceeds a given bound (which depends on the degree of the internal HFE polynomial), then return "random", else return "HFE". With the right bound on the number of neighbors, this algorithm achieves subexponential advantage.

The essential interest of linearity graphs for our purposes is that the two graphs G_f and G_g are connected by the secret matrix S.

Lemma 1. *If $T \circ g = f \circ S$ then S is a graph isomorphism that sends G_f to G_g.*

Proof. Indeed, if $x \leftrightarrow y$ in G_g, then by definition $g(x + y) = g(x) + g(y)$, and it follows that $T \circ g(x + y) = T \circ g(x) + T \circ g(y)$, and thus that $f(S \cdot x + S \cdot y) = f(S \cdot x) + f(S \cdot y)$. This in turn means that $S \cdot x \leftrightarrow S \cdot y$ in G_f. It follows that S is a graph isomorphism between G_f and G_g.

Linearity graphs thus allows a formulation of the problem where the other secret matrix T is no longer present. We have two (exponentially large) isomorphic graphs G_f and G_g, and we ultimately need to recover the whole isomorphism S. However, thanks to the dehomogenization technique of the previous section, and thanks to the ease with which inhomogeneous instances can be solved, it turns out that recovering just a little bit of information on the isomorphism

is enough to find it completely. More precisely, we just need to know how the isomorphism S transforms one arbitrary vertex.

Of course, completely building these graphs is prohibitively expensive (they have q^n vertices). It turns out that this is never necessary, because it is possible to walk in these graphs without fully knowing them.

Walking in Linearity Graphs. The function $\psi(x, y) = f(x+y)+f(x)+f(y)$ is a generalization of the *polar form* of a quadratic form to vectors thereof, in characteristic two. It is easy to check that ψ is bilinear. Given a (non-zero) vertex $x \in (\mathbb{F}_q)^n$ in the graph, the function:

$$\mathrm{D}_x f : y \mapsto \psi(x, y) = f(x + y) + f(x) + f(y)$$

is a familiar object in multivariate cryptology, called the *Differential of f at x* [27, 21, 19, 29]. It is a linear function from $(\mathbb{F}_q)^n$ to $(\mathbb{F}_q)^n$, which is then conveniently represented by a matrix. The set of nodes adjacent to x in G_f is in fact the kernel of $\mathrm{D}_x f$. Note that x always belong to $\ker \mathrm{D}_x f$, because $x + x = 0$. The main reason we chose to focus on the case where $q = 2^e$ is that this fact is not true when q is not a power of two.

The matrix $\mathrm{D}_x f$ is easy to compute given f and x. If f is a (homogeneous) quadratic map, then it is in fact a vector of n quadratic forms, which can conveniently be described by a collection of n matrices $F_1, \ldots, F_n$, that are interpreted as follows: $F_k[i, j]$ is the coefficient of $x_i x_j$ in the k-th component of f. If ${}^tM$ denotes the transpose of M, then the matrix representation of the differential of f at x is given by:

$$\mathrm{D}_x f = \left(x \cdot (F_1 + {}^tF_1) \middle| \ldots \middle| x \cdot (F_n + {}^tF_n) \right).$$

Thus, given a vector x, finding the neighbors of x in G_f can be done in time $\mathcal{O}\left(n^3\right)$: computing the matrix $\mathrm{D}_x f$ requires n matrix-vector products, and determining its kernel classically takes $\mathcal{O}\left(n^3\right)$ operations. It is thus possible to crawl the linearity graphs by spending a polynomial number of elementary operations on each traversed vertex.

Structure in Linearity Graphs. Linearity graphs possess a rich structure, thanks to their algebraic origin. Recall that in G_f, two nodes x and y are adjacent if $\psi(x, y) = 0$, where ψ is the symmetric bilinear map defined above. The bilinearity of ψ induces a lot of structure in G_f. For instance, we always have $\psi(x, x) = 0$, and by bilinearity $\psi(\lambda x, \mu x) = \lambda\mu\psi(x, x) = 0$, so that the q multiples of a vector x form a clique in G_f. The set of all multiples of x are thus topologically indifferentiable (they all have the exact same neighborhood).

Furthermore, the same reasoning shows that if two vectors x and y are adjacent in G_f, then the set of q^2 linear combinations $\lambda x + \mu y$ form a clique in G_f of size q^2.

Degree Distribution. If a quadratic map f is randomly chosen (amongst the finite number of possibilities), then the resulting linearity graph G_f follows a certain —mostly unknown— probability distribution, and any property of G_f can be seen as a random variable. One of the most interesting properties of G_f is the distribution of the *degree* (*i.e.*, of the number of neighbors) of vertices in G_f. This result is stated in terms of the probability that a random $n \times n$ matrix over $\mathbb{F}_q$ is invertible. We denote it by $\lambda(n)$:

$$\lambda(n) = \prod_{i=1}^{n} \left(1 - \frac{1}{q^i}\right)$$

Lemma 2 (theorem 2 in [21]). *Let $x \in (\mathbb{F}_q)^n$ be a non-zero vector, and $f : (\mathbb{F}_q)^n \to (\mathbb{F}_q)^n$ be a uniformly random quadratic map. Then $\mathrm{D}_x f$ is a uniformly random matrix vanishing over x. As a consequence, the probability that $\mathrm{D}_x f$ has a kernel of dimension $k \geq 1$ is:*

$$\frac{\lambda(n)\lambda(n-1)}{\lambda(k)\lambda(k-1)\lambda(n-k)} q^{-k(k-1)}$$

Because $\lambda(n)$ is a decreasing function of n that converges to a finite limit bounded away from zero, then the ratio of the λ-expressions lives in a small interval, independently of q, n and k, so that the probability is in fact of order $q^{-k(k-1)}$. Of course, over $\mathbb{F}_q$, a k-dimensional vector space contains q^k elements, so that if $\dim \ker \mathrm{D}_x f = k$, then the vertex x has q^k neighbors.

Sparsity. Computing the expectation and the variance of the degree is technical, but feasible:

$$\mathbb{E}\left[\text{degree}\right] = q - \frac{1}{q^{n-2}} \qquad \sigma^2 = q^2(q-1)\left(1 - \frac{q^2+1}{q^n} + \frac{q^2}{q^{2n}}\right)$$

Establishing these two expressions is somewhat technical, yet because both are sums of q-hypergeometric terms, they can be computed by "creative telescoping" thanks to the q-analog of Zeilberger's algorithm [56]. It follows that the expected number of edges of G_f is essentially $q^{n+1}/2$. In other terms, G_f is a very sparse graph that has barely more edges than it has vertices.

Disconnecting Linearity Graphs. A linearity graph G_f is fully connected, because all vertices are adjacent to the "zero" vertex. This "zero" vertex is not very interesting (since it is adjacent to *every* other vertex), and, as a matter of fact, it even turns out to be a bit annoying. Thus, it seems that there is nothing to lose by removing it. In addition, we could also get rif of the self-loops ; they are useless since *every* vertex has one.

We thus denote by G_f^* the simple graph G_f in which the zero vertex has been removed, and where self-edges are removed. It is interesting to note that the resulting graph is no longer connected, and that there are in fact very many

connected components. Indeed, if $\dim \ker \mathrm{D}_x f = 1$, then the only neighbors of x are its multiples, and x belong to a connected component of size $q-1$. Lemma 2 tells us that this happens with probability $\lambda(n)/\lambda(1)$, and this converges to a finite limit bounded away from zero when n goes to infinity. Thus, a constant fraction of the vertices belong to "small" connected components of size $q-1$. Working a bit on the λ functions reveals that this proportion grows like $1-1/q^2$.

4 Just Count Your Neighbors

It is well-know that if two graphs (V_1, E_1) and (V_2, E_2) are isomorphic, and if ρ is an isomorphism between them, then $u \in V_1$ and $\rho(u) \in V_2$ have the same *degree*, *i.e.*, the same number of neighbors. It follows that if $u \in V_1$ and $v \in V_2$ do not have the same degree, then they cannot be related by ρ.

We adapt this simple idea in the context of QMLE, under the form of Algorithm 2. The main idea in this algorithm is to target vertices in the linearity graphs of f and g that have a specific degree: we only look for a "right pair" $y = S \cdot x$ amongst vertices x, y that have a prescribed degree (chosen to optimise the complexity of the algorithm). The remaining of this section is devoted to establishing the properties of this algorithm, which are summarized in the following theorem.

Algorithm 2 First Birthday Based Algorithm

function SampleSet(h)
  $L \leftarrow \emptyset$
  repeat
    repeat
      $x \leftarrow$ random vertex of G_h
    until x has $q^{\sqrt{n/3}}$ neighbors
    $L \leftarrow L \cup \{x\}$
  until $|L| = \sqrt{2} q^{n/3}$
  return L

function Neighbor-Counting-QMLE(f, g)
  $U \leftarrow$ SampleSet (f)
  $V \leftarrow$ SampleSet (g)
  for all $(x, y) \in U \times V$ **do**
    $f'(z) \leftarrow f(z + y)$
    $g'(z) \leftarrow g(z + x)$
    query IQMLE-Solver with (f', g')
    if solution (S, T) found **then return** (S, T)
  return "Probably not equivalent"

Theorem 1. *Algorithm 2 performs $\mathcal{O}\left(q^{2n/3}\right)$ units of computations on average, sends at most $q^{2n/3}$ queries to the inhomogeneous solver, and succeeds with probability $1 - 1/e$.*

The helper function SampleSet returns a set of $\mathcal{O}\left(q^{n/3}\right)$ vertices of G_f (resp. G_g), each having $q^{\sqrt{n/3}}$ neighbors in the graph. It follows that there are $q^{2n/3}$ queries to the inhomogeneous solver, because this is the size of the cartesian product $U \times V$.

It remains to establish the complexity of SampleSet, and the success probability of the algorithm. As explained above, since we are looking for a "right pair" $y = S \cdot x$, it is safe to restrict our attention to vertices x, y that have a specific degree (as long as vertices with such a degree exist in the graphs).

Lemma 2 gives us the expected number iterations of the innermost loop of SampleSet that are required to find a random vertex with the required degree. Up to a constant factor, finding a vertex with degree q^k requires $q^{k(k-1)}$ trials, so that finding each new random vertex requires $\mathcal{O}\left(q^{n/3}\right)$ rank computations on $n \times n$ matrices, hence $\mathcal{O}\left(n^3 \cdot q^{n/3}\right)$ operations.

Lemma 2 also tells us that there are on average $q^{n-k(k-1)}$ vertices in G_f each having degree q^k. In Algorithm 2 we look specifically at vertices of degree $q^{\sqrt{n/3}}$, and we thus expect G_f to contain $q^{2n/3}$ of them. Since the number of iterations of the outermost **repeat**...**until** loop is roughly the square root of this number, we do not expect more than a constant number of "extra" iterations finding an already-known vector x. Putting everything together, we conclude that SampleSet terminates after $\mathcal{O}\left(n^3 q^{2n/3}\right)$ operations.

Now, the birthday bound tells us that $U \times V$ contains a "right pair" $y = Sx$ with probability greater than 63%, because both U and V contain about the square root of the total number of vertices with degree $q^{\sqrt{n/3}}$ (see [53] for a precise statement of this specific version of the birthday paradox).

Practical Results. We have implemented Algorithm 2 inside the MAGMA computer algebra system[13], running on one core of a 2.8 Ghz Xeon machine. As shown in Table 1, we found out that in practice it is difficult to balance the cost of building U and V on the one hand, and going through the candidate pair on the other hand, because the target degree can only take $\sqrt{n}$ integer values. We could nevertheless verify in practice that the complexity of building the lists and the expected number of right pairs in them is consistent with our expectations. The source code is in the public domain, and is available on the webpage of the first author. It uses an unpublished algorithm to solve the inhomogeneous instances.

n	q	generating U and V	total time	$\log_q$ (target degree)	$\lvert U \rvert$	# pairs
16	2	0s	68s	3	1	4
22	2	28s	9h45m	4	13	400
28	2	4913s	2h15m	5	8	64

Table 1. Experimental results on Algorithm 2.

5 Map Your Neighborhood

We have seen in section 3 that the linearity graphs, once deprived from the "zero" vertex, contain many small connected components. Of course, if $y = Sx$, then the connected component of x is isomorphic to the connected component of y. In this section, we describe an algorithm that builds upon this idea—instead of just looking at immediate neighbors, as we did in algorithm 2, we now try to look at the whole connected component, in order to distinguish between vertices of the same degree.

Canonical Graph Labeling. Given a graph G, a *Canonical Labeling algorithm* relabels the vertices of G, thus producing a graph $Canon(G)$, which is by definition isomorphic to G. The result is canonical in the sense that if G and H are isomorphic graphs, then $Canon(G) = Canon(H)$. The canonical labels are therefore complete invariants of the isomorphism class, and as such, computing a canonical labeling is necessarily harder than checking if two graphs are isomorphic. However, computing a canonical labeling can be done in average linear time [6], because except for an exponentially small fraction of all graphs, it can be done with a very simple linear algorithm. Deterministic algorithms that always succeed are subexponential, with complexity $\mathcal{O}\left(\exp\left(\sqrt{n \log n}\right)\right)$ [5]. The perhaps most well-known, and most practical algorithm dates back to 1978, and is implemented in the `nauty` open-source package [38]. It is known to be exponential on some specific counter-examples [39], but otherwise performs exceptionally well. There are also many relevant classes of graphs where canonical labeling is polynomial [26]: graphs of bounded degree, planar graphs, chordal graphs, graphs of bounded treewidth, etc.

Back to our more specific problem, let us denote by C_x (resp C_y) the connected component of x in G_f^* (resp. of y in G_g^*). The key idea of the algorithm presented in this section is that $y = Sx$ implies $Canon(C_x) = Canon(C_y)$. Thus, it seems that the function $H : u \mapsto Canon(C_u)$ could be used as a "hash function". In fact, in algorithm 2, we used the degree as such a "hash function", but it was not very discriminating, because the degree does not contain enough entropy. We hope that H behaves as a good hash function, and that *false positives*, *i.e.*, pairs (x, y) such that $H(x) = H(y)$ but $y \neq Sx$, should be very rare.

One problem is that H does not distinguish between vertices of the same connected component. To improve it, we would need a way to single out a specific vertex in the connected component. Fortunately, most canonical labeling algorithm return the isomorphism (say ρ) between their argument G and $Canon(G)$. To single a vertex x out in G, it is sufficient to send $\rho(x)$ along with the canonical labeling of G.

A Canonical-Labeling-Based Algorithm As discussed in section 3, G_f^* contains many small connected components that are all isomorphic to each others, since they are all cliques of size $q - 1$. Therefore, if we want our "hash function" to be discriminating, we must avoid small connected components. Our "hash

function" will thus reject the vector x if there is no simple path starting from x and of length at least r. In the other direction, we cannot exclude the existence of a giant connected component of exponential size. Therefore, we only consider the radius-r neighborhood of the vertex x we are interested in, *i.e.*, the set of all vertices that can be reached from x by crossing at most r edges. This is the basis of algorithm 3.

Algorithm 3 Canonical Labeling/Birthday Based Algorithm

function HASHABLE$^{[r]}(G, x)$
  Perform a Breadth-First Search in G starting from x
  return TRUE if the BFS hits a vertex r edges away from x

function $\mathrm{H}^{[r]}(G, x)$
  $C_x \leftarrow$ subgraph of G formed by all vertices at most r edges away from x.
  $\rho, \mathcal{G} \leftarrow$ CANONICALLABELING(C_x)
  return $\big(\mathcal{G}, \rho(x)\big)$

function SAMPLEHASHTABLE(h)
  $L \leftarrow \emptyset$
  repeat
    repeat
      $x \leftarrow$ random vertex of G_h^*
    until HASHABLE$^{[r]}(G_h^*, x)$
    $L\left[\mathrm{H}^{[r]}\left(G_h^*, x\right)\right] \leftarrow x$
  until $|L| = \sqrt{2}q^{n/2}$
  return L

function CANONICAL-LABELING-QMLE(f, g)
  $U \leftarrow$ SAMPLEHASHTABLE$\,(f)$
  $V \leftarrow$ SAMPLEHASHTABLE$\,(g)$
  for all $(h_1 \mapsto x) \in U, (h_2 \mapsto y) \in V$ **such that** $h_1 = h_2$ **do**
    $f'(z) \leftarrow f(z + y)$
    $g'(z) \leftarrow g(z + x)$
    query IQMLE-SOLVER with (f', g')
    if solution (S, T) found **then return** (S, T)
  return "Probably not equivalent"

Remarks on Algorithm 3. Establishing the complexity and success probability of algorithm 3 is surprisingly difficult, probably because is relies on topological properties of G_f^*, which is a somewhat random but very structured graph.

Algorithm 3 has been written in a generic way, independently of the actual value of q. However, we have only been able to discuss its properties when $q = 2$. We have verified that the algorithm works as we expected in this case, but the situation when $q \neq 2$ is not so clear. We tend to believe that the complexity

and/or success probability degrade exponentially fast when q grows, but we fall short of definitive conclusion.

When $q = 2$, the structure of G_f^* seems to be richer. For instance, we already alluded to the fact that the fraction of nodes whose connected component is of size only $q - 1$, grows like $1 - 1/q^2$. In addition, as we will see in the next section, setting $q = 2$ allows us to turn most more-or-less-random graphs into trees, which are much easier to deal with.

Preliminary Analysis of Algorithm 3. When $q = 2$, the correctness of the algorithm is implied by the following three heuristic statements.

Claim. *i)* $\textsc{Hashable}^{[r]}\left(G_f^*, x\right)$ is true with probability $\approx 1/r$ over the random choice of f (assuming $x \neq 0$).

ii) Both $\textsc{Hashable}^{[r]}$ and $\mathrm{H}^{[r]}$ can be evaluated in expected time $\mathcal{O}\left(rn^3\right)$.

iii) When restricted to elements that are $\textsc{Hashable}^{[r]}$, then $\mathrm{H}^{[r]}\left(G_f^*, \cdot\right)$ is an *ε^r-almost universal hash function family* (indexed by f) for some $\varepsilon < 1$.

The notion of almost universal hash function is usually useful when the hash function is "less injective" than a random function. In this paper though, $\mathrm{H}^{[r]}$ can become *more injective* than a random function, as soon as r becomes sufficiently large.

It follows from claim i that the expected number of iterations of the loop of lines 11–13 is $\mathcal{O}(r)$, and it follows from claim ii that finding one admissible vector x requires $\mathcal{O}\left(r^2n^3\right)$ operations on average. Claim iii then guarantees that if we choose r to be a bit larger than n, then the probability to find hash collisions can be made smaller than 2^{-n}, and standard birthday-type results guarantee that the number of expected hash collisions in the execution of SampleHashTable is constant. From this, we conclude that SampleHashTable runs in expected time $\mathcal{O}\left(r^2n^3q^{n/2}\right)$.

It follows from the birthday paradox [53] that there is a "right pair" in $U \times V$, *i.e.*, a pair (x, y) with $y = Sx$, with probability greater than $1 - 1/e$. This is because $\left(\mathbb{F}_q\right)^n$ has q^n elements and that the sizes of both U and V are essentially $q^{n/2}$. This guarantees the success probability of the algorithm.

Let us denote by $\mathcal{N}$ the number of bogus inhomogeneous queries, *i.e.*, the number of pairs $x \neq y \in U \times V$ with the same hash. It follows from Markov's inequality and claim iii that $\mathbb{P}\left[\mathcal{N} \geq 1\right] \leq 2q^n \cdot \varepsilon^r$. Thus, as soon as r is asymptotically larger than n, *e.g.* $r = n \log\log n$, then the probability that $\mathcal{N} \geq 1$ gets exponentially small. This concludes our preliminary analysis: algorithm 3 runs in time $\mathcal{O}\left(n^5q^{n/2}\right)$, and sends a constant number of inhomogeneous queries. It now remains to show that our claims are valid, but we first find it reassuring to show that the practical behavior of the algorithm is very consistent with our expectations.

experimental results. We have implemented Algorithm 3 using the MAGMA computer algebra system [13], and we found out that it works well in practice,

as Table 2 shows. The experiment clearly shows that $\mathcal{N}$ is constant, as expected. This justify our heuristic analysis *a posteriori*. The implementation is in the public domain and is available on the webpage of the first author.

n	q	generating U and V	finding collisions	$\lvert U\rvert$	$\mathcal{N}$
16	2	3.6 s	1s	64	6
24	2	123 s	13s	836	5
32	2	61 min	200s	11585	2
40	2	31 h	2h	165794	7

Table 2. Experimental results on Algorithm 3

6 Discussion of the Claims

Special Structure in Linearity Graphs. Any analysis of algorithm 3 will have to rely on the properties of linearity graphs. As argued above, the situation when $q = 2$ is somewhat different than that obtained with larger values of q. When $q = 2$, the connected components of G_f^* seem to enjoy a very nice structure, as illustrated by figure 1. The origin of the triangles is that any non-isolated vertex x belong to the $(q^2 - 1)$-clique formed by x, y and $x + y$ (0 has been removed). If it were not for these triangles, the connected components of G_f^* would be trees. While this structure is clearly visible on all the examples we could forge, we fall short of any rigorous explanation.

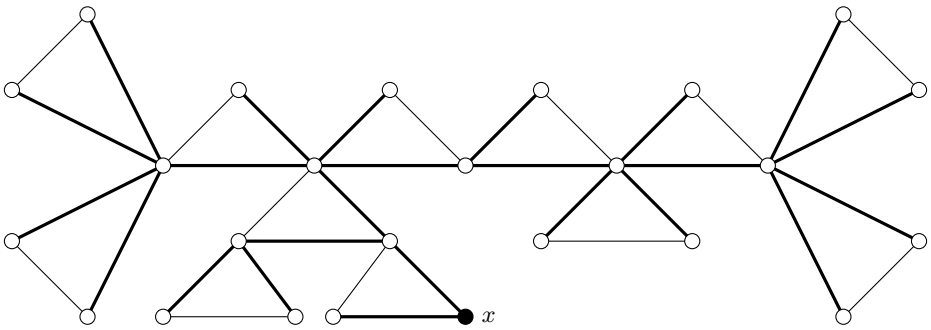

Fig. 1. A typical moderate-size connected component of G_f^* when $q = 2$. Self-edges are not shown. The thick edges show a spanning tree obtained by performing a Breadth-First Search starting from x.

Conjecture 1. When r is polynomial in n, then with high probability the radius-r neighborhood of any vertex in G_f^* does not contain cliques of size strictly greater than $q^2 - 1$. In addition, every edge belongs to at most one maximal clique with high probability.

Back to the Trees. Fig. 1 illustrates that the connected components are close to trees, and this analogy can easily be made rigorous when $q = 2$. To a vertex x in a linearity graph G_f^*, we associate the unordered, unlabeled tree $T^{[r]}(G_f^*, x)$ by performing a Breadth-First Search in G_f^* starting from x, and stopping r edges away from x. It is well-known that any graph traversal induces a spanning tree of the graph. The tree $T^{[r]}(G_f^*, x)$ is simply the spanning tree induced by the BFS (cf. fig. 1).

Lemma 3. *If G_1, G_2 satisfy the properties of Conjecture 1, then:*

$$(G_1, x) \text{ isomorphic to } (G_2, y) \iff \forall r.\ T^{[r]}(G_1^*, x) \text{ isomorphic to } T^{[r]}(G_2^*, y)$$

This transformation of connected components of G_f^* into trees serves several purposes : not only it helps understanding why our three claims hold, but is also allows a more efficient formulation of algorithm 3. Indeed, $\textsc{Hashable}^{[r]}(G, x)$ can be evaluated by checking if $T^{[r]}(G, x)$ has depth r. Lastly, it is well-known that unordered, unlabeled trees can be canonically labeled in linear time thanks to a venerable algorithm of Aho, Hopcroft and Ullman [2]. .

Random Trees From Random Linearity Graphs. When f is randomly chosen, then $T^{[r]}(G_f^*, x)$ can also be seen as a random variable. Because each vertex of G_f^* has k neighbors with some probability, then each node of $T^{[r]}(G_f^*, x)$ also has a given number of children (sometimes called "offspring" in the context of branching processes) with some probability. Everything looks as if $T^{[r]}(G_f^*, x)$ were a random tree where the number of descendant of each node was chosen at random according to a given offspring distribution. The offspring distribution of x in $T^{[r]}(G_f^*, x)$ (*i.e.*, the root of the tree) is almost exactly the degree distribution of G_f^*, which is known by lemma 2 (with the caveat that self-loops are removed). However, the offspring distribution of non-root nodes is a bit different:

$$\ell_n(i) = \mathbb{P}\left[\text{a non-root node produces } i \text{ offspring}\right] = \begin{cases} p_{n,k} & \text{when } i = q^k - q^2 \\ 0 & \text{otherwise} \end{cases}$$

where

$$p_{n,k} = \mathbb{P}\left[\dim \ker D_x f = k \middle| y \in \ker D_x f\right] = \frac{\lambda(n)\lambda(n-2)}{\lambda(k)\lambda(k-2)\lambda(n-k)} \cdot q^{-k(k-2)}$$

The expression of $p_{n,k}$ can be derived from a reasoning similar to that of the proof of lemma 2, which can also be found in [21]. It is also possible to compute

the *expected progeny* μ of each non-root node, and the variance σ^2 of the offspring distribution :

$$\mu = 1 - \frac{1}{q^{n-2}} \qquad \sigma^2 = q^2(q-1)\left(1 - \frac{q^2+1}{q^n} + \frac{q^2}{q^{2n}}\right)$$

These two expressions can be derived from the expectation and the variance of the degree in G_f without too much effort.

When a random tree is sampled by choosing independently the number of children of each node according to a fixed law, the resulting object is called a *random Galton-Watson tree*. These random trees are well-studied [4], and this wealth of results would be extremely useful to our own purposes. Unfortunately, in $T^{[r]}(G_f^*, x)$, the number of descendant of each node is not even pairwise-independent.

We nevertheless denote by $\mathfrak{P}_n$ the law of Galton-Watson trees with offspring distribution ℓ_n, and by $\mathfrak{P}_n^{[r]}$ the law of such trees conditioned to be of height at least r. We verified in practice that the following assumption holds very well.

Heuristic Assumption: Over the random choice of f, $T^{[r]}(G_f^*, x)$ has the same properties as Galton-Watson trees sampled according to $\mathfrak{P}_n^{[r]}$ and truncated at depth r.

Because $\mu \leq 1$, trees sampled according to $\mathfrak{P}_n$ are finite with probability one [4]. In addition, the probability that a tree sampled according to $\mathfrak{P}_n$ has height greater than r is equivalent to $2/(r\sigma^2) \approx 2/(r \cdot q^3)$ [4]. This justifies claim *i*.

However, it follows from this result that the expected height of trees sampled according to $\mathfrak{P}_n$ is not finite; this justifies why we stop the BFS after a (finite) depth. It is also known that in trees sampled according to $\mathfrak{P}_n$, the expected total number of nodes after h generation is $h + 1$ [42]. It follows that actually performing the BFS requires on average $\mathcal{O}(r)$ matrix operations. This justifies claim *ii*.

False Positive Rate. It remains to justify claim *iii*, the trickiest one. Under the heuristic assumption that $T^{[r]}(G_f^*, x)$ follows the law $\mathfrak{P}_n$, then claim *iii* is equivalent to the following statement: *the probability that two random trees sampled according to* $\mathfrak{P}_n^{[r]}$ *are isomorphic decreases exponentially fast with* r.

In other word, we must determine the probability that two random trees are isomorphic. While this appears to be a natural question, it has (to the best of our knowledge) not been treated in the literature. We could not establish the required exponential upper-bound in general, however we proved a strong enough bound that holds if we are allowed to reject a negligible amount of trees (*i.e.*, shrinking a bit the HASHABLE$^{[r]}$ domain).

We say that a tree has a *unique spine decomposition* if there is a unique path starting from the root and reaching a leaf of maximal depth. We also say that a tree has a unique spine decomposition *up to height* k if there is a unique path starting from the root and reaching depth k that extends to a path reaching

nodes of maximal depth. Fig 2 shows a tree with a spine decomposition up to a certain level. Note that it is easy (and efficient) to check whether a given tree has this property. We now redefine the hashable domain by saying that $x \in G$ is $\textsc{Hashable}^{[h,r]}$ if and only if $T^{[h]}(G, x)$ has depth at least h, and admits a unique spine decomposition up to height r.

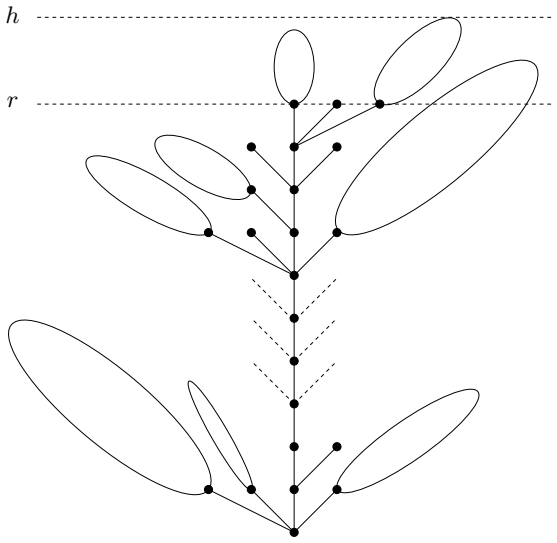

Fig. 2. A Tree of height h with a spine decomposition up height r.

Theorem 2. *There exists constants c, d such that the probability that a random tree sampled according to $\mathfrak{P}_n^{[h]}$ has a spine decomposition up to height r is greater than $1 - c \cdot (r/h) - c/r$.*

Informally speaking, this theorem means than enforcing the existence of a unique spine decomposition up to some height does not really shrink the hashable domain. For instance, one may pick $h = n \log n$ and $r = n \log \log n$. With these values, trees of height h have a unique spine decomposition up to height r asymptotically almost surely.

Theorem 3. *There is a constant $\varepsilon \in]0; 1[$ such that if two trees sampled according to $\mathfrak{P}_n^{[h]}$ have a unique spine decomposition up to height r, then the probability that they are isomorphic is upper-bounded by ε^r.*

This justifies claim *iii*. Proofs of these two theorems can be found in Appendix B. We conclude that modifying the definition of $\textsc{Hashable}(G, x)$ to only accept x if $T^{[h]}(G, x)$ has height h and a unique spine decomposition under height

r, with $h = n \log n$ and $r = n \log \log n$ is enough to make algorithm 3 work as advertised.

References

1. Agrawal, M., Saxena, N.: Equivalence of f-algebras and cubic forms. In Durand, B., Thomas, W., eds.: STACS. Volume 3884 of Lecture Notes in Computer Science., Springer (2006) 115–126
2. Aho, A.V., Hopcroft, J.E., Ullman, J.D.: The Design and Analysis of Computer Algorithms. Addison-Wesley Publishing Company (1974)
3. Alon, N., Blais, E.: Testing boolean function isomorphism. In Serna, M.J., Shaltiel, R., Jansen, K., Rolim, J.D.P., eds.: APPROX-RANDOM. Volume 6302 of Lecture Notes in Computer Science., Springer (2010) 394–405
4. Athreya, K.B., Ney, P.: Branching processes. Springer-Verlag, Berlin, New York, (1972)
5. Babai, L., Kantor, W.M., Luks, E.M.: Computational complexity and the classification of finite simple groups. In: FOCS, IEEE Computer Society (1983) 162–171
6. Babai, L., Kucera, L.: Canonical labelling of graphs in linear average time. In: FOCS, IEEE Computer Society (1979) 39–46
7. Baena, J., Clough, C., Ding, J.: Square-vinegar signature scheme. In: PQCrypto '08: Proceedings of the 2nd International Workshop on Post-Quantum Cryptography, Berlin, Heidelberg, Springer-Verlag (2008) 17–30
8. Bardet, M., Faugère, J.C., Salvy, B.: On the complexity of Gröbner basis computation of semi-regular overdetermined algebraic equations. In: Proc. International Conference on Polynomial System Solving (ICPSS). (2004) 71–75
9. Bettale, L., Faugère, J.C., Perret, L.: Cryptanalysis of the trms signature scheme of pkc'05. In Vaudenay, S., ed.: AFRICACRYPT. Volume 5023 of Lecture Notes in Computer Science., Springer (2008) 143–155
10. Billet, O., Gilbert, H.: A traceable block cipher. In Laih, C.S., ed.: ASIACRYPT. Volume 2894 of Lecture Notes in Computer Science., Springer (2003) 331–346
11. Billet, O., Macario-Rat, G.: Cryptanalysis of the square cryptosystems. In Matsui, M., ed.: ASIACRYPT. Volume 5912 of Lecture Notes in Computer Science., Springer (2009) 451–468
12. Biryukov, A., Cannière, C.D., Braeken, A., Preneel, B.: A toolbox for cryptanalysis: Linear and affine equivalence algorithms. In: EUROCRYPT. (2003) 33–50
13. Bosma, W., Cannon, J.J., Playoust, C.: The Magma Algebra System I: The User Language. J. Symb. Comput. **24**(3/4) (1997) 235–265
14. Bouillaguet, C., Faugère, J.C., Fouque, P.A., Perret, L.: Practical cryptanalysis of the identification scheme based on the isomorphism of polynomial with one secret problem. In Catalano, D., Fazio, N., Gennaro, R., Nicolosi, A., eds.: Public Key Cryptography. Volume 6571 of Lecture Notes in Computer Science., Springer (2011) 473–493
15. Clough, C., Baena, J., Ding, J., Yang, B.Y., Chen, M.S.: Square, a new multivariate encryption scheme. In Fischlin, M., ed.: CT-RSA. Volume 5473 of Lecture Notes in Computer Science., Springer (2009) 252–264
16. Cramer, R., ed.: Advances in Cryptology - EUROCRYPT 2005, 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005, Proceedings. In Cramer, R., ed.: EUROCRYPT'05. Volume 3494 of Lecture Notes in Computer Science., Springer (2005)

17. Daemen, J.: Limitations of the even-mansour construction. [34] 495–498
18. Ding, J., Wolf, C., Yang, B.Y.: -invertible cycles for multivariate quadratic public key cryptographyℓ. In Okamoto, T., Wang, X., eds.: Public Key Cryptography. Volume 4450 of Lecture Notes in Computer Science., Springer (2007) 266–281
19. Dubois, V., Fouque, P.A., Shamir, A., Stern, J.: Practical Cryptanalysis of SFLASH. In: CRYPTO. Volume 4622., Springer (2007) 1–12
20. Dubois, V., Fouque, P.A., Stern, J.: Cryptanalysis of SFLASH with Slightly Modified Parameters. In: EUROCRYPT. Volume 4515., Springer (2007) 264–275
21. Dubois, V., Granboulan, L., Stern, J.: An efficient provable distinguisher for hfe. In Bugliesi, M., Preneel, B., Sassone, V., Wegener, I., eds.: ICALP (2). Volume 4052 of Lecture Notes in Computer Science., Springer (2006) 156–167
22. Dunkelman, O., Keller, N., Shamir, A.: Minimalism in cryptography: The even-mansour scheme revisited. In Pointcheval, D., Johansson, T., eds.: EUROCRYPT. Volume 7237 of Lecture Notes in Computer Science., Springer (2012) 336–354
23. Even, S., Mansour, Y.: A construction of a cioher from a single pseudorandom permutation. [34] 210–224
24. Faugère, J.C., Joux, A., Perret, L., Treger, J.: Cryptanalysis of the hidden matrix cryptosystem. In Abdalla, M., Barreto, P.S.L.M., eds.: LATINCRYPT. Volume 6212 of Lecture Notes in Computer Science., Springer (2010) 241–254
25. Faugère, J.C., Perret, L.: Polynomial Equivalence Problems: Algorithmic and Theoretical Aspects. In Vaudenay, S., ed.: EUROCRYPT. Volume 4004 of Lecture Notes in Computer Science., Springer (2006) 30–47
26. Fortin, S.: The graph isomorphism problem. Technical report, University of Alberta (1996)
27. Fouque, P.A., Granboulan, L., Stern, J.: Differential cryptanalysis for multivariate schemes. [16] 341–353
28. Fouque, P.A., Macario-Rat, G., Perret, L., Stern, J.: Total break of the ℓ-ic signature scheme. In Cramer, R., ed.: Public Key Cryptography. Volume 4939 of Lecture Notes in Computer Science., Springer (2008) 1–17
29. Fouque, P.A., Macario-Rat, G., Stern, J.: Key Recovery on Hidden Monomial Multivariate Schemes. In Smart, N.P., ed.: EUROCRYPT. Volume 4965 of Lecture Notes in Computer Science., Springer (2008) 19–30
30. Geiger, J.: Elementary new proofs of classical limit theorems for Galton-Watson processes. J. Appl. Probab. **36**(2) (1999) 301–309
31. Geiselmann, W., Meier, W., Steinwandt, R.: An Attack on the Isomorphisms of Polynomials Problem with One Secret. Int. J. Inf. Sec. **2**(1) (2003) 59–64
32. Gligoroski, D., Markovski, S., Knapskog, S.J.: Multivariate quadratic trapdoor functions based on multivariate quadratic quasigroups. In: Proceedings of the American Conference on Applied Mathematics, Stevens Point, Wisconsin, USA, World Scientific and Engineering Academy and Society (WSEAS) (2008) 44–49
33. Goldreich, O., Micali, S., Wigderson, A.: Proofs that yield nothing but their validity and a methodology of cryptographic protocol design (extended abstract). In: FOCS, IEEE (1986) 174–187
34. Imai, H., Rivest, R.L., Matsumoto, T., eds.: Advances in Cryptology - ASIACRYPT '91, International Conference on the Theory and Applications of Cryptology, Fujiyoshida, Japan, November 11-14, 1991, Proceedings. In Imai, H., Rivest, R.L., Matsumoto, T., eds.: ASIACRYPT. Volume 739 of Lecture Notes in Computer Science., Springer (1993)
35. Joux, A., Kunz-Jacques, S., Muller, F., Ricordel, P.M.: Cryptanalysis of the tractable rational map cryptosystem. [54] 258–274

36. Kayal, N.: Efficient algorithms for some special cases of the polynomial equivalence problem. In Randall, D., ed.: SODA, SIAM (2011) 1409–1421
37. Macario-Rat, G.: Cryptanalyse de schémas multivariés et résolution du problème Isomorphisme de Polynômes. PhD thesis, Université Paris Diderot — Paris 7 (June 2010)
38. McKay, B.: Computing automorphisms and canonical labelling of graphs. In: Lecture Notes in Mathematics. (1978) 223–232
39. Miyazaki, T.: The complexity of mckay's canonical labelling algorithm. In Finkelstein, L., Kantor, W.M., eds.: Groups and computation, II. Volume 28 of DIMACS: Series in Discrete Mathematics and Theoretical Computer Science., AMS and DIMACS (1997) 239–256
40. Mohamed, M., Ding, J., Buchmann, J., Werner, F.: Algebraic attack on the mqq public key cryptosystem. In Garay, J., Miyaji, A., Otsuka, A., eds.: Cryptology and Network Security. Volume 5888 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2009) 392–401
41. Monagan, M.B., Geddes, K.O., Heal, K.M., Labahn, G., Vorkoetter, S.M., McCarron, J., DeMarco, P.: Maple 10 Programming Guide. Maplesoft, Waterloo ON, Canada (2005)
42. Pakes, A.G.: Some limit theorems for the total progeny of a branching process. Advances in Applied Probability **3**(1) (1971) 176–192
43. Patarin, J.: Hidden fields equations (hfe) and isomorphisms of polynomials (ip): Two new families of asymmetric algorithms. In: EUROCRYPT. (1996) 33–48
44. Patarin, J., Goubin, L., Courtois, N.: Improved Algorithms for Isomorphisms of Polynomials. In: EUROCRYPT. (1998) 184–200
45. Patarin, J., Goubin, L., Courtois, N.: Improved Algorithms for Isomorphisms of Polynomials – Extended Version. available at `http://minrank.org/ip6long.pdf` (1998)
46. Perret, L.: A Fast Cryptanalysis of the Isomorphism of Polynomials with One Secret Problem. [16] 354–370
47. Pointcheval, D.: A new identification scheme based on the perceptrons problem. In: EUROCRYPT. (1995) 319–328
48. Sakumoto, K.: Public-key identification schemes based on multivariate cubic polynomials. In Fischlin, M., Buchmann, J., Manulis, M., eds.: Public Key Cryptography. Volume 7293 of Lecture Notes in Computer Science., Springer (2012) 172–189
49. Sakumoto, K., Shirai, T., Hiwatari, H.: Public-key identification schemes based on multivariate quadratic polynomials. In Rogaway, P., ed.: CRYPTO. Volume 6841 of Lecture Notes in Computer Science., Springer (2011) 706–723
50. Shamir, A.: An efficient identification scheme based on permuted kernels (extended abstract). In Brassard, G., ed.: CRYPTO. Volume 435 of Lecture Notes in Computer Science., Springer (1989) 606–609
51. Stern, J.: A new identification scheme based on syndrome decoding. In Stinson, D.R., ed.: CRYPTO. Volume 773 of Lecture Notes in Computer Science., Springer (1993) 13–21
52. Stern, J.: Designing identification schemes with keys of short size. In Desmedt, Y., ed.: CRYPTO. Volume 839 of Lecture Notes in Computer Science., Springer (1994) 164–173
53. Vaudenay, S.: A Classical Introduction to Cryptography: Applications for Communications Security. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2005)
54. Vaudenay, S., ed.: Public Key Cryptography - PKC 2005, 8th International Workshop on Theory and Practice in Public Key Cryptography, Les Diablerets, Switzer-

land, January 23-26, 2005, Proceedings. In Vaudenay, S., ed.: Public Key Cryptography. Volume 3386 of Lecture Notes in Computer Science., Springer (2005)
55. Wang, L.C., Hu, Y.H., Lai, F., yen Chou, C., Yang, B.Y.: Tractable rational map signature. [54] 244–257
56. Wilf, H., Zeilberger, D.: An algorithmic proof theory for hypergeometric (ordinary and "q") multisum/integral identities. Inventiones Mathematicae **108** (1992) 575–633 10.1007/BF02100618.

A Expected Progeny and Variance

By definition the expected progeny is:

$$\mu = \sum_{k=2}^{n} p_{n,k} \left(q^k - q^2\right)$$

Via an analog of lemma 2, this can be rephrased in terms of the properties of a random linear map h. Indeed, it is shown in [21] that:

$$p_{n,k} = \mathbb{P}\left[\dim\ker D_x f = k \middle| y \in \ker D_x f\right] = \mathbb{P}\left[\dim\ker h = k \middle| x, y \in \ker h\right]$$

And therefore:

$$\mu = \left(\sum_{k=2}^{n} \mathbb{P}\left[\dim\ker h = k \middle| x, y \in \ker h\right] q^k\right) - q^2$$

The sum is in fact the expected cardinality of the kernel of a random linear map known to vanish on a fixed 2-dimensional subspace:

$$\mu = \mathbb{E}\left[\operatorname{card}\ker h \middle| x, y \in \ker f\right] - q^2$$

Thus, to establish the expression of μ, we determine the expected cardinality of the kernel of a random linear map h known to vanish on a fixed subspace F of dimension s. Even though this seems to be an elementary question, we could not find the result in the existing literature.

Lemma 4. *Let h be a uniformly random endomorphism of $(\mathbb{F}_q)^n$, vanishing on a subspace F of $(\mathbb{F}_q)^n$, with $\dim F = s$. Then:*

$$\mathbb{E}\left[\operatorname{card}\ker h \middle| F \subseteq \ker f\right] = q^s + 1 - \frac{1}{q^{n-s}}$$

This lemma establishes the expression of μ (and we postpone its proof a little bit). Let us now turn our attention to the variance σ^2:

$$\begin{aligned}
\sigma^2 &= \left[\sum_{k=2}^{n} p_{n,k} \left(q^k - q^2\right)^2\right] - \mu^2 \\
&= \left(\sum_{k=2}^{n} p_{n,k} \cdot q^{2k}\right) - 2q^2 \left(\sum_{k=2}^{n} p_{n,k} \cdot q^k\right) + q^4 - \mu^2 \\
&= \left(\sum_{k=2}^{n} p_{n,k} \cdot q^{2k}\right) - \left(\sum_{k=2}^{n} p_{n,k} \cdot q^k\right)^2
\end{aligned}$$

Thanks to the relation between $p_{n,k}$ and random linear maps outlined above, we see that σ^2 is in fact exactly the variance of the cardinality of the kernel of a random linear map known to vanish on two fixed vectors.

Lemma 5. *Let h be a uniformly random endomorphism of $(\mathbb{F}_q)^n$, vanishing on a subspace F of $(\mathbb{F}_q)^n$, with $\dim F = s$. Then the variance of the cardinality of its kernel is:*

$$q^s(q-1)\left(1 - \frac{q^s+1}{q^n} + \frac{q^s}{q^{2n}}\right)$$

This establishes the expression of σ^2. We know give the proofs of the two lemma.

Proof (of lemma 4).

$$\begin{aligned} E_n = \mathbb{E}\left[\operatorname{card}\ker f \middle| F \subseteq \ker f\right] &= \sum_{k=s}^{n} \mathbb{P}\left[\dim\ker f = k \middle| F \subseteq \ker f\right] q^k \\ &= \sum_{k=s}^{n} \frac{\lambda(n)\lambda(n-s)}{\lambda(k)\lambda(k-s)\lambda(n-k)} q^{-k(k-s)} q^k \end{aligned}$$

A combinatorial and/or elementary argument completely eluded us. We therefore use the method of "creative telescoping" to establish the result by induction on n. First, we notice that the announced results holds when $n = s$. Let us therefore assume $n > s$. We denote by $T(n,k,s)$ the hairy term under the sum. It is a q-hypergeometric term because if we set $X = q^n$ and $Y = q^k$, we see that the two following ratios are rational functions of X and Y:

$$\begin{aligned} \frac{T(n+1,k,s)}{T(n,k,s)} &= \frac{q^2X^2 - (q+q^{s+1})X + q^s}{q^2X^2 - qXY} \\ \frac{T(n,k+1,s)}{T(n,k,s)} &= q^{s+2}\frac{X+Y}{X\left(qY-q^s\right)\left(qY-1\right)} \end{aligned}$$

We thus used the q-analog of Zeilberger's algorithm [56] (as implemented in Maple [41]), and it found the nice recurrence relation:

$$a \cdot T(n+1,k,s) - b \cdot T(n,k,s) = g(n,k+1,s) - g(n,k,s) \qquad (\star)$$

where:

$$\begin{aligned} a &= q^{n+1} + q^{n+s+1} - q^{s+1} \\ b &= q^{n+1} + q^{n+1+s} - q^s \\ g(n,k,s) &= \frac{\left(q^k - q^s\right)\left(q^k - 1\right)\left(q^{n+s+1} - q^{n+s+2} - q^{k+s} + q^{n+k+1} + q^{n+k+s+1}\right)}{q^{2k}\left(q^{n+1} - q^k\right)} T(n,k,s) \end{aligned}$$

The point is that summing $(\star)$ over $k = s, \ldots, n-1$ yields:

$$a\left(E_{n+1} - T(n+1,n+1,s) - T(n+1,n,s)\right) - b\left(E_n - T(n,n,s)\right) = g(n,n,s) - g(n,s,s)$$

At this point, it is easy to find that $g(n,s,s)=0$, and we check (using a computer algebra system!) that:

$$g(n,n,s)+a\cdot(T(n+1,n+1,s)+T(n+1,n,s))+b\cdot T(n,n,s)=0$$

Thus, we have established that:

$$\left(1+q^s-\frac{1}{q^{n-s}}\right)E_{n+1}=\left(1+q^s-\frac{1}{q^{n+1-s}}\right)E_n$$

Thus, if the result holds at rank n, then it also holds at rank $n+1$. □

Proof (of lemma 5). The variance is:

$$V_n=\underbrace{\sum_{k=s}^{n}\left(\frac{\lambda(n)\lambda(n-s)}{\lambda(k)\lambda(k-s)\lambda(n-k)}q^{-k(k-s)}\right)q^{2k}}_{U_n}-\left(q^s+1-\frac{1}{q^{n-s}}\right)^2$$

We will first demonstrate by induction on $n\geq s$ that:

$$U_n=q^{2s}+1+(1+q)\left(q^s-\frac{1}{q^{n-s}}-\frac{1}{q^{n-2s}}\right)+\frac{1}{q^{2n-1-2s}}\qquad(\clubsuit)$$

When $n=s$, we should have $U_n=q^{2n}$, and looking at ($\clubsuit$) carefully reveals that our expression of U_n simplifies to this value. Let us therefore assume $n>s$, and let us again denote by $T(n,k,s)$ the hairy term under the sum. It is again a q-hypergeometric term, and running the q-analog of Zeilberger's algorithm yields:

$$a\cdot T(n+1,k,s)-b\cdot T(n,k,s)=g(n,k,s)-g(n,k+1,s)\qquad(\star)$$

where:

$$a=-q^{n+s+2}+q^{s+1+2n}+q^{1+2n}+q^{2s+2}-q^{2s+n+1}-q^{s+1+n}-q^{2s+2+n}+q^{2s+2n+1}+q^{s+2+2n}$$

$$b=-q^{1+2n}+q^{n+s}-q^{s+1+2n}+q^{s+1+n}-q^{2s}+q^{2s+n+1}+q^{2s+n}-q^{2s+2n+1}-q^{s+2+2n}$$

g is a complicated term with a singularity when $n+1=k$. We again notice that $g(n,s,s)=0$ and that:

$$a\cdot T(n+1,n+1,s)+a\cdot T(n+1,n,s)-b\cdot T(n,n,s)=g(n,n,s)$$

So that summing ($\star$) over $k=s,\ldots,n-1$ and exploiting the previous equation yields:

$$a\cdot U_{n+1}=b\cdot U_n$$

By induction hypothesis, ($\clubsuit$) holds at rank n. Plugging the expression of U_n into this recurrence relation and simplifying shows that ($\clubsuit$) holds at rank $n+1$ — please use a computer algebra system if you really want to verify this. Moving back to the expression of V_n, it is not difficult to verify that the result of the lemma holds. □

B Isomorphism of Random Trees

For any $n \geq 3$, let $\mathfrak{T}$ be a tree sampled according to $\mathfrak{P}$ (*i.e.*, with offspring distribution ℓ), and let $\mathfrak{P}^{[h]}$ be the law of $\mathfrak{T}$ conditioned to have height at least h.

In this section, all quantities depend on n (the random tree $\mathfrak{T}$, the law $\mathbb{P}^{[h]}$, the offspring distribution ℓ, the height h, etc.), but we do not always make this dependency explicitly visible by writing subscripts or superscripts, in order to make notations less cumbersome. In addition, we also write $\mathbb{P}^{[h]}[\cdot]$ instead of $\mathbb{P}\left[\cdot \middle| \textsc{Height}(\mathfrak{T}) \geq h\right]$.

We need a criterion to decide whether two conditioned trees are isomorphic or not, and we need it to be simple enough so that we may evaluate the probability that it holds. The criterion we will use is the following: two isomorphic trees with a unique spine decomposition must have empty subtrees emanating from the backbone at the exact same heights. Of course, if the spine decomposition is unique up to height r, then this holds only up to height r. This will intuitively show that two random trees with a unique spine decomposition up to height r are isomorphic with a probability that gets exponentially small in r. We will make this intuition formal later, but we must first introduce some properties of the spine decomposition.

We decompose a conditioned tree (*i.e.*, a tree of law $\mathfrak{P}^{[h]}$) into a *backbone* (or *spine*) going from the root to height h, on which we graft a given number of unconditioned Galton-Watson trees at each of its nodes. Looking at all nodes of height r, if only one of them has descendants at height h then the spine up to height r is uniquely determined: necessarily, it is the path in the tree going from the root to this node (fig. 2 illustrates this).

Let us work for a moment with ordered Galton-Watson trees. That is, we also record who is the descendant of each parent and offspring are ordered (so that we can talk about brothers to the left or to the right of an individual). In [30], Geiger shows that if we define the sequence of independent random variables $(V_m, Y_m), m \in \mathbb{N}$ by

$$\mathbb{P}\left[V_m = j, Y_m = k\right] = \frac{\mathbb{P}\left[\textsc{Height}(\mathfrak{T}) \geq m-1\right]}{\mathbb{P}\left[\textsc{Height}(\mathfrak{T}) \geq m\right]} \cdot \mathbb{P}\left[\textsc{Height}(\mathfrak{T}) < m-1\right]^{j-1} \cdot \ell(k),$$

for $1 \leq j \leq k < \infty$, then $\mathfrak{T}_n$ conditioned to have height at least h has the same law as the random tree constructed inductively as follows:

- The root (i.e., the first node of the spine) has Y_h offspring.
- To each of the $V_h - 1$ first offspring node we graft a Galton-Watson tree with offspring distribution ℓ and conditioned to have height (strictly) less than $h-1$. These $V_h - 1$ trees are independent of each other (and of the rest of the construction). These subtrees are on the left of the backbone on fig. 3.
- To each of the $Y_h - V_h$ last offspring, we graft an unconditioned Galton-Watson tree with offspring distribution ℓ (again, these trees are independent of each other and of the rest of the construction). These subtrees are on the right of the backbone on fig. 3.

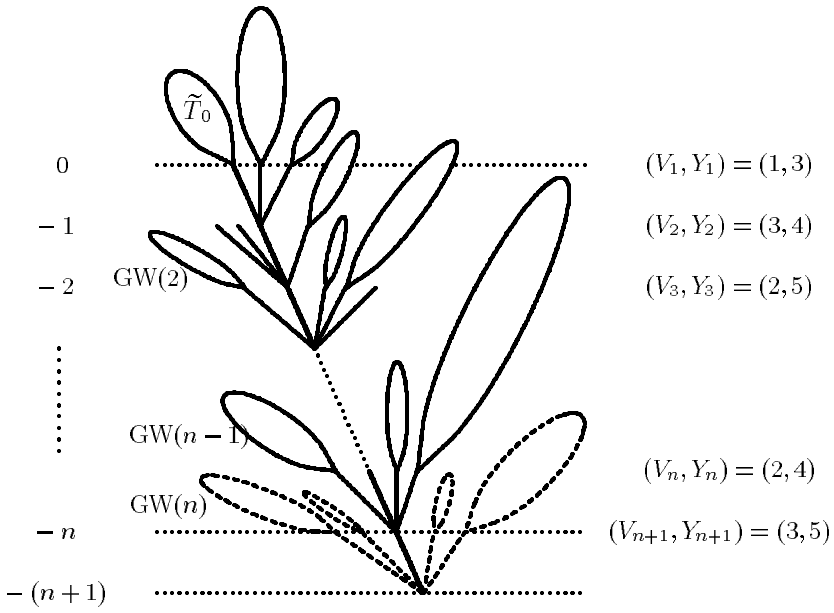

Fig. 3. Illustration of the spine decomposition (this is Figure 1 from [30]). This shows the Galton-Watson tree conditionned on non-extinction at generation n and $n+1$ respectively. $GW(k)$ denotes a Galton-Watson tree conditioned to be extinct at generation k. The subtrees to the right of the line of descent of the left-most particle are ordinary Galton-Watson trees.

– The V_h-th offspring node continues the spine. It has Y_{h-1} offspring, the first V_{h-1} ones are the roots of i.i.d. Galton-Watson trees conditioned to have height less than $h-2$, the last $Y_{h-1}-V_{h-1}$ are the roots of i.i.d. unconditioned Galton-Watson trees and the spine carries on with the V_{h-1}-th offspring, which has Y_{h-2} offspring nodes, and so on.

Observe that the marginal distribution of Y_m is given by

$$\mathbb{P}\left[Y_m = y\right] = \frac{1-\mathbb{P}\left[\textsc{Height}(\mathfrak{T}) < m-1\right]^y}{\mathbb{P}\left[\textsc{Height}(\mathfrak{T}) \geq m\right]} \cdot \ell(y), \tag{1}$$

The spine can be seen as a "prolific" line of descent that survives up to generation h by producing a biased number of offspring, while the other individuals of the population reproduce essentially according to the initial offspring distribution (we refer to [30] for an explanation of the fact that trees emanating from brothers to the left of the spine are conditioned not to have descendants at generation h).

Proof (proof of theorem 2). We show that in a tree sampled according to $\mathfrak{P}^{[h]}$, with high probability only one path from the root to height r extends to a path reaching height h. Call this event A. Since this property is purely topological, then it does not matter whether the tree is ordered or not. We obtain the desired result by bounding from below the probability of A by the probability that all trees emanating from the spine under height r are of height less than $h-r$. The independence of this family of trees, together with the fact (easy to check) that for every integer i in the interval $\{1,\dots,r-1\}$

$$\mathbb{P}\left[\textsc{Height}(\mathfrak{T}) < h-r \middle| \textsc{Height}(\mathfrak{T}) < h-i\right] \geq \mathbb{P}\left[\textsc{Height}(\mathfrak{T}) < h-r\right],$$

enables us to write

$$\begin{aligned}\mathbb{P}\left[A\right] &\geq \prod_{i=0}^{r-1} \mathbb{E}\left[\mathbb{P}\left[\textsc{Height}(\mathfrak{T}) < h-r\right]^{Y_{h-i}-1}\right] \\ &\geq \mathbb{E}\left[\mathbb{P}\left[\textsc{Height}(\mathfrak{T}) < h-r\right]^{\sum_{i=0}^{r-1} Y_{h-i}}\right]. \end{aligned} \tag{2}$$

Now, as $n \to +\infty$, all the $p_{n,k}$ (for $k \in \{3,\dots,n\}$) converge to a finite limit $p_{\infty,k}$, the expected progeny μ converges to 1 (recall that $\mu < 1$ for every n), and finally the variance σ_n^2 converges to $q^3 - q^2$. The last two convergences happen exponentially fast in n, therefore the same proof as that of Theorem 3.1 in [30] (in which $\mu = 1$ for all n) shows that whenever $(m_n)_{n\geq 1}$ tends to infinity at most polynomially, we have

$$\lim_{n\to\infty} m_n \cdot \mathbb{P}\left[\textsc{Height}(\mathfrak{T}) \geq m_n\right] = \frac{2}{\sigma^2}. \tag{3}$$

Furthermore, we have the following lemma.

Lemma 6. *There exist constants $C_3, C_4 > 0$ such that for every $n \geq 3$,*

$$\mathbb{P}\left[\sum_{i=0}^{r-1} Y_{h-i} > rC_3\right] \leq \frac{C_4}{r}.$$

We postpone the proof of Lemma 6 until the end of the proof of Theorem 2. Armed with (3) and Lemma 6, we can come back to (2) and write for every n

$$\begin{aligned}\mathbb{P}[A] &\geq \mathbb{E}\left[\mathbb{P}[\textsc{Height}(\mathfrak{T}) < h-r]^{rC_3} \cdot \mathbf{1}_{\left\{\sum_{i=0}^{r-1} Y_{h-i} \leq C_3 r\right\}}\right] \\ &\geq \left(1 - \frac{C_4}{\sigma^2 h}\right)^{rC_3} \times \mathbb{P}\left[\sum_{i=0}^{r-1} Y_{h-i} \leq rC_3\right] \\ &\geq e^{-\frac{r}{h}C_6}\left(1 - \frac{C_4}{r}\right) \\ &\geq 1 - \frac{r}{h}C_7. \end{aligned} \tag{4}$$

Note that for the third inequality, use the fact that $1 - x \geq e^{-2x}$ for every $x \in [0, 1/2]$. What (4) shows is that for every $n \geq 3$, if we sample a Galton-Watson tree $\mathfrak{T}$ according to $\mathfrak{P}^{[h]}$, then with probability at least $1 - C_7\frac{r}{h}$ there will be a unique spine decomposition under height r.

Proof (of lemma 6). We use Markov's inequality (in a Chebychev-like fashion) as follows: if $C_3 > 0$, we have for each $n \geq 3$

$$\begin{aligned}\mathbb{P}\left[\sum_{i=0}^{r-1} Y_{h-i} > rC_3\right] &= \mathbb{P}\left[\sum_{i=0}^{r-1} (Y_{h-i} - \mathbb{E}[Y_{h-i}]) > C_3 \cdot r - \sum_{i=0}^{r-1} \mathbb{E}[Y_{h-i}]\right] \\ &\leq \frac{\mathbb{E}\left[\left(\sum_{i=0}^{r-1} (Y_{h-i} - \mathbb{E}[Y_{h-i}])\right)^2\right]}{\left(C_3 \cdot r - \sum_{i=0}^{r-1} \mathbb{E}[Y_{h-i}]\right)^2}. \end{aligned} \tag{5}$$

Let us show that the numerator in the right-hand side of (5) is of order r, while the denominator is of order r^2 whenever $C_3 > 0$ is large enough. These two points rely on appropriate bounds on the first two moments of all Y_{h-i}'s (observe that the numerator is in fact the sum of the variances of the Y_{h-i}'s). Indeed, recall from (1) that for every $k \in \{3, \ldots, n\}$,

$$\mathbb{P}\left[Y_{h-i} = q^k - q^2\right] = \frac{1 - \mathbb{P}[\textsc{Height}(\mathfrak{T}) < h - i - 1]^{q^k - q^2}}{\mathbb{P}[\textsc{Height}(\mathfrak{T}) \geq h - i]} \cdot p_{n,k}$$

and these are the only possible values for Y_{h-i}. Because $1 - e^{-x} \leq x$ for all $x \geq 0$, we can write for every $i \leq r - 1$:

$$\begin{aligned}1 - \mathbb{P}[\textsc{Height}(\mathfrak{T}) < h - i - 1]^{q^k - q^2} &\leq -\left(q^k - q^2\right) \log \mathbb{P}[\textsc{Height}(\mathfrak{T}) < h - i - 1] \\ &\leq -\left(q^k - q^2\right) \log \mathbb{P}[\textsc{Height}(\mathfrak{T}) < h - r - 2]. \end{aligned}$$

We thus have for every such integer i

$$\frac{1 - \mathbb{P}[\textsc{Height}(\mathfrak{T}) < h - i - 1]^{q^k - q^2}}{(q^k - q^2) \cdot \mathbb{P}[\textsc{Height}(\mathfrak{T}) \geq h - i]} \leq -\frac{\log \mathbb{P}[\textsc{Height}(\mathfrak{T}) < h - r - 2]}{\mathbb{P}[\textsc{Height}(\mathfrak{T}) \geq h]}.$$

Moreover, because $\lambda(\cdot)$ is decreasing,

$$\frac{\lambda(n)\lambda(n-2)}{\lambda(k)\lambda(k-2)\lambda(n-k)} \leq \lim_{n\to\infty} \frac{1}{\lambda(n)} =: C_q.$$

Combining the above, we arrive at

$$\mathbb{P}\left[Y_{h-i} = q^k - q^2\right] \leq -\frac{\log \mathbb{P}\left[\textsc{Height}(\mathfrak{T}) < h-r-2\right]}{\mathbb{P}\left[\textsc{Height}(\mathfrak{T}) \geq h\right]} \cdot C_q \cdot \left(q^k - q^2\right) q^{-k(k-2)}$$

for every $n \geq 3$ and $k \in \{2, \ldots, n\}$. This yields

$$\mathbb{E}\left[Y_{h-i}\right] \leq -\frac{\log \mathbb{P}\left[\textsc{Height}(\mathfrak{T}) < h-r-2\right]}{\mathbb{P}\left[\textsc{Height}(\mathfrak{T}) \geq h\right]} \cdot C_q \cdot \sum_{k=3}^{n} \left(q^k - q^2\right)^2 q^{-k(k-2)}$$

$$\mathbb{E}\left[\left(Y_{h-i}\right)^2\right] \leq -\frac{\log \mathbb{P}\left[\textsc{Height}(\mathfrak{T}) < h-r-2\right]}{\mathbb{P}\left[\textsc{Height}(\mathfrak{T}) \geq h\right]} \cdot C_q \cdot \sum_{k=2}^{n} \left(q^k - q^2\right)^3 q^{-k(k-2)}$$

Now, by (3) we have

$$\lim_{n\to\infty} -\frac{\log \mathbb{P}\left[\textsc{Height}(\mathfrak{T}) < h-r-2\right]}{\mathbb{P}\left[\textsc{Height}(\mathfrak{T}) \geq h\right]} = 1,$$

and furthermore,

$$\sum_{k=3}^{\infty} \left(q^k - q^2\right)^2 q^{-k(k-2)} =: m_1 < \infty \qquad \text{and} \qquad \sum_{k=3}^{\infty} \left(q^k - q^2\right)^3 q^{-k(k-2)} =: m_2 < \infty.$$

As a consequence, there exists $C > 0$ such that for every $n \geq 3$, we have

$$\sum_{i=0}^{r-1} \mathbb{E}\left[Y_{h-i}\right] \leq C m_1 r,$$

and (using the independence of all Y_m's)

$$\mathbb{E}\left[\left(\sum_{i=0}^{r-1} Y_{h-i} - \mathbb{E}\left[Y_{h-i}\right]\right)^2\right] = \sum_{i=0}^{r-1} \mathrm{Var}\left(Y_{h-i}\right) \leq rC',$$

for a constant $C' > 0$ depending on m_1 and m_2. Choosing $C_3 > Cm_1$ and coming back to (5), we obtain the existence of $C_4 > 0$ such that for every $n \geq 3$,

$$\mathbb{P}\left[\sum_{i=0}^{r-1} Y_{h-i} \geq rC_3\right] \leq \frac{C_4}{r}.$$

This completes the proof of Lemma 6. □

Proof (proof of theorem 3). Let us use again $\mathfrak{T}$ (from the proof of theorem 2) and its spine decomposition under the additional conditioning that all trees emanating from the spine under height r are of height smaller than $h-r$. We write $\tilde{\mathbb{P}}^{[h]}[\cdot]$ as a shorthand for this conditionnal probability. By construction, each brother of the i-th node of the spine ($0 \leq i \leq r-1$) has no offspring with probability

$$e := \mathbb{P}\left[\mathfrak{T} = \emptyset \middle| \textsc{Height}(\mathfrak{T}) \leq h-r\right] = \frac{\mathbb{P}\left[\mathfrak{T} = \emptyset\right]}{\mathbb{P}\left[\textsc{Height}(\mathfrak{T}) \leq h-r\right]} = \frac{\ell_n(0)}{\mathbb{P}\left[\textsc{Height}(\mathfrak{T}) \leq h-r\right]}. \tag{6}$$

Brother to the right or to the left does not matter here since the condition at the denominator is stronger than $\textsc{Height}(\mathfrak{T}) < h-i-1$ for our range of integers i. Let us use (6) to obtain some bounds (away from 0 and 1), uniform in n and $i \leq r-1$, for the probability that all of the $Y_{h-i}-1$ brothers of the i-th node of the spine have zero offspring. Because $\ell(0) = p_{n,2}$ and using (3), the right-hand side of (6) is equivalent as $n \to \infty$ to

$$\frac{\ell(0)}{1-2/(\sigma^2 \cdot h)} \underset{n\to\infty}{\simeq} \lim_{n\to\infty} \frac{\lambda(n)}{\lambda(2)} =: \mathbf{e} \in]0,1[. \tag{7}$$

Thus, if we denote $\alpha = \tilde{\mathbb{P}}^{[h]}\left[\text{no nephews at height } i\right]$, then by definition

$$\alpha \geq \mathbb{P}\left[Y_{h-i} = q^3 - q^2\right] \cdot (e)^{q^3-q^2-1}$$

Using (1) and (3),

$$\alpha \geq \frac{1 - \left(1 - \dfrac{2}{\sigma^2(h-i-1)} + o\left(\frac{1}{h}\right)\right)^{q^3-q^2}}{\dfrac{2}{\sigma^2(h-i)} + o\left(\frac{1}{h}\right)} \cdot p_{n,2} \cdot (e)^{q^3-q^2-1}$$

The fraction is equal to $q^3 - q^2 + o(1/(h))$, and given the expression of $p_{n,3}$ as well as (7), the lower bound on α is equivalent to

$$\frac{q^3-q^2}{q^3} \cdot \frac{\mathbf{e}^{q^3-q^2-1}}{\lambda(1)\lambda(3)} \cdot \prod_{j=1}^{\infty}\left(1 - \frac{1}{q^j}\right) = \mathbf{e}^{q^3-q^2-1} \prod_{j=4}^{\infty}\left(1 - \frac{1}{q^j}\right) \in]0,1[.$$

Likewise,

$$\begin{aligned}\tilde{\mathbb{P}}^{[h]}[\text{at least one nephew at height } i] &\geq \mathbb{P}\left[Y_{h-i} = q^3 - q^2\right] \left(1 - (e)^{q^3-q^2-1}\right) \\ &\simeq (1 - \mathbf{e}^{q^3-q^2-1}) \prod_{j=4}^{\infty}\left(1 - \frac{1}{q^j}\right) \in]0,1[.\end{aligned}$$

Hence, since these two probabilities belong to $]0,1[$ for all $n \geq 3$ and $i \leq r-1$, and belong to a smaller interval of $]0,1[$ bounded away from 0 and 1 whenever

n is large enough, this provides the existence of $\kappa_l, \kappa_u \in]0,1[$ such that for every $n \geq 3$ and $i \in \{0, \dots, r-1\}$,

$$1 - \kappa_l \leq \tilde{\mathbb{P}}^{[h]}[\text{no nephews at height } i] \leq \kappa_u. \tag{8}$$

Now, let $\mathfrak{T}, \mathfrak{T}'$ be two trees of height at least h and such that their spine decompositions are unique under height r. For every $i \in \{0, r-1\}$, let γ_i (resp. ${\gamma'}_i$) be the indicator function of the event that all brothers of the i-th node of the spine have no offspring. It follows from the properties of the spine decomposition that for every $n \geq 3$, $\{\gamma_i,\ 0 \leq i \leq r-1\}$ form a family of independent random variables and by (8), we have

$$\tilde{\mathbb{P}}^{[h]}\big[\gamma_i = 1\big] \leq \kappa_u \qquad \text{and} \qquad \tilde{\mathbb{P}}^{[h]}\big[\gamma_i = 0\big] \leq \kappa_l.$$

Comparing the absence or presence of nephews of the spine in $\mathfrak{T}$ and in $\mathfrak{T}'$, and defining the constant $\kappa = \max(\kappa_l, \kappa_u) < 1$, we obtain:

$$\tilde{\mathbb{P}}^{[h]}[\mathfrak{T} = \mathfrak{T}'] \leq \kappa^r.$$

□