Why you cannot even hope to use Gröbner Bases in Public Key Cryptography

An open letter to a scientist who failed

and a challenge to those who have not yet failed

Boo Barkee, Deh Cac Can, Julia Ecks, Theo Moriarty, R.F. Ree *

> In the magical art of Steganography, there is nothing frivolous, nor contrary to the Gospels and the Catholic faith; nor have we taught superstitious beliefs. Everything is based on natural, lawful and honest principles; the mystery which veils the precepts of this art and the names of the spirits, requires a cultivated reader; to hide the secrets of this art, which could be harmful if made known to wicked men, we avail ourselves of the services of the spirits.
> Johannes Trithemius, *Steganographia*
>
> The air is influenced by astral emanations. So one can, naturally and without spiritual help, communicate his thoughts to another man, however large the distance between them. This I have seen done, I did myself and was done by Trithemius... In the same way, one can broadcast in the air any image, however far, by means of mirrors... The image will be, through large distances, seen by a conscious reader in the lunar disc; this artifice was used by Pythagoras.
> H.C. Agrippa *De Occulta Philosophia*
>
> It's better to have loved and lost
> than to have liked and tied for second
> Anonymous

Dear Deluded Author,

you are proposing to use the fact that Gröbner bases are hard to compute to devise a public key cryptography scheme. We are firmly convinced, instead, that no scheme using Gröbner bases will ever work. The following notes are an attempt to explain why.

Let us start by recalling the basic facts related to Gröbner bases (*cf.* [**B, BWK**]). One has an ideal $I \subset k[X_1, \ldots, X_n]$ (where k is a field) and a well-ordering compatible with product on the semigroup $\mathbf{T}$ of terms (monic monomials) in $k[X_1, \ldots, X_n]$. This ordering allows to represent uniquely each $f \in k[X_1, \ldots, X_n]$ as an ordered linear combination of elements of $\mathbf{T}$:

$$f = \sum_{i=1}^{r} c_i t_i \quad c_i \in k \setminus \{0\}, t_i \in \mathbf{T}, t_1 > \cdots > t_r$$

so to each non-zero element $f \in k[X_1, \ldots, X_n]$, we can associate $T(f) := t_1$, the *maximal term* of f; its coefficient c_1 is called the *leading coefficient* of f and denoted $lc(f)$.

Moreover, the ordering allows to associate to the ideal I the semigroup ideal $T(I) := \{T(f) \in \mathbf{T} : f \in I \setminus \{0\}\} \subset \mathbf{T}$, and its complement, the *order ideal* $O(I) := \mathbf{T} \setminus T(I)$.

The most important fact which can be derived by this setting is the following:

* Postal address: Prof. B. Barkee, Math White Hall, Cornell University, Ithaca NY 14853
E-mail: 100142.3240@compuserve.com
Partially supported by SPECTRE

Fact

1) $k[X_1, \ldots, X_n] = I \oplus Span_k(O(I))$.

2) There is a k-vector space isomorphism between $k[X_1, \ldots, X_n]/I$ *and* $Span_k(O(I))$.

3) For each $f \in k[X_1, \ldots, X_n]$ *there is a unique* $g := Can(f, I) \in Span_k(O(I))$ *s.t.* $f - g \in I$.

Moreover:

a) $Can(f, I) = Can(g, I)$ *if and only if* $f - g \in I$.

b) $Can(f, I) = 0$ *if and only if* $f \in I$.

Remark immediately a very important point, which is obscured in most presentations of Gröbner bases: the canonical form of an element is defined just in terms of the ideal and of the ordering; conceptually the notion of a Gröbner basis is not needed to define canonical forms. In fact, we have not yet defined Gröbner bases: we are going to do it now.

A *Gröbner basis* of I is a set of generators $G := \{g_1, \ldots, g_s\} \subset I$ s.t. $T(G) := \{T(g_1), \ldots, T(g_s)\}$ generates $T(I)$.

If a Gröbner basis G of I is known, given $f \in k[X_1, \ldots, X_n]$, $Can(f, I)$ can be computed by the following algorithm (Buchberger Reduction):

Red(f, G)
  $h := 0$
  While $f \neq 0$
    If $T(f) \in T(G)$ **then**
      choose $g \in G$ **s.t.** $T(f) = tT(g)$
      $f := f - lc(f)lc(g)^{-1}tg$
    else
      $h := h + lc(f)T(f), f = f - lc(f)T(f)$
  $Can(f, I) := h$

which is a procedure analogous to Gaussian reduction in vector spaces and whose cost is therefore quadratic in the size of the input polynomial f, i.e., in a dense representation, in the number of terms which are not greater than $T(f)$. If the ordering is compatible with degree, i.e. $deg(t_1) < deg(t_2) \Longrightarrow t_1 < t_2$, this number is $\mathbf{O}(d^n)$ where $d = deg(f)$.

On the other side, the Buchberger Algorithm to compute a Gröbner basis of an ideal I, knowing a basis $\{f_1, \ldots, f_t\}$ of I, has a worst-case complexity $d^{2^{\mathbf{O}(n)}}$, where $d = maxdeg(f_i)$.

Notwithstanding the complexity result above, the basic assumption of your paper, i.e. that Gröbner bases are hard to compute, is based on a misunderstanding.

It is true that there are ideals whose Gröbner bases have elements whose degree is doubly exponential in the degrees of the input basis; but such examples are rare, with bad algebraic properties, and absolutely *not* random, while randomness must obviously be somehow part of your scheme.

In fact in one of the two papers which at least partially settled the complexity of the Buchberger algorithm ([**G**]), the following is proved:

Theorem 1 *"Most" of the ideals generated by* s *polynomials in* n *variables of degree bounded by* d *are such that their Gröbner bases have degree bounded by* $(n+1)d - n$.

Most here means that coefficients are randomly chosen and that the result holds except for a set of measure zero in the space parametrizing the coefficients.

The major misunderstanding of your paper is however confusing the problem of deciding ideal membership with the problem of computing Gröbner bases; a solution to the second problem gives a solution to the first one, but an easier solution to the first problem could be at hand. It is this easier solution which allows to break a Gröbner cryptographic scheme.

Let us begin by describing a basic cryptographic scheme using Gröbner bases; this scheme, or variants of it, has popped up many times and never reached existence, since we dutifully explained its authors the attacks we are going to describe below.

Somehow Archibald has produced an ideal $I \subset k[X_1, \ldots, X_n]$ of which he secretly owns a hard-to-compute Gröbner basis, so allowing him to compute canonical forms in polynomial time.

Archibald makes public a set of terms $\mathcal{T} \subset O(I)$ (either the whole of it, or, for added security, a subset of it) and a set of low-degree polynomials $\{g_1, \ldots, g_l\} \subset I$, which are a basis of either I or of some ideal properly contained in it.

When Balthazar wants to send Archibald a message, he encodes it as a linear combination $M = \sum_{t_i \in \mathcal{T}} c_i t_i$. The polynomial M therefore satisfies $M = Can(M, I)$.

To encrypt it, Balthazar produces randomly polynomials $p_1, \ldots, p_l$ and broadcasts the polynomial $C := M + \sum_{i=1}^{l} p_i g_i$.

Since $C - M \in I$, $Can(C, I) = Can(M, I) = M$, so Archibald can use his secret Gröbner basis to compute $M = Can(C, I)$.

What our f(r)iend Fantomas knows is $\mathcal{T}$, the g_i's and C; but he also knows that M, while unknown to him, is the canonical form of C.

Before describing our attacks, we need a few more assumptions, in order to have a complexity measure: we assume that everything is dense, i.e. we assume that Archibald makes public:

1) l *dense* polynomials of degree at most d in n variables, $g_1, \ldots, g_l$.
2) A set of monomials $\mathcal{T} = \{m_1, \ldots, m_s\} \subset O(I)$ of degree at most d.

To send a message, Balthazar chooses each p_i to be a dense polynomial of degree r. C is therefore a dense polynomial of degree $R \leq D := d + r$ [1] ; if we denote by τ the number of terms of degree at most D, the complexity of Balthazar's encoding and Archibald's decoding is therefore between $\mathbf{O}(\tau)$ and $\mathbf{O}(\tau^2)$ [2].

The value of r is probably a public parameter, but Fantomas would not be hampered even if it were a secret choice of Balthazar. In fact, because of randomness in the choice of p_i, there will be only a few cancellations in the sum $\sum_{i=1}^{l} p_i g_i$, so that Fantomas has a good guess of r too.

Fantomas has another and stronger advantage: because of the uniqueness of the canonical form, Fantomas doesn't need to find the same p_i's used by Balthazar; any choice of q_i's s.t. $C = M + \sum_{i=1}^{l} q_i g_i$ is equally fine for him. In particular, Fantomas can look for the minimal degree representation $C = M + \sum_{i=1}^{l} q_i g_i$. If we set $D' := maxdeg(q_i g_i)$, one has $R \leq D' \leq D = d + r$; because of randomness, however, it is to be expected that $R = D' = D$.

To break the system, Fantomas can now use the following result about Gröbner bases ([**DFGS**]), which definitely shows that the ideal membership problem is not necessarily as hard as computing a Gröbner basis.

Theorem 2 *Let $I = (g_1, \ldots, g_l)$ and let h be s.t. $deg(h) \leq D$, $h - Can(h, I) = \sum_{i=1}^{l} p_i g_i$ with $deg(p_i g_i) \leq D$.*

Let $\mathcal{G}$ be the output of the Buchberger algorithm, modified so that each computation involving polynomials of degree higher than D is not performed.

Then $Can(h, I)$ can be computed by Buchberger reduction of h via $\mathcal{G}$.

As a consequence, Fantomas computes a Gröbner basis of I postponing all computations involving polynomials of degree higher than $R = deg(C)$, thus obtaining a set $\mathcal{G}$; then he computes $h := \mathbf{Red}(C, \mathcal{G})$.

If $h \in Span_k(\mathcal{T}) \subset O(I)$, then $h = Can(C, I) = M$, since $C - h \in I$ and because of the uniqueness of the canonical form.

Otherwise, Fantomas knows that his guess $R = deg(C)$ for D' was underestimated. He then makes the guess $D' = R + 1$. Of course he doesn't have to restart the Gröbner basis computation from scratch, but just perform those computations of degree $R + 1$ which had been postponed in the first run, together with any new computation of degrees not larger than $R + 1$. Also, at the end of the second run, he will not reduce C but h.

Because of this, the computations performed by Fantomas are exactly the same as if he knew the actual value D' and ran the algorithm above just once, skipping computations of degree higher than D'.

---

[1] D is the degree C would have if no cancellation occurs and if there is at least a g_i of degree d, such that the degree of the corresponding random p_i is exactly r. That's exactly what one should expect to happen.

[2] to achieve this complexity in decoding, either Balthazar has to choose a degree-compatible ordering or he must use some clever version of Buchberger reduction. Since this has nothing to do with security, and is a bit involved, we don't enter into many details and we just briefly discuss the case in which the ordering is degree-compatible: because of density, Archibald has to scan all τ terms for reducing them if possible, which gives the $\mathbf{O}(\tau)$ lower bound; since each reduction costs $\mathbf{O}(\tau)$, the upper bound is obtained too.

Remark that the algorithm above gives Fantomas the knowledge of exactly that part of the Gröbner basis which is needed to decode *any* possible message. It can be proved that its complexity is $\mathbf{O}(\tau^4)$, see the Appendix.

The attack our f(r)iend James Moriarty (the one who wrote *A treatise on the binomial theorem* [**M**]) favoured, did not even use the notion of Gröbner bases, but just old, plain linear algebra, and it had even cubic complexity! Its main importance lies in the fact that it shows plainly that the connection between Gröbner bases and the ideal membership problem is much weaker than currently believed.

Like Fantomas, Moriarty makes a first reasonable guess on D', choosing R, and solves the polynomial equation $M + \sum_{i=1}^{l} q_i g_i = C$, where the q_i's are polynomials of degree $R - deg(g_i)$ with unknown coefficients; this is a linear system of equations whose unknowns are the coefficients of M and of the q_i's and whose equations are just the coefficients of each term in the polynomial $M + \sum_{i=1}^{l} q_i g_i - C = 0$.

Once solved it, either:

- there are solutions; then even if there are more than one, the coefficients of M are uniquely determined (because M is unique) and J.M. reads the message
- there are no solution; the guess on D' has proved wrong (otherwise Balthazar's input would be a solution); J.M. then starts again with a higher guessed value; however, since the initial subset of equations is the same as before, using say gaussian elimination, the job already done is part of the job he should do now, so he can skip it. Therefore, as in Fantomas' attack, this doesn't really increase the complexity which is $\mathbf{O}(\tau^3)$.

While we have discussed how Moriarty can decode a single message, it is evident that matrix inversion allows him to break the system in cubic time; so avoiding the use of Gröbner bases gives you an order of advantage[3].

And now the challenge.

Both Fantomas and Moriarty attacks are correct and efficient for the basic Gröbner scheme we have outlined above.

The underlying ideas are that to solve a bounded ideal membership problem, one just needs a partial Gröbner basis and one can even avoid the use of it by making recourse to linear algebra.

The high complexity of Gröbner bases is in fact strictly related with the existence of polynomials in an ideal whose minimal degree representation in terms of a given basis is doubly exponential in the degree of the basis elements. Since such polynomials cannot be used as encoded messages, a cryptographic scheme applying the complexity of Gröbner bases to an ideal membership problem is bound to fail.

Is our reader able to find a scheme which overcomes this difficulty?

In particular our reader could think (perhaps with some reason) that a *sparse* scheme could work. We believe (perhaps without reason) that sparsity will make the scheme easier to crack. We would be glad to test our belief on specific sparse schemes.

References

[**BWK**] T. Becker, V. Weispfenning, in cooperation with H. Kredel *Gröbner bases*, Springer Verlag (1993)

[**B**] B.Buchberger, *Gröbner bases: an algorithmic method in polynomial ideal theory*, in N.K. Bose (ed.) *Recent trends in multidimensional system theory*, Reidel (1985)

[**DFGS**] A. Dickenstein, N. Fitchas, M. Giusti, C. Sessa, *The membership problem for unmixed polynomial ideals is solvable in single exponential time*, Discrete Applied Mathematics **33** (1991) 73–94

[**G**] M. Giusti, *Some effectivity problems in polynomial ideal theory*, EUROSAM 84, J. Fitch (Ed.), Springer L.N.C.S. 174 (1984) 159–171

[**M**] J.Moriarty, *A treatise on the binomial theorem*, s.l., s.d.

[3] In fact, by fast linear algebra algorithms, Moriarty could even break the system in $\mathbf{O}(\tau^\omega)$, where currently $\omega = 2.4\ldots$, provided he knows the value of r.
But Moriarty is a conservative so, as he doesn't like to use Gröbner bases, he doesn't like to use fast linear algebra as well.

Appendix

We collect here those technical results which are needed for a proper understanding of Fantomas attack. Of course, we begin by recalling a (simplified) version of the Buchberger algorithm.

Given $f, g \in k[X_1, \ldots, X_n] \setminus \{0\}$ we denote:

$$S(f,g) := lc(f)^{-1} \frac{l.c.m.(T(f), T(g))}{T(f)} f - lc(g)^{-1} \frac{l.c.m.(T(f), T(g))}{T(g)} g$$

$$deg(f,g) = maxdeg(\frac{l.c.m.(T(f), T(g))}{T(f)} f, \frac{l.c.m.(T(f), T(g))}{T(g)} g)$$

Theorem 3 (*cf.* [**BWK**], Theorem 5.62, Cor. 5.63) *For $G \subset I$, the following conditions are equivalent:*

1) *G is a Gröbner basis of I.*
2) *Each $f \in I$, $f \neq 0$, has a representation $f = \sum p_i g_i$, $g_i \in G$, with $T(f) = T(p_1 g_1) > \ldots > T(p_i g_i) > T(p_{i+1} g_{i+1}) > \ldots$, which is called a strong Gröbner representation*
3) *G is a basis of I and $\forall f, g \in G$, $S(f,g)$, if not zero, has a strong Gröbner representation.*

Proof: 1) $\Rightarrow$ 2) Buchberger reduction shows in fact that for each $f \in k[X_1, \ldots, X_n]$, $f - Can(f, I)$ has a strong Gröbner representation.

2) $\Rightarrow$ 1) The definition of Gröbner basis is immediately verified since for each $f \in I$, $T(f)$ is then a multiple of $T(g_1)$.

2) $\Rightarrow$ 3) This is obvious since $S(f,g) \in I$.

3) $\Rightarrow$ 2) f has clearly at least a representation $f = \sum p_i g_i$, where w.l.o.g. $T(p_1 g_1) \geq T(p_2 g_2) \geq \ldots$; if it is not a strong Gröbner representation, there are at least two summands $p_i g_i$, $p_j g_j$, s.t. $T(p_i g_i) = T(p_j g_j) =: t$. To each representation which is not Gröbner, we can therefore associate a term t, to be the highest term s.t. there are i, j, with $T(p_i g_i) = T(p_j g_j) = t$, and an integer $m \geq 2$ to be the total number of summands $p_k g_k$ s.t. $T(p_k g_k) = t$.

We now assume that f has no strong Gröbner representation; then among the representations of f for which t is minimal, we choose one with minimal m. Let this representation be $f = \sum p_i g_i$ and let $i, i+1, \ldots, i+m-1$ be the indexes s.t. $T(p_{i+j} g_{i+j}) = t$.

Denote $t_i := \frac{l.c.m.(T(g_i), T(g_{i+1}))}{T(g_i)}$, $t_{i+1} = \frac{l.c.m.(T(g_i), T(g_{i+1}))}{T(g_{i+1})}$.

Clearly there is τ s.t. $T(p_i) = \tau t_i$

Moreover the assumption implies that $S(g_i, g_{i+1}) = lc(g_i)^{-1} t_i g_i - lc(g_{i+1})^{-1} t_{i+1} g_{i+1} = \sum q_\lambda h_\lambda$ with $h_\lambda \in G$ and $T(t_i g_i) = T(t_{i+1} g_{i+1}) > T(q_1 h_1) > \ldots$, i.e.

$$lc(g_i)^{-1} t_i g_i = lc(g_{i+1})^{-1} t_{i+1} g_{i+1} - \sum q_\lambda h_\lambda$$

$$lc(g_i)^{-1} T(p_i) g_i = lc(g_{i+1})^{-1} \tau t_{i+1} g_{i+1} - \sum \tau q_\lambda h_\lambda$$

are both strong Gröbner representations

Therefore if we substitute in the chosen representation $T(p_i) g_i$ with the representation above, either t will decrease (in case $m = 2$ and $lc(p_i g_i) = lc(p_{i+1} g_{i+1})$) or m will decrease.

In both cases we have a contradiction with the minimality of the chosen representation. ∎

Buchberger algorithm to compute a Gröbner basis G of I, given a basis F of I is a direct consequence of the theorem above:

$G =$ **Gröbner**(F)
  $G := F$
  $B := \{\{f, g\} : f, g \in G\}$
  While $B \neq \emptyset$ **do**
    Choose $\{f, g\} \in B$
    $B := B \setminus \{\{f, g\}\}$
    $h :=$ **Red**$(S(f,g), G)$
    If $h \neq 0$ **then**
      $B := B \cup \{\{f, h\} : f \in G\}$

$G := G \cup \{h\}$

The modified algorithm of Theorem 2 is obtained by not performing the computation and the reduction of those $S(f,g)$ s.t. $deg(f,g) > D$. Its proof is an easy modification of the one of Theorem 3: one has just to consider only the representations of $f - Can(f,I)$ in terms of the basis $\mathcal{G}$ (since it contains the input basis, such representations exist); once a minimal one is chosen, one has just to remark that if $maxdeg(p_i g_i) \leq D$, then $deg(g_i, g_{i+1}) \leq D$, so $S(g_i, g_{i+1})$ has a strong Gröbner representation in terms of $\mathcal{G}$. Once a strong Gröbner representation of $f - Can(f,I)$ in terms of $\mathcal{G}$ is found, it is obvious that $Can(f,I) = \mathbf{Red}(f,\mathcal{G})$.

The complexity evaluation is also easy; if τ is the number of terms of degree at most D, there are at most τ elements in $\mathcal{G}$, $\mathbf{O}(\tau^2)$ elements in B, whose reduction has a $\mathbf{O}(\tau^2)$ cost.