On the Generation of Multivariate Polynomials which are Hard to Factor

Adi Shamir
Department of Applied Mathematics
The Weizmann Institute of Science, Rehovot 76100, Israel
email: shamir@wisdom.weizmann.ac.il

Abstract. In this paper we consider the difficulty of factoring multivariate polynomials $F(x, y, z, \ldots)$ modulo n. We consider in particular the case in which F is a product of two randomly chosen polynomials P and Q with algebraically specified coefficients, and n is the product of two randomly chosen primes p and q. The general problem of factoring F is known to be at least as hard as the factorization of n, but in many restricted cases (when P or Q are known to have a particular form) the problem can be much easier. The main result of this paper is that (with one trivial exception), the problem of factoring F is at least as hard as the factorization of n whenever P and Q are randomly chosen from the same sample space, regardless of what may be known about its form.

1. Introduction

If you pick two random primes p and q and multiply them together, the result is believed to be very difficult to factor. Even though this complexity is unproven (both in the worst case sense and in the average case sense), this widely accepted assumption serves as a basis for the security of several cryptographic schemes (such as the RSA — see Rivest, Shamir and Adleman [1978]).

In this paper we consider the following related problem: If you pick two random multivariate polynomials $P(x, y, z, \ldots)$ and $Q(x, y, z, \ldots)$ and multiply them together, is the result hard to factor? We are interested in particular in the case where the polynomial product $F = PQ$ is computed modulo a numeric product $n = pq$, and try to relate the difficulty of factoring F with the difficulty of factoring n.

Our interest in this problem arose in an attempt to construct new cryptographic schemes, which are more efficient than the RSA. They are based on the problem of factoring multivariate polynomials modulo a composite n, and we were surprised to discover that this natural problem received little attention in the literature. There are many papers on the problems of factoring such polynomials over Z, Q, or over finite fields, but not over rings such as Zn. It is easy to prove that in the worst case the problem of factoring F cannot be easier than the problem of factoring n, but there seems to be no analysis of:

1. The average case complexity of the problem (F may be easy to factor for random choices of P and Q).
2. The problem of factoring multivariate polynomials of particular forms (e.g., when only certain monomials appear in the given polynomial F).
3. The problem of factoring multivariate polynomials when side information is available (e.g., that the unknown factors P and Q have a particular form).

All these problems arise naturally in cryptographic applications: worst case results are meaningless, compactly representable multivariate polynomials are desirable since they lead to smaller keys and more efficient implementations, and side information can sometimes be extracted by exploiting weaknesses in the protocols. However, the problems are also of great interest in the context of computer algebra systems, which are often required to factor large multivariate polynomials with specialized forms.

Here are some simple examples which illustrate the intricacies of this problem:

1. $P = (x + ay)$, $Q = (x - ay)$, where a is a randomly chosen constant in $[0, n)$. The factorization of the product $F = (x^2 - a^2y^2)$ $\pmod n$ requires square root extraction, and thus it is as difficult as the factorization of the modulus n.

2. $P = (x + ay)$, $Q = (x + ay)$. The modified product $F = (x^2 + 2axy + a^2y^2) \pmod n$ becomes trivial to factor, since a occurs explicitly in the coefficient of xy. This demonstrates that whereas the extraction of square roots of numbers mod n is provably as difficult as the factorization of the modulus, the extraction of square roots of polynomials mod n can be very easy.

3. $P = (x^3 + ax + b)$, $Q = (x^3 + cx + d)$, where a, b, c, d are randomly chosen constants in $[0, n)$. The factorization of their product $F = (x^6 + (a + c)x^4 + (b + d)x^3 + acx^2 + (ad + bc)x + bd) \pmod n$ is as difficult as the factorization of n.

4. $P = (x^3 + ax + b)$, $Q = (x^4 + cx + d)$. This slightly modified product is trivial to factor: Given $F = (x^7 + ax^5 + (b + c)x^4 + dx^3 + acx^2 + (ad + bc)x + bd)$, we can obtain a and d by inspection, and then b and c by division.

5. $P = (ax + by + cz + \ldots)$, $Q = (dx + ey + fz + \ldots)$, where $a, b, \ldots$ are randomly chosen constants in $[0, n)$. The factorization of their product $F \pmod n$ is as difficult as the factorization of the modulus n whenever the number of variables is at least 2.

6. In example 5, assume that the cryptanalyst knows that $a = 0$. This side information makes the factorization of F trivial. On the other hand, if the cryptanalyst knows that $a = 1$, the problem remains hard.

7. The single variable case of example 5, $P = ax$ and $Q = dx$, is trivial to factor. The factorization remains easy when $P = ax$ and $Q = (dx + 1)$, but becomes as difficult as the factorization of n when $P = (ax + 1)$ and $Q = (dx + 1)$.

8. Let P and Q be two randomly chosen polynomials of the form $(a^2x^3y + b^3xz + (a + b^2)yz)(ax + by + 8z) + 1$. In spite of the complex relationships between the various coefficients and the fact that the factorization problem can be translated into 35 equations in only 4 unknown parameters, the general result proven in this paper shows that the problem of factoring $F = PQ \pmod n$ is at least as difficult as the factorization of the modulus n.

Before we proceed, a word of caution: The ring of (univariate or multivariate) polynomials over a field is known to be a unique factorization domain, and we have a good intuition about what the factors can look like. This intuition can be misleading when we consider polynomials over the ring Zn, which has zero divisors. Consider, for example, the simplest conceivable polynomial $F(x) = x$. It is clearly irreducible over any field or over the ring of integers Z. However, x *is reducible over* Zn for $n = pq$, and can be written as the product of the two nontrivial polynomials $P = c(px + q)$ and $Q = (qx + p) \pmod n$ where $c = (p^2 + q^2)^{-1} \pmod n$. This curious counterexample demonstrates that the factors of F need not have a lower degree, and raises the question whether there are *any* irreducible polynomials modulo a composite n (for example, can we continue to factor these P and Q?). Later in this paper we'll prove that these P and Q are in fact irreducible, and provide a complete characterization of all the irreducible multivariate polynomials modulo a composite n.

2. Definition of the Problem

Definition: An *algebraic form* is a multivariate polynomial in the variables $x, y, \ldots$ whose coefficients are rational functions (i.e., ratios of polynomials) in the free parameters $a, b, \ldots$. An *algebraic collection of polynomials* (mod n) is the set of all the polynomials in the variables $x, y, \ldots$ with numeric coefficients which can be obtained from an algebraic form by substituting values in $[0, n)$ for all the free parameters $a, b, \ldots$. This collection can be considered as a probability space by making all the choices of the parameters equally likely. With some abuse of notation, we denote both the form and the collection by the same letter C.

Example: $C = (a^2x + (b/a^2)y^2 + (a+b)^2xy)$ is an algebraic form. The n^2 possible substitutions of values in $[0, n)$ for a and b define an algebraic collection of polynomials. Some polynomials of the form $(ix + jy^2 + kxy)$ do not occur at all in this collection (e.g., when i is a quadratic non-residue mod n), some polynomials are chosen several times (e.g., when $b = 0$ there are four choices of a which result in the same polynomial), some polynomials are chosen only once (e.g., whenever a and b are invertible (mod n)), and some polynomials are ignored (e.g., whenever a is not relatively prime to n and thus b/a^2 is ill defined). As a result, the number of polynomials in C is smaller than n^2, and they have a non-uniform probability distribution.

Definition: An algebraic form C is called *monic* if the coefficient of its leading monomial (under lexicographic order) is 1, and *trivial* if its collection of polynomials consists of a single polynomial.

Definition: An *algebraic factorization problem* consists of two monic algebraic forms C_1 and C_2 with the same modulus n. They define a probability space of factorization problems, in which random polynomials P and Q are chosen from the respective probability spaces C_1 and C_2, and multiplied together mod n. The product F is then presented as input to the factorization algorithm, whose goal is to find some P' in C_1 and some Q' in C_2 whose product mod n is F.

Remarks:

1. For the sake of simplicity, we assume in this paper that the syntactic complexity of the polynomials (i.e., the number of variables, the number of terms, the degrees, etc) is fixed, and the only parameter which is allowed to grow asymptotically is the value of n. However, with carefully defined extensions, the results remain valid even when the syntactic complexity of the polynomials is allowed to change as well.

2. The goal of the factorization problem is not necessarily to find the original P and Q, since F may have several indistinguishable factorizations. However, we insist that the factors found should have the proper algebraic form. For example, if P and Q have the algebraic form $(x^2 + ax)$, it is trivial to factor the product $(x^4 + (a' + a'')x^3 + a'a''x^2)$ into the product of x^2 and $(x^2 + (a' + a'')x + a'a'')$, but the factorization will be considered illegal except in the unlikely case $a'a'' = 0$ (mod n) (in which the two factors are of the form $(x^2 + ax)$ for the particular values of $a = 0$ and $a = (a' + a'')$, respectively). As a result, the algebraic factorization problem may be difficult even if each one of P and Q can be factored independently.

3. The goal of the factorization is to find P' and Q', but not necessarily the values of the parameters $a, b, \ldots$ in their algebraic definitions. For example, if the coefficient of x in P' is a^2, it suffices to compute a^2 rather than to find the value of the underlying parameter a by root extraction.

4. The definition of the factorization problem allows algebraic relationships between the various coefficients in P, and algebraic relationships between the various coefficients of Q, but does not allow related parameters in P and Q. For example, the problem of factoring $(x + ay + a^2z)(bx + b^2y + z)$ (mod n) with independently chosen a and b is admissible, but both the easy factoring problem $(x + ay)(x + ay)$ (mod n) and the hard factoring problem $(x + ay)(x - ay)$ (mod n) are inadmissible problems in our framework. In Section 5 we extend the set of admissible

problems to include such examples as well.

3. Irreducible Factors Modulo A Composite

As demonstrated in the introduction, even the simplest polynomial $F = x \pmod{n}$ can be factored into $(p^2+q^2)^{-1}(px+q)(qx+p) \pmod{n}$. In this section we clarify the situation by characterizing the irreducible polynomials mod n, and upper bound the number of ways in which PQ can be factored mod n. This bound will be crucial in the proof of the main result.

Due to the Chinese remainder theorem, the ring Zn can be viewed as the direct product of the two finite fields Zp and Zq. Each polynomial $F \pmod{n}$ can thus be viewed as a pair of polynomials (F_1, F_2) where $F_1 = F \pmod{p}$ and $F_2 = F \pmod{q}$. In particular, the polynomial $F = x \pmod{n}$ can be viewed as the pair of polynomials (x, x). Since the product of such pairs is computed componentwise, we can write (x, x) as the product of $(x, 1)$ and $(1, x)$. By using the Chinese remainder theorem it is easy to show that this is essentially the factorization demonstrated above.

Lemma 1: The set of irreducible multivariate polynomials mod n is exactly the set of polynomials whose pair form is either $(P, 1)$ or $(1, Q)$, where P is irreducible mod p and Q is irreducible mod q (ignoring multiplication by constants).

Proof: We first show that $(P, 1)$ is irreducible. If not, it can be written as $(P, 1) = (P', Q')(P'', Q'')$, and thus $P = P'P'' \pmod{p}$ and $1 = Q'Q'' \pmod{q}$. However, P was assumed to be irreducible mod p, and thus either P' or P'' are a constant. Since Zq is a field, 1 can only be the product of constants mod q, and thus both Q' and Q'' are constants. By chinese remaindering two constants mod p and mod q we get a constant mod n, and thus one of the two factors of $(P, 1)$ is a constant and the factorization is trivial. A similar proof shows that $(1, Q)$ is also irreducible.

For any other multivariate polynomial (P, Q) where neither P nor Q is a constant, the factorization $(P, Q) = (P, 1)(1, Q)$ is nontrivial, and proves that (P, Q) is reducible mod n. QED.

Corollary 2: All the irreducible polynomials mod n have non-constant coefficients which are either multiples of p or multiples of q. Consequently, there are no monic irreducible polynomials mod n, and knowing any non-monic irreducible polynomial mod n is equivalent to factoring n.

Proof: If P is irreducible mod n, then all its non-constant coefficients (including its leading coefficient) are obtained by chinese remaindering a non-zero coefficient from $P \pmod{p}$ with a zero coefficient from $1 \pmod{q}$ or vice versa. Such a coefficient is either a multiple of p or a multiple of q. QED.

This corollary is the main reason we restrict the algebraic factorization problem of F to the recovery of P' and Q' rather than to a complete factorization of F into irreducible factors — otherwise the problem becomes uninteresting.

Over fields such as Zp and Zq, a polynomial of degree d can be factored into at most d factors. Over the ring Zn this bound is no longer true. For example, the polynomial $P = x \ (mod\ n)$ of degree d=1 can be written as the product of two linear factors. However, it is easy to establish an upper bound on the number of possible factorizations:

Lemma 3: A monic multivariate polynomial F which has a constant degree d can have at most a constant number 2^{2d} of factorizations into a pair of monic polynomials $F = PQ \pmod{n}$.

Proof: Modulo a prime, a monic polynomial F of degree d can be uniquely factored into at most d irreducible monic factors. Modulo n, F has at most $2d$ irreducible factors of the form $(P, 1)$ and $(1, Q)$. The number of ways in which F can be split into a product of two monic polynomials cannot exceed the number of partitions of the set of irreducible factors of F, which is 2^{2d}. QED.

Remarks:

1. Without normalizing the polynomials to monic form, there can be exponentially many

ways to write F as a product of P and Q, which differ by their multiplicative constants.

2. The number of partitions which result in P and Q of the desired algebraic form can be much smaller, but there are examples of C_1 and C_2 for which the number of proper factorizations is exponential in d. When d is allowed to grow logarithmically with the size of n, the number of factorizations is polynomial, and the proof of our main result remains valid. When n is the product of more than two primes, we can again handle constant or logarithmic number of factors, but not cases in which this number grows faster than $O(\log(|n|)) = O(\log\log n)$.

4. The Main Result

Theorem 4: Any algebraic factorization problem with non-trivial $C_1 = C_2$ is at least as hard as the factorization of the modulus n.

This can be viewed as the proper extension from numbers to polynomials of the statement that "extracting square roots modulo a composite is equivalent to factoring". The obvious extension (i.e., computing P from $P^2 \pmod n$) is clearly inadequate, since in many cases it is actually an easy computational task. Instead, we consider the case where the algebraic form of the two polynomials is the same, but each one is chosen with independent random parameters. The only exception is the case of trivial algebraic forms, since we can easily show:

Lemma 5: If either C_1 or C_2 is trivial, the algebraic factorization problem is solvable in polynomial time.

Proof: A trivial form generates only one polynomial P, which can be found by substituting random values into the parameters of the given algebraic form. Once P is known, we can divide the given polynomial F by P in polynomial time, and obtain Q. By definition, it has the proper algebraic form. QED.

One possible approach to the proof of the main theorem is to embed a numeric square root computation in the polynomial factorization problem, so that the later cannot be easier than the former. This can be done in some particularly simple cases, but for arbitrary algebraic factorization problems it seems to be very difficult, since the many known relationships between the unknown parameters can supply a lot of side information.

Our approach is based on the invariance of the factorization problem with respect to the order of P and Q: F can be written both as $PQ \pmod n$ and as $QP \pmod n$. Since P and Q are assumed to be random members of the same algebraic collection, these two factorizations are (information theoretically) indistinguishable.

Let us consider now the transformation which switches the positions of P and Q modulo p, and keeps their positions modulo q. By Chinese remaindering $P \pmod p$ with $Q \pmod q$ we get one polynomial $S \pmod n$, and by Chinese remaindering $Q \pmod p$ with $P \pmod q$ we get another polynomial $T \pmod n$. It is easy to show:

Lemma 6: If $C_1 = C_2 = C$, then $F = ST \pmod n$, and S and T also belong to C.

Proof: F is clearly equal to ST modulo both p and q, and thus also modulo n. To prove that S belongs to C, Chinese remainder the values mod p of the parameters $a', b', \ldots$ used to define P with the values mod q of the parameters $a'', b'', \ldots$ used to define Q. The results $a, b, \ldots$ will define S. A similar proof shows that T also belongs to C. Note the crucial role of the algebraic nature of the relationships between the coefficients — the result will no longer be true if the coefficients of P would be related to each other by bit reversal, for example. QED.

If someone knows both the PQ and the ST factorizations of F, he can try to factor n by using the fact that P is equal to S mod p but P is not likely to be equal to S mod q. One technical difficulty is that the later is not guaranteed (unlike the case of numeric root extractions, when the two roots $+r$ and $-r \pmod p$ can never be the same for $r \neq 0$). However, we can use the following classical property of low degree multivariate polynomials:

Lemma 7: A multivariate polynomial $G(a, b, \ldots)$ of constant degree d modulo a large prime q is either everywhere zero or almost nowhere zero.

We need the following corollary:

Corollary 8: Two independent random assignments of values to the free parameters $a, b, \ldots$ of a multivariate rational function $R(a, b, \ldots)$ of constant degree d modulo a large prime q lead with overwhelming probability to two different values, unless $R(a, b, \ldots)$ is equal to some constant v_0 whenever it is defined.

Proof: Express the rational function $R(a, b, \ldots)$ as the ratio of two low degree multivariate polynomials $R'(a, b, \ldots)/R''(a, b, \ldots)$. Assume that it is not a constant, and let v_1 be its value for one random assignment. Then the low degree polynomial $G(a, b, \ldots) = R'(a, b, \ldots) - v_1 R''(a, b, \ldots) \pmod p$ is not identically zero, and thus a second random assignment to the free parameters is extremely unlikely to make G and R'' zero. We can thus divide G by R'' mod q, and get with overwhelming probability a defined value v_2 of R which is different than v_1. QED.

Proof of Theorem 4: Assume the existence of a "magic box" which can factor the products of two randomly chosen monic polynomials from C. If we pick a random product $PQ \pmod n$ and give it as input to the magic box, there are at most 2^{2d} monic factorizations which can possibly be returned as output. Since ST is one of them and it is (information theoretically) indistinguishable from the original factorization PQ, the probability that all of P, Q, S, T will become known is at least 2^{-2d}, which is assumed to be a constant. The algebraic form C was assumed to be non-trivial, and thus at least one of the coefficients of the monomials in it is a rational function which is non-constant modulo at least one of p and q. Assume without loss of generality that the last coefficient in C is a non-constant rational function $R(a, b, \ldots)$ modulo q. The corresponding numeric coefficients in P mod q and in S mod q are derived from two independent random assignments of values to the free parameters $a, b, \ldots$ in R, and by Corollary 8 they are different with overwhelming probability. On the other hand, these numeric coefficients in P mod p and S mod p are equal by definition, and thus the GCD of n and the difference between the values of these coefficients in P mod n and in S mod n is likely to be p. By repeating this experiment sufficiently many times for the same n and randomly chosen pairs of multivariate polynomials P and Q from C, we can factor n with overwhelming probability. QED.

Remarks:

1. Note again the difference between the numeric and polynomial versions of the problem: a number a can have at most 4 square roots mod $n = pq$, whereas a multivariate polynomial can have many more factorizations modulo the same n. In fact, if we let d grow with the size of n, then with overwhelming probability the known and computed factorizations of F would be different modulo both p and q, and thus n would not be split.
2. The proof can be easily extended to cases in which the magic box manages to factor a non-negligible fraction of its inputs (rather than all of them). This just increases the expected number of repetitions required in order to factor n.
3. The general proof fails when P and Q are chosen from different algebraic forms C_1 and C_2, since the two factorizations become distinguishable and can no longer be "half mixed". This is a real rather than a technical problem, as demonstrated by the ease with which products such as $(x^3 + ax + b)(x^4 + cx + d) \pmod n$ can be factored.

5. The Difficulty of Solving Systems of Algebraic Equations Modulo A Composite n

An alternative way of looking at an algebraic factorization problem is to consider the set of modular equations obtained by equating the algebraic coefficients of PQ with the numeric coefficients of F.

Example: If C is defined as $(x+ay+bz) \pmod n$, then the factorization problem of $(x+a'y+b'z)(x+a''y+b''z) \pmod n$ is equivalent to solving the five equations:

$$a'+a''=i_1, \quad b'+b''=i_2, \quad a'a''=i_3,$$
$$b'b''=i_4, \quad a'b''+a''b'=i_5 \pmod n$$

where $i_1, \ldots i_5$ are the given coefficients of F. Note that this is a system of five equations in only four unknowns, which can be simplified to a system of three equations in two unknowns by exploiting the linearity of two of the equations. As proven in the previous section, its solution is at least as difficult as the factorization of n.

The difficulty of solving systems of algebraic equations is a major research area in mathematics and computer science, and there are many known results. The diophantine case of solving one polynomial equation in 13 unknowns over Z is undecidable (Matijasevic and Robinson [1975]). Solving one quadratic equation in two variables over N is NP-complete (Manders and Adleman [1978]). Solving a system of quadratic equations modulo a prime p is NP-complete even when $p=2$ (Fraenkel and Yesha [1977]), but small systems of equations can be solved efficiently by Grobner base techniques. Solving one quadratic equation in one unknown modulo a composite n is as difficult as the factorization of n, but solving two quadratic equations in one unknown is easy (since the solver can eliminate the non-linear term x^2 from the two equations). Some of the techniques developed for solving such systems of equations are quite sophisticated. For example, Pollard developed surprisingly efficient heuristics for solving one quadratic equation in two unknowns mod n even when the factorization of n is unknown, and extended the technique in a limited way to one cubic equation in three unknowns.

Once again, most of the negative results (which prove the unsolvability or the NP-completeness of the problem) are worst case complexities, and little seems to be known about the complexities of solving random systems of equations of some particular form. Our main result (recast as an equation solving problem) can be viewed as one step in this direction.

Can we extend the proof technique from factorization-based equations to more general types of algebraic equations? Our proof technique relied on five basic ingredients:

1. Invariance. Square root extraction modulo a composite n is difficult since squaring $(\bmod\ n)$ is invariant under negation. Multivariate polynomials F are hard to factor into $PQ \pmod n$ since the product is invariant under order. There are many other examples of invariance which we may try to exploit.

2. Modularity. The operation should be applicable not only mod n, but also separately modulo each one of its factors p and q. The chinese remainder theorem makes this possible whenever the operation is a rational function defined in terms of +,-,*,/.

3. Invertibility. The input and output of the operation should have the same probability of being chosen. This is easy to achieve if the operation is uniquely invertible (with the possible exception of a negligible subset of inputs, where the output may be undefined). In the case of negation, both v and $-v$ are uniformly distributed. In the case of commutativity, PQ and QP are equally likely choices of factors.

4. Non-triviality. The input and output of the operation should be different with non negligible probability. In our examples, v and $-v$ are different except for the unlikely choice of $v=0$, and the pairs (P,Q) and (Q,P) are different with overwhelming probability whenever the algebraic form C is non-trivial.

5. Boundedness. The size of all the equivalence classes under the operation should grow at most polynomially with the size of n. In the case of squaring modulo n each equivalence class contains at most four values, and in the case of polynomial multiplication each equivalence class contained at most 2^{2d} monic polynomials.

Our goal now is to use these ideas in order to prove that solving additional types of multivariate rational equations is at least as difficult as factoring the modulus n. We first define:

Definition: A system of multivariate rational equations of fixed degree d of the form $R_i(a, b, \ldots) = v_i \pmod{n}$ for $i = 1 \ldots k$ is said to be *randomly solvable* if the v_i values are computed by substituting uniformly distributed random values for the variables $a, b, \ldots$ in the R_i functions.

Randomly solvable systems of equations have fixed left-hand sides, and a (possibly non-uniform) probability distribution on their right-hand sides. They are presented without the underlying choice of values for $a, b, \ldots$, and the goal is to find some values for $a, b, \ldots$ which satisfy the k equations. The existence of at least one solution is guaranteed by the construction.

The main result in this section is:

Theorem 9: If a randomly solvable system of equations is invariant under some invertible rational transformation of the variables which is not the identity, and the system is known to have at most a polynomial number of solutions, then finding any one of them is at least as difficult as the factorization of the modulus n.

The proof is essentially the same as the proof of Theorem 4, and is left to the reader.

Examples:

1. The factorization problem can be viewed as a special case of this theorem: Each coefficient in F gives rise to an equation which remains invariant if we exchange the unknown variables $a', b', \ldots$ defining P with the corresponding unknown variables $a'', b'', \ldots$ defining Q, and Lemma 3 provides an upper bound on the number of solutions.
2. The transformation does not have to affect all the variables. For example, $a^2 + b^2 + abc + ad + bd = v$ remains invariant if we exchange a and b, but leave c and d unchanged.
3. The transformation can be more general than exchanging pairs of variables. For example, the equation $a^2 + (a + b)^2 = v$ remains invariant under the invertible linear transformation which replaces a by $a + b$ and b by $-b$.
4. The transformation can be rational rather than linear. For example, the equation $a^3 - b^{-3} = v$ remains invariant if we replace a by $-1/b$ and b by $-1/a$.

As demonstrated in these examples, there are many types of operations which satisfy the invariance, modularity, invertibility, and non-triviality conditions. The tricky part in applying Theorem 9 is to bound the number of solutions of the given system of equations. We expect the number of solutions to be small if the equations are "randomly looking" and there are more equations than variables. Further evidence can be obtained by actually solving the given equations modulo several random moduli n' with known factorizations, and counting the number of solutions. However, such heuristic arguments can not be used to actually prove that solving a given system of equations modulo a given n with unknown factorization is as hard as the factorization of the modulus. What makes factorization-based equations special is that the unique factorization theorem provides an automatic upper bound on their number of solutions, and there is no need to apply a case-by-case analysis.

The generalized technique developed in Theorem 9 can greatly extend the class of multivariate polynomials whose factorization can be proven to be as difficult as the factorization of the modulus. For example, the hard factorization problem of $F = (x + ay)(x - ay) \pmod{n}$ was considered an inadmissible case in our original formulation, since the free parameters in P and Q were not independently chosen. However, mapping a to $-a$ is an invertible rational transformation which exchanges the definitions of P and Q and is not the identity. The difficulty of factoring such a F can thus be deduced from our extended formulation. We can not use the same technique to prove the (false) difficulty of factoring $F = (x + ay)(x + ay) \pmod{n}$ since the mapping which exchanges P and Q is the identity mapping, which is disallowed. Another example which was beyond the scope of our original formulation is the problem of factoring $F = (x^2 + ay + bz)(x^2 - ay + (a+b)z)$ (mod n). The simultaneous mapping of a to $-a$ and of b to $a+b$ is an invertible rational transformation which exchanges the definitions of P and Q and is not the identity, and thus the problem of factoring such F is also as difficult as the factorization of the modulus.

Bibliography:

1. A. S. Fraenkel and Y. Yesha[1977], "Complexity of Problems in Games, Graphs, and Algebraic Equations", unpublished manuscript.

2. K. Manders and L. Adleman[1978], "NP-Complete Decision Problems for Binary Quadratics", J. Comput. System Sci. 16, 168-184.

3. Y. Matijasevic and J. Robinson[1975], "Reduction of an arbitrary Diophantine Equation to One in 13 Unknowns", Acta Arith. 27, 521-553.

4. R. Rivest, A. Shamir and L. Adleman[1978], " A Method for Obtaining Digital Signatures and Public Key Cryptosystems", Comm. ACM 21, 120-126.