NTNU
Norwegian University of
Science and Technology

Noncommutative Gröbner bases in Polly Cracker cryptosystems

Andreas Helde

Master of Science in Physics and Mathematics
Submission date: June 2009
Supervisor: Aslak Bakke Buan, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem Description

Fellows and Koblitz have suggested a cryptosystem based on Gröbner-bases for commutative polynomial rings where we consider NP-complete problems. There exists several known attacks on these kind of cryptosystems, and it appears to be very hard to efficiently create a secure system such that finding a Gröbner-basis for an ideal is infeasible. Tapan S. Rai and others have come up with a generalization of this cryptosystem with non-commutative polynomial rings. Ideals in these kind of rings do not necessarily have a finite Gröbner-basis. The aim of this project is to study Rai's system with respect to security and look at suggested attacks on these kind of cryptosystems.

Assignment given: 26. January 2009
Supervisor: Aslak Bakke Buan, MATH

Preface

This master thesis is sort of a continuation of a project which was about commutative Gröbner bases in Polly Cracker cryptosystems. In the project I only wrote from my own head and had more of a scholastic angle which lead to inaccuracy in the presentation of the theory. One of the subjects was to consider graph-3-coloring in the construction of a public key. Later, this gave me an idea of a way to solve and create sudoku puzzles using the algorithm of computing Gröbner bases. Due to some occupation with this thesis, I have not done any further research on the following system of equations:

- Regard every route as a variable, $\{x_i\}_{i=1}^{81}$.

- Regard the numbers 2-9 as prime numbers 2,3,5,7,11,13,17,19

- Set up 81 equations, one for each variable, x_i, on the form:
$(x_i-1)(x_i-2)(x_i-3)(x_i-5)(x_i-7)(x_i-11)(x_i-13)(x_i-17)(x_i-19)=0$

- Set up 27 equations, one for each row, column and square on the form:
$x_{j_1}\cdot x_{j_2}\cdot x_{j_3}\cdot x_{j_4}\cdot x_{j_5}\cdot x_{j_6}\cdot x_{j_7}\cdot x_{j_8}\cdot x_{j_9}-9699690=0$

The variables, $\{x_{j_s}\}_{s=1}^{9}$, are in the same row, column or square, and we have $1\leq j\leq 27$. Notice that $2\cdot 3\cdot 5\cdot 7\cdot 11\cdot 13\cdot 17\cdot 19=9699690$. Computing a reduced Gröbner basis for these 108 equations should make it easy to solve any sudoku (because they only have one solution), and by considering the reduced basis one could get ideas of how to construct your own sudokus of different levels.

The chapter 2 of this thesis is given without proof or examples, and I refer the reader to my project for more details. Of course, it may not be available for the general public, but there are other works to read about the subject. Later, the reader will observe that some of the results that are presented, comes from research using *Opal* [GHK]. This is a system for computing noncommutative Gröbner bases, but the details of how this work will not be presented. This is partly due to the lack of availability, but also because it is not a part of the objective in this thesis.

At last, I would like to thank myself for doing all the writing and thinking.

Abstract

We present the noncommutative version of the Polly Cracker cryptosystem, which is more promising than the commutative version. This is partly because many of the ideals in a free (noncommutative) algebra have an infinite Gröbner basis, which can be used as the public key in the cryptosystem. We start with a short brief of the commutative case which ends with the conclusion that the existence of "intelligent" linear algebra attacks ensures that such cryptosystems are left insecure.

Further, we see that it is hard to prove that noncommutative ideals have an infinite reduced Gröbner basis for all admissible orders. Nevertheless, in chapter 4 we consider some ideals for which it seems infeasible to realize a finite Gröbner basis. These are considered further in a cryptographic setting, and there will be shown that one class of ideals seems more promising than the others with respect to encountering attacks on the cryptosystem. In fact, at the end of this thesis we are proposing a way of constructing a cryptosystem based on this class of ideals, such that any linear algebra attack will not be successful.

However, many of the results are on experimental level, so there remains a serious amount of research in order to conclude that we have found a secure cryptosystem.

Contents

1 Introduction

Noncommutative Polly Cracker cryptosystems are a class of public key or asymmetric cryptosystems. The advantage of these systems compared to private key or symmetric cryptosystems, is that there is no need to secretly exchange a common decryption key. Any user of a public key cryptosystem can keep the decryption key, f^{-1}, to themselves and create an encryption key, f, which is public. The message space, M, and the ciphertext space, C, are also public, and we have

$$\begin{aligned} f &: M \longrightarrow C \\ f^{-1} &: C \longrightarrow M \end{aligned}$$

where f^{-1} should be infeasible to obtain for a cryptanalyst by looking at f (or any other way for that matter). The encryption key, f, is also called the public key, and the decryption key, f^{-1}, is also called the private key.

In later years, public key cryptosystems have grown in significance due to the increasing use of electronic communications. Here, it is hard to find a trusted channel to exchange decryption keys in, and therefore a private key cryptosystem is not considered secure. An example of a public key cryptosystem in use on the internet today is RSA, which is based on the infeasibility of factoring large integers. However, this system is *deterministic*, which means that any message, $m \in M$, will always be encrypted into the same ciphertext, $f : m \to c \in C$. If the message space is relatively small, a cryptanalyst can exploit this weakness by encrypting all possible messages and find the corresponding ciphertexts. Now he has knowledge of all the possible forms of c, and any ciphertext he intercepts is readable.

This disadvantage of deterministic encryption has lead to the introduction of *probabilistic* cryptosystems. The public key, f, in these systems can be viewed as a "one-to-many" function where a message, m, can be encrypted into infinitely many variations of ciphertexts. This means that the function, f, is not injective. Of course, decrypting a certain ciphertext with the private key, f^{-1}, will always give the same message.

The Polly Cracker cryptosystem was presented in 1993 by Fellows and Koblitz [FeKo], and is an asymmetric probabilistic cryptosystem. It is based on the theory of Gröbner bases, which has received much attention following the growth in computational power. Now, let p be an arbitrary element in a commutative algebra over a finite field, R, where we have the ideal, $I \subset R$. In general, the security of a Polly Cracker cryptosystem is based on the intractability of deciding if $p \in I$ or $p \notin I$, also called the ideal membership problem. This is later extended to also regard noncommutative algebras.

The main threat to a commutative Polly Cracker cryptosystem is the linear algebra attack, which is basically a way to correctly decrypt the ciphertext only by looking at the public key. In fact, there exists no known methods which provide that this system is resistant to such an attack. This is why people have tried to generalize this cryptosystem in order to make it secure, by introducing the noncommutative Polly Cracker cryptosys-

tem. This cryptosystem gives two promising advantages compared to the commutative case:

- Ideals in a free (noncommutative) algebra are not noehterian, and may therefore have infinite Gröbner bases.

- The ciphertext are constructed by $c = p + m$, where $p = \sum_{i=1}^{s} \sum_{j=1}^{k_i} F_{ij} q_i H_{ij}$. The coefficient of any monomial in c is given by two unknown coefficients of F_{ij} and H_{ij}, and the linear algebra attack will hopefully not work since the equations are quadratic.

These two points are mainly what caused the research on noncommutative Polly Cracker cryptosystems. The basic approach is to construct ideals, $J \subsetneq I$, where J has an infinite Gröbner basis, but I has not.

2 Commutative Gröbner bases and Polly Cracker

We start up with presenting commutative Gröbner bases used in the Polly Cracker cryptosystem. This is a cryptosystem which is considered to be insecure due to the existence of a linear algebra attack. There exists many good references for this material, such as [AdLo] and [Ko] which we refer the reader to for the proofs of the results we present here.

Commutative reduced Gröbner bases are always finite and easier to treat intuitively than the noncommutative. When you get an understanding of the basics, the step over to the noncommutative case will go relatively smooth, because you will see that it is quite similar to the commutative.

2.1 Monomial orders

As you will see later, the computation of Gröbner bases is relying on a well defined multivariate polynomial division, and for that, we need a well defined order of the monomials. We start with defining a monomial.

Definition 2.1. *A monomial is a product of powers of variables and is denoted* $\mathbf{x}^\alpha$*, where* $\boldsymbol{x} = (x_1, x_2, .., x_n)$ *and* $\alpha = (a_1, a_2, .., a_n) \in \mathbb{N}^n$*.* a_i *is the power of the variable* x_i*. The degree of a monomial is the sum of the powers* $|\alpha| = a_1 + a_2 + .. + a_n$*.*

Remark: It is the vector α which defines the monomial $\mathbf{x}^\alpha$. If the monomial is without a variable x_k, then a_k i simply equal zero.

For an order of such monomials to be well defined, it has to satisfy the following conditions:
(i) $\mathbf{x}^\alpha > 1$ for all $|\alpha| \neq 0$.
(ii) If $\mathbf{x}^\alpha > \mathbf{x}^\beta$, then $\mathbf{x}^\alpha \mathbf{x}^\delta > \mathbf{x}^\beta \mathbf{x}^\delta$ for all $\delta \in \mathbb{N}^n$.
(iii) Given any $\alpha, \beta \in \mathbb{N}^n$, exactly one of the orders $\mathbf{x}^\alpha > \mathbf{x}^\beta$, $\mathbf{x}^\alpha = \mathbf{x}^\beta$ or $\mathbf{x}^\alpha < \mathbf{x}^\beta$ holds.

Now we are ready to give the two most important total monomial orders.

The lexicographic order: We have an arbitrary order $x_1 > x_2 > ... > x_n$ and two vectors $\alpha = (a_1, a_2, .., a_n)$ and $\beta = (b_1, b_2, .., b_n)$, both in $\mathbb{N}^n$. The order of the two monomials, $\mathbf{x}^\alpha$ and $\mathbf{x}^\beta$, is depending on the vector $\omega = \alpha - \beta$. If the first element from the left, unequal zero, in ω is positive, then we have the order $\mathbf{x}^\alpha > \mathbf{x}^\beta$. If it is negative, the order will be $\mathbf{x}^\alpha < \mathbf{x}^\beta$. (Simple summarize: Look first at the order, then the degree).

The degree-lexicographic order: We have an arbitrary order $x_1 > x_2 > ... > x_n$ and two vectors $\alpha = (a_1, a_2, .., a_n)$ and $\beta = (b_1, b_2, .., b_n)$, both in $\mathbb{N}^n$. The monomial order is $\mathbf{x}^\alpha > \mathbf{x}^\beta$ if $|\alpha| > |\beta|$. If $|\alpha| = |\beta|$, we use the lexicographic order. (Simple summarize: Look first at the degree, then the order.)

Definition 2.2. *We have a polynomial* $p(\boldsymbol{x}) = c_1\boldsymbol{x}^{\alpha_1} + c_2\boldsymbol{x}^{\alpha_2} + .. + c_r\boldsymbol{x}^{\alpha_r}$ *where* $\boldsymbol{x}^{\alpha_1} > \boldsymbol{x}^{\alpha_2} > .. > \boldsymbol{x}^{\alpha_r}$ *for a given order. We define:*
The leading term of $p(\boldsymbol{x})$*:* $\mathrm{LT}(p(\boldsymbol{x})) = c_1\boldsymbol{x}^{\alpha_1}$
The leading monomial of $p(\boldsymbol{x})$*:* $\mathrm{LM}(p(\boldsymbol{x})) = \boldsymbol{x}^{\alpha_1}$
The leading coefficient of $p(\boldsymbol{x})$*:* $\mathrm{LC}(p(\boldsymbol{x})) = c_1$

Remarks
(*i*) The coefficients of the monomials has no influence on the order.
(*ii*) The leading monomial in a polynomial can vary depending on the monomial order that are being used.

2.2 Multivariate polynomial division

The multivariate polynomial division is very similar to the one variable polynomial division, but there are three important differences. Nevertheless, this is fairly easy to get hold of, so i will not present a division algorithm here. The three differences are:

(*i*) The outcome of the polynomial division depends on which order we use on the divisors, and is in general not unique. (In the case with one variable, the outcome is always unique.)

(*ii*) You need to have a total order on the monomials so the polynomial division is consistent. (In the case with one variable, a total order is given automatically by the degree of the monomials.)

(*iii*) When dividing a polynomial, p, by a polynomial, f, consider if each term in p is divisible by $\mathrm{LT}(f)$. (In the case with one variable, the division stops if $\mathrm{LT}(f)$ does not divide $\mathrm{LT}(p)$, without considering the rest of the terms in p.)

Definition 2.3. *Let* p *be a polynomial and* $F = \{f_1, f_2, .., f_r\}$ *a set of* r *polynomials in a certain order. We denote the multivariate polynomial division of* p *by* F *as*

$$p \xrightarrow{F} p'$$

where none of the terms in p' *are divisible by any term in* $\mathrm{LT}(F) = \{\mathrm{LT}(f_1), \mathrm{LT}(f_2), .., \mathrm{LT}(f_r)\}$. *We say that* p *is reduced to* p' *modulo* F.

Remark: In general, p' depends on the order on the set, F, but the next section presents special sets of polynomials where a polynomial order has no impact on the outcome of polynomial division by the set.

2.3 Gröbner bases

Let $R = \mathbb{F}[\mathbf{x}]$ be a polynomial ring where $\mathbf{x} = (x_1, x_2, .., x_n)$ (n commuting variables) and $\mathbb{F}$ is any field. The elements of R are polynomials which consist of finite sums of terms on the form $c\mathbf{x}^{\alpha}$, where $c \in \mathbb{F}$ is a coefficient of a monomial $\mathbf{x}^{\alpha}$.

All ideals of R are finitely generated, which means that every ideal, I, can be generated by a finite set of polynomials. The proof of this is based on the Hilbert basis theorem [AdLo, page 5]. An arbitrary subset $S \subset R$, generates an ideal, I_s, which we denote $\langle S \rangle = I_s$. The ideal I_s is the smallest ideal of R that contains S, and S is called a *generating set* or a *basis* for I_s.

Definition 2.4. *[Ra] Let $G = \{g_1, g_2, .., g_t\}$ be a set of t polynomials which generates an ideal $\langle G \rangle = I \subset R = \mathbb{F}[\boldsymbol{x}]$. If for all $p \in I$ and any order on G we have $p \xrightarrow{G} 0$, then G is called a Gröbner basis for I. Equivalently, if G is a Gröbner basis for I, then:*
(i) $\langle \mathrm{LT}(G) \rangle = \langle \mathrm{LT}(I) \rangle$.
(ii) For $f \in R$ and any order on G, the remainder $f \xrightarrow{G} f'$ is unique and is called the normal form of f which we denote $N(f) = f'$.
(iii) $p \in I$ if and only if $p = \sum_{i=1}^{t} h_i g_i$ with $\mathrm{LM}(p) = \max_{1 \le i \le t}(\mathrm{LM}(h_i)\,\mathrm{LM}(g_i))$.

Remark: In part *(iii)* each h_i is given from the polynomial division of p by the corresponding g_i. If G was not a Gröbner basis for I, the polynomial division by G would in most cases not reveal a linear combination $p = \sum_{i=1}^{t} h_i g_i$ for some $p \in I$. This means that $p \xrightarrow{G} p' \neq 0 = N(p)$, which would make us conclude that that $p \notin I$.

It is not hard to see the disadvantage in not having a Gröbner basis for the ideal when we are operating in a quotient ring R/I. The polynomial division modulo I is clearly not consistent if the basis of I is not a Gröbner basis. This is why we need a way to obtain a Gröbner basis for an ideal, I, given $\langle f_1, .., f_r \rangle = I$ where $r \geq 2$.

Constructing a Gröbner basis

The essential step in finding a way to construct Gröbner bases for ideals, is Buchberger's Theorem where the S-polynomial is introduced.

Definition 2.5. *Let $f_1, f_2 \in R = \mathbb{F}[\boldsymbol{x}]$ and $L = lcm(\mathrm{LM}(f_1), \mathrm{LM}(f_2))$. The linear combination $S(f_1, f_2) = \frac{L}{\mathrm{LT}(f_1)} f_1 - \frac{L}{\mathrm{LT}(f_2)} f_2$ is called the S-polynomial of f_1 and f_2.*

Theorem 2.6. ***Buchberger's Theorem*** *[Bu]*
Let $G = \{g_1, g_2, .., g_t\}$ be a set of polynomials in $\mathbb{F}[\mathbf{x}]$. G is a Gröbner basis for the ideal $I = \langle g_1, g_2, .., g_t \rangle$ if and only if $S(g_i, g_j) \xrightarrow{G} 0$ for all $i \neq j$.

Buchberger's theorem leads to Buchberger's algorithm. The input of the algorithm is a basis for an ideal, $\langle F \rangle = \langle f_1, f_2, .., f_r \rangle = I$. It uses all the different combinations of the f_i's to create S-polynomials, which are reduced modulo F to a polynomial p. Every $p \neq 0$ is added to the extended basis F^*, and makes combinations with the other polynomials in the basis to create other S-polynomials. This continues until all S-polynomials can be reduced to zero modulo F^*. Then F^* is a Gröbner basis for I. The termination of this algorithm is given by using the Hilbert basis theorem.

Buchberger's Algorithm

INPUT: Ideal generating polynomials, $F = \{f_1, f_2, .., f_s\} \in I \subset \mathbb{F}[\mathbf{x}]$
OUTPUT: A Gröbner basis, $G = \{g_1, g_2, .., g_t\}$ for I.
INITIALIZE: $G := F$, $H := \{\{f_i, f_j\} | f_i \neq f_j \in F\}$.
(H consists of all possible pairs of the input polynomials)

WHILE $H \neq \emptyset$ DO

Choose arbitrarily $\{f, g\} \in H$
(Going through all polynomial pairs)

$H := H - \{f, g\}$
(Removing our chosen polynomial pair)

$S(f, g) \xrightarrow{G} h$
(Constructing the S-polynomial and reducing it to $h \mod G$)

IF $h \neq 0$ THEN
(Checking if the basis, G, needs to be extended)

$H := H \bigcup \{\{g_i, h\} |$ for all $g_i \in G\}$
(Finding all possible pair combinations with h)

$G := G \bigcup \{h\}$
(Extending the basis, G, by h)

Remark: The Gröbner basis produced by this algorithm is not unique, in fact, an ideal can have infinitely many Gröbner bases.

Reduced Gröbner basis

When we compute a Gröbner basis for an ideal from Buchberger's algorithm, it will most likely contain redundant polynomials. These polynomials can be left out from the basis, without loosing the property of a Gröbner basis.

Definition 2.7. *We have an ideal $I = \langle G \rangle$, where $G = \{g_1, g_2, .., g_t\}$ is a Gröbner basis. G is called a reduced Gröbner basis for I, if $g_i \xrightarrow{G^*} g_i$ where $G^* = G \setminus \{g_i\}$ and $\mathrm{LC}(g_i) = 1$ for all $g_i \in G$. Every ideal, $I \in R$, has a unique and reduced Gröbner basis for a given order.*

Remark: To compute a reduced Gröbner basis from a non-reduced Gröbner basis, we only need to consider the following:
(i) If $g_k \xrightarrow{G^*} 0$, then g_k has to be left out from the basis.
(ii) If $g_k \xrightarrow{G^*} g'_k$, we remove g_k from the basis and replace it with g'_k.

This can be done after using Buchberger's algorithm, but the fastest way to get the unique and reduced Gröbner basis for an ideal, is to modify the algorithm so that you remove redundant polynomials during the computation. It is obvious that polynomial division by a reduced Gröbner basis is in general more efficient than by a non-reduced Gröbner basis.

2.4 The commutative Polly Cracker cryptosystem

Polly Cracker is a cryptosystem introduced by [FeKo] where the theory of Gröbner bases is essential. The security is based on the intractability of finding a Gröbner basis for an ideal in polynomial time. A person without access to a Gröbner basis for an ideal, $I = \langle F \rangle$, can not be sure of what the normal form of $p \in R/I$ is. In other words, the polynomial division $p \xrightarrow{F} p'$ is not consistent. We give now a description of this cryptosystem.

We have a reduced Gröbner basis, $G = \{g_1, g_2, .., g_t\}$, for an ideal in a polynomial ring, $I \subset \mathbb{F}_q[\mathbf{x}]$, as the private key of the cryptosystem. $\mathbb{F}_q$ is a finite field.

The public key is a set of polynomials $B = \{p_j\}_{j=1}^s \in I$ where every p_j is chosen as a linear combination of the polynomials g_i in the private key. The goal is to construct these linear combinations in such a way that finding a Gröbner basis for $\langle B \rangle$ computationally infeasible.

The message space, M, consists of polynomials on the normal form of the residue classes in the quotient ring $\mathbb{F}_q[\mathbf{x}]/I$. We wish to send the message $m \in M$. The encryption is achieved by adding the message with a linear combination of the polynomials in the public key. We obtain a ciphertext, $c = r + m$, where $r = \sum_{j=1}^s u_j p_j$. The polynomials $\{u_j\}_{j=1}^s$ are chosen arbitrarily.

Decryption is achieved by polynomial division of c by a Gröbner basis G'. As we know, the outcome of this polynomial division is the unique normal form of c, namely m ($c \xrightarrow{G'} m$, because $r \in I[\mathbf{x}]$).

Remark: If you obtain a Gröbner basis, G', for an ideal, $\langle G' \rangle = I$, the computations to get the reduced Gröbner basis, G, can be done efficiently. This means that *any* Gröbner basis should be computational infeasible to achieve.

We summarize:
Private key: The reduced Gröbner basis $G = \{g_1, g_2, .., g_t\}$ for an ideal, $I[\mathbf{x}] \in \mathbb{F}_q[\mathbf{x}]$.
Public key: A set of polynomials, $B = \{p_j\}_{j=1}^s$, where $p_j = \sum_{i=1}^t h_i g_i$, is "cleverly" constructed.
Message space: All polynomials $q(\mathbf{x})$, where $q(\mathbf{x}) \xrightarrow{G} q(\mathbf{x})$.

Encryption: $c = r + m$, where $r = \sum_{j=1}^{s} u_j p_j$, message: $m \in \mathbb{F}_q[\mathbf{x}]/I[\mathbf{x}]$.
Decryption: Reduction of c modulo G: $c \xrightarrow{G} m$.

Attacks on the Polly Cracker cryptosystem

When attacking this kind of cryptosystem, there are two possible goals of achievement.

(i) You examine the polynomials in the public key to find weakness in their construction. This can help you reveal the private key. If you succeed, every future message can be read.

(ii) You obtain a encrypted message, c, and compare it to the polynomials in the public key. If c is poorly constructed, it can be sorted out which terms of c is the message, m. If you succeed here, you only get to read this one message.

There exists no results which states that the private key (a Gröbner basis) will stay unrevealed for an attack of the type described in part (i), but Koblitz [Ko] suggests ways of constructing the public key such that finding a Gröbner basis is a NP-complete problem. Nevertheless, it is the part (ii) which has caught the most attention from cryptanalysts. Koblitz describes an attack of the type in part (ii), called a "linear algebra attack". The original form of this attack has a weakness, but H. W. Lenstra, Jr. proposed an improved version of it, called the "intelligent linear algebra attack".

The intelligent linear algebra attack

We obtain an encrypted message $c = r + m$, where $r = \sum_{j=1}^{s} u_j p_j$ and $m \in M$. It is reasonable to believe that a monomial multiple of an arbitrary polynomial in the public key, p_k, is presented as terms in c.

When constructing the ciphertext, we see that an unknown monomial, d, in u_k is multiplied with p_k such that $d \cdot p_k$ appears in c, and then from p_k, it should be easy to recognise the polynomial $d \cdot p_k$. It follows that the coefficient of d will be revealed. This can be done several times for each polynomial, p_j, in the public key. The monomials in c that can not be written as any monomial multiple of some polynomial in $\{p_j\}_{j=1}^{s}$, are most likely the message, m.

Remark: This method can be extended to also consider the coefficients of the polynomials in the public key. A scalar multiple of the coefficients of the monomials in p_k may be presented as coefficients of some monomials, $d \cdot p_k$, in c.

Lenstra has come up with a suggestion of what to do to provide security from this attack. When constructing the ciphertext, the encryptor should build monomials, $d_{j,i}$, into the polynomials, u_j, such that $d_{k,i}$ (appearing in u_k) times some term in p_k is cancelled in the entire sum. The number of the monomials, $d_{j,i}$, should not be too small. If none of the monomials are cancelled, the attack will succeed without any problems at all.

Despite Lenstra's suggestion, the existence of this attack has made T.Mora and others [Mo] conjecture that a commutative Polly Cracker cryptosystem can not be constructed in a way that gives sufficient security.

3 Noncommutative Gröbner bases

This section will present the basics of how to compute a Gröbner basis for a (noncommutative) free algebra over a finite field. As mentioned before, many of these Gröbner bases will be infinite, and that is what prompted the studies of noncommutative Gröbner bases in a Polly Cracker cryptosystem.

Most of the theory we present here holds for more general classes of noncommutative algebras, but a proper understanding of the material should come fairly easy if you are well aquainted with commutative Gröbner bases over polynomial rings. The notation we will use for a free algebra is $R = \mathbb{F}_q\langle \mathbf{x} \rangle$, where $\mathbb{F}_q$ is a finite field and $\mathbf{x} = x_1, x_2, .., x_n$ represent n non-commuting variables.

3.1 Noncommutative monomial orders

Monomials in a free algebra have the same nature as words in the alphabet, and are therefore also called *strings*. The only difference to be aware of, is that in this setting, the dictionary order is not an acceptable order to use.

Definition 3.1. *A monomial or string in a free algebra, R, is a noncommutative product of variables and is denoted* $\vec{\mathbf{x}} = x_{\sigma_1} x_{\sigma_2} .. x_{\sigma_s}$ *where* $1 \leq \sigma_i \leq n$.

Remark: A monomial/string, $\vec{\mathbf{x}}$, is defined by the variable index values $\{\sigma_i\}_{i=1}^{s}$ and the size of $0 \leq s < \infty$ (the string length). We define a monomial where $s = 0$ as 1.

The monomial orders in the noncommutative case needs to satisfy some conditions to avoid infinite descending chains of monomials. These conditions are written differently from the commutative case, but the content is the same. A monomial order is called *admissible* if it for all monomials, p, q, r, s, satisfies:

(i) If $p > q$, then $s \cdot p \cdot r > s \cdot q \cdot r$.
(ii) If $p = s \cdot r$, then $p > s$ and $p > r$.

Now we are ready to give the two most important noncommutative admissible monomial orders.

The (left) length-lexicographic order: We have an arbitrary order $x_1 > x_2 > ... > x_n$ and two monomials $\vec{\mathbf{x}}_1 = x_{a_1} x_{a_2} .. x_{a_{s_1}}$ and $\vec{\mathbf{x}}_2 = x_{b_1} x_{b_2} .. x_{b_{s_2}}$ where we denote the monomial length as $l(\vec{\mathbf{x}}_i) = s_i$. The order is $\vec{\mathbf{x}}_1 > \vec{\mathbf{x}}_2$
(i) if $l(\vec{\mathbf{x}}_1) > l(\vec{\mathbf{x}}_2)$ $(s_1 > s_2)$.
(ii) if $l(\vec{\mathbf{x}}_1) = l(\vec{\mathbf{x}}_2)$ $(s_1 = s_2)$ and for some $1 \leq k \leq s_1$, we have $x_{a_k} > x_{b_k}$ where $x_{a_1} x_{a_2} .. x_{a_{k-1}} = x_{b_1} x_{b_2} .. x_{b_{k-1}}$.

The (left) weight-lexicographic order: We have an arbitrary order $x_1 > x_2 > ... > x_n$ where every variable, x_i, has a scalar weight $w(x_i) \in \mathbb{N}$. The weight of a monomial is denoted $W(\vec{\mathbf{x}}_i) = \sum_{j=1}^{s_i} w(x_{\sigma_j})$. We order two monomials $\vec{\mathbf{x}}_1 > \vec{\mathbf{x}}_2$
(i) if $W(\vec{\mathbf{x}}_1) > W(\vec{\mathbf{x}}_2)$.
(ii) if $W(\vec{\mathbf{x}}_1) = W(\vec{\mathbf{x}}_2)$, by $\vec{\mathbf{x}}_1 > \vec{\mathbf{x}}_2$ by the (left) length-lexicographic order.

Remark: If every variable in a weight-lexicographic order has weight $w(x_i) = 1$, it is the same as the length-lexicographic order. Both orders can be modified by considering the strings from the right.

Definition 3.2. *We have a polynomial $f = c_1\vec{\mathbf{x}}_1 + c_2\vec{\mathbf{x}}_2 + .. + c_r\vec{\mathbf{x}}_r \in R$ where $\vec{\mathbf{x}}_1 > \vec{\mathbf{x}}_2 > .. > \vec{\mathbf{x}}_r$ for an arbitrary admissible order, and an ideal $I \subset R$. We define*

the tip of f:	$\mathrm{tip}(f) = \vec{\mathbf{x}}_1$
the tip coefficient of f:	$\mathrm{Ctip}(f) = c_1$
the tail of f:	$\mathrm{tail}(f) = f - \mathrm{Ctip}(f)\,\mathrm{tip}(f)$
the tips of a set F:	$\mathrm{tip}(F) = \{\mathrm{tip}(f) \mid f \in F\}$
the set of residue classes in R/I:	$\mathrm{NonTip}(I) = \{N(f) \mid f \in R\}$

Remark: The tip of a polynomial can vary depending on the monomial order, and corresponds to the leading monomial in the commutative case.

3.2 Noncommutative polynomial division

Polynomial division in the noncommutative case is based on the same principals as in the commutative, but some of the aspects are more complicated. Say we want to divide (reduce) a polynomial, p, by a set of polynomials $F = \{f_1, f_2, .., f_s\}$. Then we need to find monomials, u_{ij} and v_{ij}, and integers $k_1, k_2, .., k_s \in \mathbb{N}$ such that:
(i) $\quad p \xrightarrow{F} r \quad \Rightarrow \quad p = \sum_{i=1}^{s}\sum_{j=1}^{k_i} u_{ij} \cdot f_i \cdot v_{ij} + r.$
(ii) The remainder is a polynomial, r, which is not divisible by any f_i.

Remark: The monomials u_{ij} and v_{ij} are elsewhere in literature considered as polynomials. That is in my opinion not necessary. If u_{ij} and v_{ij} are polynomials, $\sum_{j=1}^{k} u_j f v_j$ can always be rewritten as a sum over monomials $\sum_{j=1}^{d} u'_j f v'_j$ where $d \geq k$ for monomials u'_j and v'_j in the polynomials u_j and v_j respectively.

Example 1
In this example we only divide a polynomial, p, by one polynomial, f, which compared to the formula above, gives us $i = 1$ and leads to $p = \sum_{j=1}^{k} u_j \cdot f \cdot v_j + r$. The order in use is the length-lexicographic order with $x > y$.

a) We want to divide the monomial $p = x^2y^3$, by the polynomial $f = xy - x$. Observe that since $xy > x$, it follows that $\mathrm{tip}(f) = xy$ and $\mathrm{Ctip}(f) = 1$. By the first division we get $p = x \cdot f \cdot y^2 + x^2y^2$, and further we see that there are three steps of division:

$$p = x \cdot f \cdot y^2 + x \cdot f \cdot y + x \cdot f \cdot 1 + x^2 = x \cdot f \cdot (y^2 + y + 1) + x^2$$

which we also can write $p \xrightarrow{f} x^2$. In this case we see that the monomials from the three steps of division can be merged into one ($k = 1$) multiple of f with polynomials $u_1 = x$ and $v_1 = (y^2 + y + 1)$. As mentioned above, this can also be written as three ($k = 3$) multiples of f with the monomials $v_1 = y^2$, $v_2 = y$ and $v_3 = 1$ on the right side.

b) Now we divide the monomial $p = x^2zy^3$ by the polynomial $f = xzy - z$ and get two steps of division,

$$p = x \cdot f \cdot y^2 + 1 \cdot f \cdot y + zy$$

which can be written $p \xrightarrow{f} zy$. It is clear that this sum ($k = 2$) over monomials can not be rewritten as polynomials. If we try, we see that $p = (x+1)\cdot f\cdot(y^2+y)-x\cdot f\cdot y-1\cdot f\cdot y^2+zy$, which does not contribute to more simplicity. This observation supports my remark of using sums over monomials rather than polynomials. In comparison with the commutative case, p can be written as $(xy^2 + y) \cdot f + zy$.

We now present a noncommutative division algorithm in pseudocode[1]. As in the commutative case, the outcome of this polynomial division depends on the order of the divisors $\{f_i\}_{i=1}^s$, and there is no guarantee that $p \xrightarrow{F} r = N(p)$. As we saw in definition 2.4, $N(p)$ is the normal form of p.

[1]Some of the disagreements with the algorithm in [Ra] are that it is not restarting the division by the set, F, after reducing the tip of the dividend. If further the tip is not divisible by F, the algorithm stops because it does not change from "true" to "false" in the checking of division occurred.

The noncommutative division algorithm

INPUT: $f_1, f_2, .., f_s, p$
OUTPUT: t_i, u_{ij}, v_{ij}, r, for $1 \leq i \leq s$ and $1 \leq j(i) \leq t_i$
INITIALIZE: $k_1, k_2, .., k_s := 0, r := 0, h := p, i := 1$

WHILE $(h \neq 0)$ DO

WHILE $(i \leq s)$ DO
(Going through all the f_i's)

IF $\text{tip}(h) = u \cdot \text{tip}(f_i) \cdot v$ for monomials u, v, THEN
(Checking if $\text{tip}(h)$ can be divided by f_i)

$k_i := k_i + 1$
(Counting one more division by f_i)

$u_{ik_i} := (\text{Ctip}(h)/\,\text{Ctip}(f_i)) \cdot u$
(Adapting the coefficient so we can eliminate $\text{tip}(h)$)

$v_{ik_i} := v$
(The proper coefficient has been taken care of in u_{ik_i})

$h := h - u_{ik_i} \cdot f_i \cdot u_{ik_i}$
(Reducing h by f_i so that $\text{tip}(h)$ is eliminated)

IF $(h \neq 0)$ THEN $i := 1$
(When h is reduced to $h' \neq 0$, we need to start over with all the f_i's)

ELSE $i := s + 1$
(Terminating the while loop if h is reduced to 0)

ELSE $i := i + 1$
(Checking if the next f_i divides h)

IF $(h \neq 0)$ THEN

$r := r + \text{Ctip}(h) \cdot \text{tip}(h)$
(If none of the f_i's divide the tip of h, we put it in the remainder)

$h := \text{tail}(h)$
(Removing the non-dividable tip of h so the algorithm can continue)

3.3 Noncommutative Gröbner bases

Let $R = \mathbb{F}_q\langle \mathbf{x} \rangle$ be a free algebra in n non-commuting variables where $\mathbb{F}_q$ is a finite field. An ideal, $I \subset R$, is *two-sided* if $f \cdot g \cdot h \in I$ for all $g \in I$ and $f, h \in R$. Left or right ideals will not be considered here. If $G = \{g_1, g_2, g_3, ..\}$ is a generating set or basis for I, the elements of $I = \langle G \rangle$ can be written as finite sums on the form $\sum_{i=1}^{t} \sum_j f_{ij} g_i h_{ij}$ where $f_{ij}, h_{ij} \in R$ are monomials. As earlier mentioned, an ideal in a noncommutative algebra might be infinitely generated.

Definition 3.3. *[Ra] Let $G = \{g_1, g_2, g_3, ..\}$ be a set of polynomials which generates a two-sided ideal $\langle G \rangle = I \in R = \mathbb{F}_q\langle \boldsymbol{x} \rangle$. If for all $p \in I$ and any order on G we have $p \xrightarrow{G} 0$, then G is called a Gröbner basis for I. Equivalently, if G is a Gröbner basis for I, then:*
(i) $\langle \mathrm{tip}(G) \rangle = \langle \mathrm{tip}(I) \rangle$.
(ii) For $f \in R$ and any order on G, the remainder $f \xrightarrow{G} f'$ is unique, and is called the normal form of p which we denote $N(f) = f'$.
(iii) $p \in I$ if and only if $p = \sum_i \sum_{j=1}^{k_i} f_{ij} g_i h_{ij}$ with $\mathrm{tip}(p) = \max_{i,1 \le j \le k_i} (\mathrm{tip}(f_{ij})\,\mathrm{tip}(g_i)\,\mathrm{tip}(h_{ij}))$.

Remarks
(i) For any ideal, $I \subset R$, we have $R = I \oplus \mathrm{NonTip}(I)$, which means that every $f \in R$ can be written uniquely as $f = i_f + N(f)$ where $i_f \in I$ and $N(f) \in \mathrm{NonTip}(I)$. Only division by a Gröbner basis for I can guarantee the unique normal form as outcome.
(ii) We will of course only consider finitely generated ideals, but the Gröbner bases may be infinite.

Constructing a noncommutative Gröbner basis

The construction of a noncommutative Gröbner basis for an ideal follows the same procedure as in the commutative case. The essential difference is that this can go on forever due to the possibility that the ideal has an infinite Gröbner basis. We start with presenting the noncommutative analogue to the S-polynomial.

Definition 3.4. *Let $f_1, f_2 \in R = \mathbb{F}_q\langle \boldsymbol{x} \rangle$ and a, b be monomials such that*
(i) $\mathrm{tip}(f_1) \cdot a = b \cdot \mathrm{tip}(f_2)$.
(ii) a is not divisible by f_2 and b is not divisible by f_1.
Then f_1 and f_2 have a overlap relation (or overlap) which we write

$$O(f_1 \cdot a, b \cdot f_2) = \frac{f_1 \cdot a}{\mathrm{Ctip}(f_1)} - \frac{b \cdot f_2}{\mathrm{Ctip}(f_2)}$$

Remarks
(i) If $a = 1$, then f_2 divides f_1. If $b = 1$, then f_1 divides f_2.
(ii) $\mathrm{tip}(O(f_1 \cdot a, b \cdot f_2)) < \mathrm{tip}(f_1) \cdot a = b \cdot \mathrm{tip}(f_2)$
(iii) In contrary to the commutative S-polynomial, each polynomial pair can have several overlaps and there may also be self-overlaps. This you have to consider when constructing a Gröbner basis for an ideal.

Example 2
Let $R = \mathbb{Z}_7\langle x, y, z\rangle$ where we use the length-lexicographic order with $x > y > z$. We look at an ideal, $I = \langle G\rangle = \langle g_1, g_2\rangle$, where $g_1 = yxz - yz$ and $g_2 = zy - x$. In order to find a Gröbner basis for I, we compute the two overlap relations between g_1 and g_2 and reduce them modulo I:

$$O(g_1 \cdot y, yx \cdot g_2) = yxzy - yzy - (yxzy - yxx) = yxx - yzy \xrightarrow{g_2} yxx - yx = g_3$$
$$O(g_2 \cdot xz, z \cdot g_1) = zyxz - xxz - (zyxz - zyz) = xxz - zyz \xrightarrow{g_2} xxz - xz = g_4$$

We add g_3 and g_4 to the basis G and find four more overlaps in G:

$$O(g_2 \cdot xx, z \cdot g_3) = xxx - zyx \xrightarrow{g_2} xxx - xx = g_5$$
$$O(g_4 \cdot y, xx \cdot g_2) = xzy - xxx \xrightarrow{g_2, g_5} 0$$
$$O(g_3 \cdot z, y \cdot g_4) = yxz - yxz = 0$$
$$O(g_3 \cdot xz, yx \cdot g_4) = yxxz - yxxz = 0$$

Further we see that g_5 has two overlaps with each of g_3 and g_4, and two self-overlaps where all of them are reduced to zero modulo $G = \{g_1, g_2, g_3, g_4, g_5\}$. This means that G is a finite Gröbner basis for I.

As the example shows, the overlap relations are used to construct a Gröbner basis for an ideal in the same way as the S-polynomials do in the commutative case. But there is one additional detail to take in concern, and that is why we present the following definition.

Definition 3.5. *Let $F = \{f_1, f_2, f_3, ..\}$ be a generating set for an ideal, $I \subset R = \mathbb{F}_q\langle \mathbf{x}\rangle$. We say F is tip-reduced if no* $\mathrm{tip}(f_i)$ *divides* $\mathrm{tip}(f_j)$ *for all $f_i \neq f_j \in F$.*

If in addition, all monomials $h \in \mathrm{tip}(I)$ can be written as $h = r_1 \cdot t \cdot r_2$ for any $t \in \mathrm{tip}(F)$ and $r_1, r_2 \in R$, we say that F is completely tip-reduced.

Remark: Every completely tip-reduced set, F, has a unique set of tips, $\mathrm{tip}(F)$, which may be infinite, and it follows that if a monomial is divisible by any element in $\mathrm{tip}(I)$, it is also divisible by some element in $\mathrm{tip}(F)$.

Example 3
Let $F = \{f_1, f_2\}$ generate an ideal I, where $f_1 = xxz - z$ and $f_2 = xxxzy - y$, using the length-lexicographic order with $x > y > z$. We obtain a tip-reduced generating set for the ideal by dividing f_2 with f_1, such that $f_2 \xrightarrow{f_1} xzy - y := f_2$ (updated). We can find out if the set is completely tip-reduced by considering overlaps in $F = \{f_1 = xxz - z, f_2 = xzy - y\}$. One overlap relation is $O(f_1 \cdot y, x \cdot f_2) = xy - zy$, which gives us the tip $xy \in \mathrm{tip}(I)$. We see clearly that xy can not be written as a multiple of $\mathrm{tip}(f_1)$ or $\mathrm{tip}(f_2)$, thus F is not a completely tip-reduced set.

Now we present the termination theorem which is the noncommutative version of Buchberger's theorem and leads to a noncommutative version of Buchberger's algorithm. The proof is given using G. Bergman's Diamond Lemma [Be]. The algorithm may not terminate because the input ideal, $I = \langle f_1, f_2, .., f_s \rangle$, has possibly a infinite Gröbner basis.

Theorem 3.6. ***The termination theorem***
Let $G = \{g_1, g_2, g_3, ..\}$ be a possible infinite set of tip-reduced polynomials in $\mathbb{F}_q\langle \mathbf{x} \rangle$. G is a Gröbner basis for the ideal $I = \langle g_1, g_2, g_3, .. \rangle$ if $O(g_i \cdot a, b \cdot g_j) \xrightarrow{G} 0$ for all $g_i, g_j \in G$ where a and b are monomials.

Buchberger's algorithm modified for noncommutative polynomial rings [Bu2]

INPUT: A tip-reduced set $F = \{f_1, f_2, .., f_s\}$.
OUTPUT: A Gröbner basis, $G = \{g_1, g_2, g_3, ..\}$ for $I = \langle F \rangle$.
INITIALIZE: $G := F$, $H := \{\{f_i, f_j\} | f_i \neq f_j \in F\}$.
(H consists of all possible pairs of the input polynomials)

WHILE $H \neq$ DO

Choose arbitrary $\{u, v\} \in H$
(Going through all polynomial pairs)

$H := H - \{u, v\}$
(Removing our chosen polynomial pair)

FOR each overlap relation of u, v DO

$O(u \cdot a, b \cdot v) \xrightarrow{G} h$
(Constructing the overlap relations and reducing it to h *mod* G)

IF $h \neq 0$ THEN
(Checking if the basis, G, needs to be extended)

$H := H \bigcup \{\{g, h\} |$ for all $g \in G\}$
(Finding all possible pair combinations with h)

$G := G \bigcup \{h\}$
(Extending the basis, G, by h)

Remarks
(*i*) As in the commutative case, this algorithm does not create a unique Gröbner basis. The uniqueness is given for the reduced Gröbner basis.
(*ii*) If the input ideal does not have a finite Gröbner basis, we can still obtain a partial finite Gröbner basis by terminating the algorithm after a certain number steps.

The observant reader will notice that the demand of a tip-reduced polynomial set differs from the commutative case. This demand is to guarantee that the noncommutative version of Buchberger's algorithm gives us a Gröbner basis as output, and prevent a possible infinitely long running time of the algorithm, even if the input ideal has a finite Gröbner basis. If you do not use tip-reduced polynomials, you have to take some precautions, as the following example shows.

Example 4
a) Let $\mathbb{F}_q\langle x, y, x\rangle$ be a free algebra where we use the length-lexicographic order with $x > y > z$. We want to find a Gröbner basis for the ideal $I = \langle G\rangle = \langle g_1, g_2, g_3\rangle$ where $g_1 = xy - x$, $g_2 = yxyz - xz$ and $g_3 = zxyyxy - x$.

We easily see that $\text{tip}(g_1)$ divides $\text{tip}(g_2)$ and $\text{tip}(g_3)$, and we get
$g_2 \xrightarrow{g_1} yxz - xz := g_2$ (reduced) and
$g_3 \xrightarrow{g_1} zxx - x := g_3$ (reduced)

Then we find the overlap relations:
$O(g_1 \cdot xz, x \cdot g_2) = xxz - xxz = 0$
$O(g_2 \cdot xx, yx \cdot g_3) = xzxx - yxx \xrightarrow{g_3} yxx - xx = g_4$
$O(g_3 \cdot y, zx \cdot g_1) = xy - zxx \xrightarrow{g_3, g_1} 0$

We add the polynomial $g_4 = yxx - xx$ to the basis, G, and find the new overlap relations:
$O(g_1 \cdot xx, x \cdot g_4) = xxx - xxx = 0$
$O(g_4 \cdot y, yx \cdot g_1) = xxy - yxx \xrightarrow{g_4, g_1} 0$

This means that $\{g_1, g_2, g_3, g_4\} = G$ is indeed a finite Gröbner basis for I.

b) Now we try to find a Gröbner basis without tip-reducing G at the beginning. We find the overlap relations:
$O(g_1 \cdot xyz, x \cdot g_2) = xxyz - xxz \xrightarrow{g_1} xxz - xxz = 0$
$O(g_3 \cdot xyz, zxyyx \cdot g_2) = zxyyxxz - xxyz \xrightarrow{g_1} zxxxz - xxz = g_4$
$O(g_3 \cdot z, zxy \cdot g_2) = zxyxz - xz \xrightarrow{g_1} zxxz - xz = g_5$
$O(g_2 \cdot xyyxy, yxy \cdot g_3) = xzxyyxy - yxyx \xrightarrow{g_3, g_1} = yxx - xx = g_6$
$O(g_3 \cdot 1, zxyy \cdot g_1) = zxyyx - x \xrightarrow{g_3} zxx - x = g_7$

Here we see that overlap relations between g_4 and g_5 will give infinitely many polynomials on the form $zx^iz - x^{i-1}z$. To prevent this, it is crucial that we discover g_7, which can reduce these polynomials to zero.

If we compute every overlap relation by turn and checking if the remainder, $g_i \neq 0$,

reduces any other polynomial in the basis, this problem will most likely not occur. (If we did that to begin with, we would have started with a tip-reduced set of polynomials.) An arbitrary selection between all possible overlap relations could clearly cause some problems by not discovering $O(g_3 \cdot 1, zxyy \cdot g_2) \xrightarrow{g_3} zxx - x = g_7$.

Further. we use g_7 to reduce g_4 and g_5 to zero, and add the polynomials $g_6 = yxx - xx$ and $g_7 = zxx - x$ to the basis, G, and find the new overlap relations:
$O(g_7 \cdot y, zx \cdot g_1) = zxx - xy \xrightarrow{g_7,g_1} 0$
$O(g_6 \cdot y, yx \cdot g_1) = xxy - yxx \xrightarrow{g_1,g_6} 0$
$O(g_1 \cdot xx, x \cdot g_6) = xxx - xxx = 0$
$O(g_2 \cdot xx, yxy \cdot g_7) = xzxx - yxyx \xrightarrow{g_7,g_1,g_6} 0$
$O(g_3 \cdot xx, zxyyx \cdot g_6) = zxyyxxx - xxx \xrightarrow{g_1,g_7} xxx - xxx = 0$

It is clear that $\{g_1, g_2, g_3, g_6, g_7\} = G'$ is not a Gröbner basis for I, and a simple example shows that $yxz \xrightarrow{G'} yxz$ but $yxz \xrightarrow{G} xz$. If we now reduce the polynomials we had at the start, g_2 and g_3, we see that $g_3 \xrightarrow{g_1,g_7} = 0$ and $g_2 \xrightarrow{g_1} = yxz - xz$ and we obtain the same Gröbner basis we found in part *a*) of the example.

We see that the demand of tip-reduction is just to avoid issues with reduction in the algorithm. The same Gröbner basis can be obtained from a polynomial set which is not tip-reduced if you modify the algorithm with the necessary reductions, but it is obvious that it is far more easy to just tip-reduce the set before starting on the algorithm.

c) We now look at the commutative version of the ideal, $I = \langle g_1, g_2, g_3 \rangle$, where we rewrite the polynomials $g_1 = xy - x$, $g_2 = xy^2z - xz$ and $g_3 = x^2y^3z - x$. We use the degree-lexicographic order with $x > y > z$. Without tip-reducing, we find the S-polynomials:
$S(g_1, g_2) = xyz - xz = g_4$
$S(g_1, g_3) = x^2y^2z - x \xrightarrow{g_1} x^2z - x = g_5$
$S(g_2, g_3) = x^2yz - x \xrightarrow{g_1} g_5$

We continue computing the S-polynomials
$S(g_1, g_4) = xz - xz = 0$
$S(g_2, g_4) = xyz - xz \xrightarrow{g_4} 0$
$S(g_3, g_4) = x^2y^2z - x \xrightarrow{g_2,g_5} 0$
$S(g_1, g_5) = x^2z - xy \xrightarrow{g_1,g_5} 0$
$S(g_2, g_5) = x^2z - xy^2 \xrightarrow{g_1,g_5} 0$
$S(g_3, g_5) = xy^3 - x \xrightarrow{g_1} x - x = 0$
$S(g_4, g_5) = x^2z - xy \xrightarrow{g_1,g_5} 0$

$G' = \{g_1, g_2, g_3, g_4, g_5\}$ is a Gröbner basis for I with several redundant polynomials. If we look at the noncommutative case in part *b*), the noncommutative polynomial $yxz - xz$ was not discovered and was the reason why we did not obtain a Gröbner basis. In both cases we can find this polynomial ($yxz - xz$ or $xyz - xz$) by dividing g_2 by g_1, but in the commutative case, it was also revealed by $S(g_1, g_2)$.

Nevertheless, we find the commutative reduced Gröbner basis $G = \{g_1, g_2\}$ where

$g_1 = xy - x$ and $g_2 = x^2z - x$, and observe that the commutative analogues to the noncommutative polynomials, $yxz - xz$ and $yxx - xx$, are reduced to zero by g_1.

Reduced noncommutative Gröbner basis

The definition of a noncommutative reduced Gröbner basis is formulated differently than in the commutative case, but both definitions provides uniqueness. The noncommutative definition has a lack of focus on the computational aspects and is not that intuitive.

Definition 3.7. *We have an ideal $I \subset \mathbb{F}_q\langle \mathbf{x} \rangle$ and F as a completely tip-reduced generating set of I with* $\mathrm{Ctip}(f) = 1$ *for all $f \in F$. Then the reduced Gröbner basis for I is $G = \{t_i - N(t_i)\}$ for all $t_i \in \mathrm{tip}(F)$.*

Remark: For all $g_i \in G$, $\mathrm{tip}(g_i) = t_i \in \mathrm{tip}(F)$ and $\mathrm{tail}(g_i) = N(t_i) \in \mathrm{NonTip}(I)$.

This gives the same result as the definition in the commutative case, except for the possibility of an infinite basis, $1 \leq i \leq \infty$. We see that the tips of the reduced Gröbner basis are equal the tips of the completely tip-reduced set, F, and the tail of a polynomial in the basis is not divisible by any element in $\mathrm{tip}(I)$, which is exactly the result of the definition in the commutative case.

Example 5
Look at $I = \langle G \rangle$ where $G = \{g_1 = xyz - xzy, g_2 = xz - yy\}$ using the length-lexicographic order with $x > y > z$. We observe that there are no overlaps between g_1 and g_2, which means that G is a finite Gröbner basis for I. In addition we see that $\mathrm{tip}(g_1)$ and $\mathrm{tip}(g_2)$ does not divide each other, and we can then conclude that G is a completely tip-reduced set.

Since there are only two monomials in the unique set of tips, $\mathrm{tip}(G) = \{t_1 = xyz, t_2 = xz\}$, the reduced Gröbner basis will also only consist of two polynomials, $g_1 = t_1 - N(t_1) = xzy - N(t_1)$ and $g_2 = t_2 - N(t_2) = xz - N(t_2)$. The normal forms of t_1 and t_2 can fortunately be computed because we have a Gröbner basis. We get

$N(t_1)$: $xyz \xrightarrow{g_1} xzy \xrightarrow{g_2} yyy$
$N(t_2)$: $xz \xrightarrow{g_2} yy$

which gives us the reduced Gröbner basis $G = \{g_1 = xyz - yyy, g_2 = xz - yy\}$

Remark: Using the commutative definition, we would have obtained the reduced Gröbner basis by $g_1 \xrightarrow{g_2} xyz - yyy$.

It is worth noticing that the noncommutative definition is taking starting point in the unique set of tips from a completely tip-reduced generating set for an ideal, in contrary to the commutative case where you start with a Gröbner basis. The two problems that occur is how to find a completely tip-reduced generating set and how to compute the normal form $N(t)$ in R/I of a monomial $t_i \in \mathrm{tip}(F)$.

It is obvious that both these problems can be solved with a Gröbner basis, G, for I, by reducing the set $\mathrm{tip}(G)$ to find $\mathrm{tip}(F)$, and reducing $t_i \xrightarrow{G} N(t_i)$. In practice, you are doing the same thing as in the commutative case, namely removing all redundant polynomials and monomial factors, and it seems clear that the definition from the commutative case can be used in the noncommutative case, and vice versa.

4 Infinite Gröbner bases

In a free algebra, $\mathbb{F}_q\langle \mathbf{x} \rangle$, there exists ideals with infinite Gröbner bases which is of cryptographic interest. In fact, most of the ideals are believed to have infinite Gröbner bases, but proving them to be infinite is hard due to a lack of research on this subject. However, there exists techniques to realize finite Gröbner bases, which has to be taken in consideration.

T. S. Rai [Ra] has made use of a system called *Opal* [GHK] to compute Gröbner bases and partial Gröbner bases in search for a proper class of ideals to use in the Polly Cracker cryptosystem. Due to the existence of techniques to realize finite Gröbner bases, we need ideals which have infinite Gröbner bases for all admissible orders.

In comparison to the commuative case, we see that a cryptanalyst in both cases wants to find a Gröbner basis, but the methods of finding them are different. In the commutative case the security is based on the computational time, which has to be so large that computing a Gröbner basis for a given order is infeasible. In the noncommutative case, the security is based on using an ideal with infinite Gröbner bases for all admissible orders. It is obvious that when you try to compute a Gröbner basis for such an ideal, the running time is infinite. The cryptanalyst tries out all possible admissible orders in search for a finite Gröbner basis.

4.1 Ideals with infinite Gröbner bases

It is pretty easy to find ideals which have an infinite reduced Gröbner basis. If you consider a set of polynomials with several overlaps, you will most likely succeed. In this section we look at some principal ideals and their infinite Gröbner bases in the free algebra $\mathbb{F}_q\langle x, y \rangle$ where we use a length-lexicographic order with $x > y$.

An ideal generated by xx -xy

The ideal $I = \langle xx - xy \rangle$ has an infinite reduced Gröbner basis on the form $G = \{g_i = xy^{i-1}x - xy^i | i \geq 1\}$. The polynomial $g_1 = xx - xy$ has a self-overlap, $O(g_1 \cdot x, x \cdot g_1) = xxy - xyx \xrightarrow{g_1} xyx - xy^2 = g_2$, and further we see that

$$O(g_i \cdot y^{j-1}x, xy^{i-1} \cdot g_j) = xy^i y^{j-1}x - xy^{i-1}xy^j \xrightarrow{g_i} xy^{j+i-1}x - xy^{i+j} = g_{i+j}$$

for all $i, j \geq 1$.

Remark: There are exactly two overlap relations for each polynomial pair g_i, g_j which are reduced to the same polynomial, g_{i+j}. You can say that the overlap function commutes for this ideal.

It is easy to see that if we change the order to $y > x$, we get the polynomial, $xy - xx$, which has no (self-)overlaps. We now have a finite Gröbner basis, $G' = \{xy - xx\}$, for the ideal, and a way to make use of it is presented later. Observe that $xx \xrightarrow{G} xy$ and $xx \xrightarrow{G'} xx$.

An ideal generated by xyx -xy

In the same way as above, it can be shown that $I = \langle xyx - xy \rangle$ has an infinite reduced Gröbner basis on the form $G = \{g_i = xy^i x - xy^i | i \geq 1\}$.

If we consider all possible forms of the admissible orders presented earlier, we see that no order can change the tip of the polynomial $xyx - xy$, and the infinite reduced Gröbner basis holds for any admissible order. Still there is a way of realizing a finite Gröbner basis for this ideal, which we are coming back to.

An ideal generated by xTx -a · xW

We now present ideals of a more general form in the free algebra $\mathbb{F}_q\langle \mathbf{x} \rangle$ with n variables, namely $I = \langle xTx - a \cdot xW \rangle$, where $a \in \mathbb{F}_q - \{0\}$ and T, W are monomials.

Remark: If $T = W = y$ we have the special case presented above.

The ideal has an infinite reduced Gröbner basis on the form
$G = \{g_i = xW^{i-1}Tx - a \cdot xW^i | i \geq 1\}$ if
(i) $T \geq W$
(ii) The set $\{T, W\}$ has no overlaps.
(iii) T and W do not begin or end with x.

This can be shown by proving the following statements:
a) Every two polynomials, g_i, g_j, has exactly two overlaps for all i, j which are contributing to the infinite reduced Gröbner basis.
b) No $\text{tip}(g_i)$ divides $\text{tip}(g_j)$ for all $i \neq j$

We see that part a) provides the construction of G with overlap relations. How these overlaps are computed and how many they are, is not important as long as they contribute to G. However, in this case, we eliminate the possibility that there are more overlaps than the two obvious ones. From the following proof we will see that every g_i is constructed by the overlap relation of some polynomials, g_r and g_s, where $r + s = i$ for $i \geq 2$. Part b) provides that elements in G do not reduce each other, and it follows that the proof of these points are sufficient in order to verify that $\langle G \rangle = I$ has an infinite Gröbner basis.

Proof.
a) We denote an overlap between two monomials, $m_1 = ABx$ and $m_2 = BxC$, in a new way as $O[AB(x), B(x)C] \backsim O(m_1 \cdot C, A \cdot m_2)$. Note that Bx is the actual overlap, but it is the last variable, x, in the first monomial which determines where the overlap ends, and is therefore put in parenthesis.

Every two polynomials, $g_i, g_j \in G$, has obviously two overlaps by the x on each side of the tips. If there are more than these overlaps, there must be at least one x in at least

one of W and T such that $T = AxB$ and/or $W = CxD$ for some monomials A, B, C, D. From condition (iii), we know that A and C do not start with x, and B and D do not end with x.

We see that such an overlap between xW^iTx and xW^jTx has to be on the form
(1.) $O[xW^iT(x), xW^j \cdot A(x)B \cdot x]$ or
(2.) $O[xW^iT(x), xW^k \cdot C(x)D \cdot W^lTx]$ where $k + l + 1 = j$.

- If $x \in T$, we see that $xW^iT(x) = xW^iAxB(x)$, which makes it clear that B has to end with A or C so that Ax or Cx is a part of the overlap. Notice that B has an x on each side, and it is easy to see that because of condition (i) $(C, D < T)$, there are only four alternative forms of B:
$B = W^j \cdot A$
$B = D \cdot W^e \cdot A$, where $0 \leq e \leq j - 1$
$B = W^k \cdot C$
$B = D \cdot W^f \cdot C$, where $0 \leq f \leq k - 1$

- If $x \notin T$, we only have overlap relations on form (2.), and by condition (i) we see that the possible forms of T are:
$T = ExW^kC$, where E does not start with x by condition (iii).
$T = W^{k-i}C$, if $k \geq i$

Every one of these forms of B and T contradicts condition (ii), namely that the set $\{T = AxB, W = CxD\}$ has no overlaps, and thus we have shown that there are only two overlap relations between some $\text{tip}(g_{i+1}) = xW^iTx$ and $\text{tip}(g_{j+1}) = xW^jTx$, which are on the form $O[xW^iT(x), (x)W^jTx] \backsim O(g_{i+1} \cdot W^jTx, xW^iT \cdot g_{j+1})$ for any $i, j \in \mathbb{N}$. Finally, we show that the overlap relations between any two polynomials, $g_i, g_j \in G$, is reduced to the Gröbner basis polynomial g_{i+j}.

$O(g_i \cdot W^{j-1}Tx, xW^{i-1}T \cdot g_j) = a \cdot xW^iW^{j-1}Tx - a \cdot xW^{i-1}TxW^j =$
$a(xW^{i+j-1}Tx - xW^{i-1}TxW^j) \xrightarrow{g_i} a(xW^{i+j-1}Tx - a \cdot xW^iW^j) =$
$a^{-1}a(xW^{i+j-1}Tx - a \cdot xW^iW^j) = xW^{i+j-1}Tx - a \cdot xW^{i+j} = g_{i+j}$

Remark: Also for this ideal, the overlap relation can be considered as commutative $(g_{j+i} = g_{i+j})$.

b) If some monomial, xW^iTx, divides any polynomial xW^jTx for $i \neq j$, then $j > i$ and $A \cdot xW^iTx \cdot B = xW^jTx$ for some monomials A, B.

- If $B \neq 1$, we use the same arguments as in the proof of part a), that T has to be on such a form that there will be overlaps in the set $\{T, W\}$, contradicting condition (ii).

- If $B = 1$, we see that when $j > i$, we get $W = Cx$ for some monomial C and $A = xW^{j-i-1}C$ such that

$$A \cdot xW^iTx = xW^{j-i-1}CxW^iTx = xW^jTx$$

This is a contradiction to condition (iii) where W can not start or begin with x, and thus we have shown that G is an infinite reduced Gröbner basis for I for all admissible orders. □

In search for a finite Gröbner basis, the order of the monomials, $T \geq W$, can possibly be changed by choosing another admissible order. If $W > T$, it may lead to $a \cdot xW > xTx$, and the ideal, $I = \langle a \cdot xW + xTx \rangle$, has indeed a finite reduced Gröbner basis because W has no overlaps.

Lemma 4.1. *Consider two monomials in $\mathbb{F}_q\langle x_1, x_2, .., x_n\rangle$ with the ordering $\vec{\mathbf{x}}_1 > \vec{\mathbf{x}}_2$. We define $|x_i \in \vec{\mathbf{x}}|$ as the number of times x_i appears in $\vec{\mathbf{x}}$. Given any variation of the monomial orders presented in chapter 3, the ordering $\vec{\mathbf{x}}_1 > \vec{\mathbf{x}}_2$ will not change if*
(i) $l(\vec{\mathbf{x}}_1) > l(\vec{\mathbf{x}}_2)$
(ii) $|x_i \in \vec{\mathbf{x}}_1| \geq |x_i \in \vec{\mathbf{x}}_2|$ *for all* $1 \leq i \leq n$

Proof. We consider the weight-lexicographic order because it is a generalization of the length-lexicographic order. For any value of $w(x_i)$, of every variable x_i, part (ii) ensures that the total weight of $\vec{\mathbf{x}}_1$ will be at least as big as $\vec{\mathbf{x}}_2$. It follows from part (i) that $|x_k \in \vec{\mathbf{x}}_1| > |x_k \in \vec{\mathbf{x}}_2|$ for at least one x_k.

This means that $w(\vec{\mathbf{x}}_1) + w(x_k) \geq w(\vec{\mathbf{x}}_2)$, and thus, $\vec{\mathbf{x}}_1 > \vec{\mathbf{x}}_2$ for any variation for the monomial orders presented in chapter 3. □

An example of two such monomials is $xyx > xy$, which was presented above.

4.2 String rewrite system

This section describes a way to find a finite reduced Gröbner basis for some ideals in $\mathbb{F}_q\langle \mathbf{x} \rangle = R$ called string rewrite system. In this setting, we use the technique to rewrite strings into new variables and compute a Gröbner basis out of this. The same principal ideals as in previous section will be considered, and it will be shown that it is possible to realize a finite Gröbner basis for every one of them using this technique.

The ideal generated by xx -xy

The ideal $I = \langle xx - xy \rangle$ has an infinite reduced Gröbner basis on the form $G = \{g_i = xy^{i-1}x - xy^i | i \geq 1\}$ using the length-lexicographic order with $x > y$. We now rewrite $xx = z$, which is the same as reducing g_1 with respect to $p = xx - z$, and we get $g_1 \xrightarrow{p} xy - z$. Notice that the outcome of this rewrite is the change of monomial order to $xy > z = xx$ because of $l(xy) > l(z)$. We compute all overlaps from the set $\{xx - z, xy - z\}$ and find a finite reduced Gröbner basis $G' = \{g'_1 = xx - z, g'_2 = xy - z, g'_3 = zx - xz, g'_4 = zy - xz\}$ for the order $z > x > y$.

We now give an example where we present three methods for reducing a polynomial by the set G.

Example 6
We want to find the normal form of a monomial, $p = yxy^3xy$, in the quotient ring R/I by reducing p by a Gröbner basis of I. This can be done by an infinite Gröbner basis, the string rewrite system or in this particular case, changing the order on the variables.

a) If we have knowledge of the infinite Gröbner basis, we see that $g_4 = xy^3x - xy^4$ can be used, and we get $p \xrightarrow{g_4} yxy^5 = N(p)$.

b) An infinite Gröbner basis can be infeasible to obtain for more complex ideals. Instead, we can use the finite reduced Gröbner basis, G', we found by introducing the polynomial $xx - z$ ($xx = z$) for the order $z > x > y$. We see that

$$p = yxy^3xy \xrightarrow{g_2'} yzy^2z \xrightarrow{g_4'} yx^2z^2 \xrightarrow{g_1'} yz^3 = p'$$

which can be written $p \xrightarrow{G'} yz^3$. Now we switch the monomials in g_1' from $xx - z$ to $z - xx$, because we want to reverse the string rewrite $xx = z$. We now reduce p' by $z - xx$ and the generating polynomial of G, $g_1 = xx - xy$. Now there are two important steps:

(i) We first reduce p' by $z - xx$, to reverse the string rewrite, and get $p' \xrightarrow{z-xx} yx^6$
(ii) Then we divide by g_1 **from the right** and get $yx^6 \xrightarrow{g_1*} yxy^5 = N(p)$

It is crucial that the division by g_1 is done from the right to get $N(p)$. Another way of dividing may have sent us back to the start. When deciding how you shall do this division, we look at $g_1 = xx - xy = x(x - y)$ and observe that to reduce as many x's to y as possible (by $(x - y)$), we need an x on the left side of an reducible x. It is clear that if we then start reducing from the right, every x but one of yx^6 is reduced to y.

c) Now we only consider the finite Gröbner basis $\{g = xy - xx\}$ obtained by changing the order to $y > x$. Observe that $g_2 = xy - z$ from the Gröbner basis G', gives the order $xy > z = xx$. It seems like this is the same as changing the order to $y > x$, as we did in the previous section, but the finite Gröbner basis is of course different. Nevertheless, the method of reducing p is quite similar, and we see that $p = yxy^3xy \xrightarrow{g} yx^6$. If we now change the order back, we can obtain the normal form of p (for $x > y$) by following the same procedure as in part b). One of the reasons why this method works here, is that xy does not divide xx.

The way we reversed the string rewrite system to find $N(p)$ described in part b), can be much more complicated and maybe not efficient when having several large generating polynomials of the ideal.

The ideal generated by xyx -xy

The ideal $I = \langle xyx - xy \rangle$ has an infinite reduced Gröbner basis on the form $G = \{g_i = xy^ix - xy^i | i \geq 1\}$ using the length-lexicographic order with $x > y$. We notice that the

tips of these polynomials can not be changed be choosing another of the monomial orders from chapter 3, based on lemma 4.1.

Example 7
We want to find the normal form of the monomial $p = xy^4xy$ in the quotient ring R/I by using string rewrite system. We see that from the infinite Gröbner basis we get $p \xrightarrow{g_4} xy^5 = N(p)$.

We rewrite $xy = z$ and use the polynomial $g_1' = xy - z$ to reduce $xyx - xy \xrightarrow{g_1'} zx - z = g_2'$. From this we find a finite reduced Gröbner basis $G' = \{g_1' = xy - z, g_2' = zx - z, g_3' = zy - zz\}$ for a chosen order $y > z$, and we reduce p modulo G':

$p \xrightarrow{g_1'} zy^3z \xrightarrow{g_3'} z^5 = p'$

Now we reverse the string rewrite, using $z - xy$ and the generator of the Gröbner basis, $g_1 = xyx - xy$:

$p' \xrightarrow{z-xy} xyxyxyxyxy \xrightarrow{g_1*} xy^5 = N(p)$

The division by $g_1 = xyx - xy = xy(x - 1)$ is done from right for the same arguments as in example 6.

Observe that the string rewrite gives us the order $y > z = xy$, which can be seen upon as a way of omitting the rules of admissibility. We can rewrite xyx by for example $yx = v$ with $y > v$ such that we get a non-admissible ordering, $xy > xyx$, on an admissible form, $xy > xv$. It follows by the use of this string rewrite that any monomial ordering can be changed, and we have to consider that any monomial in a polynomial, can appear as the tip of the polynomial.

If there exists at least one monomial with no self-overlaps in a polynomial, it seems reasonable to think that it is possible that the principal ideal, generated by that polynomial, has a finite Gröbner basis on some admissible form by using string rewrite system. Notice that the generating set of a principal ideal can only have self-overlaps.

The ideal generated by xTx +a · xW

It can be proven that one can find a finite Gröbner basis for the ideal, $I = \langle xTx - a \cdot xW \rangle$ where $a \in \mathbb{F}_q - \{0\}$ and $T \geq W$ are monomials with no overlaps and which do not start or end with x. The proof will be presented later because we need some results from the next section.

4.3 The search for an ideal of cryptographic interest

It seems clear that if it shall be impossible to realize a finite reduced Gröbner basis for an ideal, a Gröbner basis must be proven infinite for all admissible orders, but as we have shown, this may still not be enough. Trial and failure with the string rewrite system can take a lot of time in the attempt of proving that an ideal can not have a finite Gröbner basis for any admissible order or any monomial ordering. Therefore it is not easy to be sure if a finite Gröbner basis is infeasible to realize for an ideal.

An ideal where it is infeasible to realize a finite Gröbner basis

We now present an ideal, $I = \langle G \rangle$, which is proven to have an infinite Gröbner basis for any admissible order. The proof is given in [Ra, page 29], but will not be duplicated here.

$$G = \{g_1 = xzy + yz, g_2 = yzx + zy\} \tag{1}$$

Here we must consider $yz > xzy$ and $zy > yzx$ as admissible monomial orderings. They do not follow lemma 4.1, but can still be shown to be admissible. If $yz > xzy$ shall be legitimate in a monomial order, we could by example modify the definition of the weight-lexicographic order such that $w(yz) \neq w(zy)$. Of course, in a free algebra we can not assume the variables to commute when ordering them[2]. We present an example to show how such monomial orderings can be admissible.

Example 8
We look at the monomial $\vec{\mathbf{x}} = xzxyyz$ and observe that we have five different substrings of length 2 in the set $S_2(\vec{\mathbf{x}}) = \{xz, zx, xy, yy, yz\}$. By the rules of admissibility, we know that $\vec{\mathbf{x}}$ is ordered before any of these substrings for any order on x, y, z. Now observe that $zy \notin S_2(\vec{\mathbf{x}})$ and we can write $zy > xzxyyz$ if $zy > s$ for all $s \in S_2(\vec{\mathbf{x}})$, which means that we need the order $z > y > x$.

Consider monomials with an ordering $\vec{\mathbf{x}}_1 > \vec{\mathbf{x}}_2$ by lemma 4.1, where $l(\vec{\mathbf{x}}_2) = k$. In general, the ordering may still be changed for some admissible order if $\vec{\mathbf{x}}_2 \notin S_k(\vec{\mathbf{x}}_1)$. We see that such a monomial order is not presented in chapter 3.

Lemma 4.2. *Consider two monomials in $\mathbb{F}_q\langle x_1, x_2, .., x_n\rangle$ with the ordering $\vec{\mathbf{x}}_1 > \vec{\mathbf{x}}_2$. The monomial ordering will not change for any admissible order if and only if we can write $\vec{\mathbf{x}}_1 = a \cdot \vec{\mathbf{x}}_2 \cdot b$ where at least one of the nonzero monomials, a and b, are different from 1. In other words: if and only if $\vec{\mathbf{x}}_2$ divides $\vec{\mathbf{x}}_1$.*

Proof. Say $\vec{\mathbf{x}}_2 > \vec{\mathbf{x}}_1$ and $\vec{\mathbf{x}}_2$ divides $\vec{\mathbf{x}}_1$. Then we can write $\vec{\mathbf{x}}_1 = u \cdot \vec{\mathbf{x}}_2 \cdot v$ for some monomials u, v. We use the rules of admissibility and get $\vec{\mathbf{x}}_2 > \vec{\mathbf{x}}_1 = u \cdot \vec{\mathbf{x}}_2 \cdot v > \vec{\mathbf{x}}_2$, which contradicts the monomial ordering $\vec{\mathbf{x}}_2 > \vec{\mathbf{x}}_1$. □

This lemma makes it easy to accept that $yz > xzy$ is an admissible monomial ordering (for $yz > xz$ and $yz > zy$), and if we try out all admissible orders, we see that there are three possible sets of $\mathrm{tip}(G)$, namely $\{xzy, yzx\}$, $\{xzy, zy\}$ and $\{yz, yzx\}$. For each of these three sets, there has been proved in [Ra] that the overlap relations in G generates infinite sequences of monomials which is not divisible by any other polynomial in the Gröbner basis, similar to what we had in the ideal $I' = \langle xyx - xy \rangle$ [3].

[2] In [Ra], this is not brought to attention, and the order $yz > xzy$ is stated as admissible without explanation.

[3] A weakness in [Ra] appears after he proves that the ideals, $I = \langle G \rangle$ and $I' = \langle xyx - xy \rangle$, have infinite reduced Gröbner bases for all admissible orders. He presents a way of realizing a finite Gröbner

Remark: If we use $yz > xzy$ and $zy > yzx$ with the rules of admissibility $xzy > zy$ and $yzx > yz$, we see that $yz > xzy > zy > yzx > yz$, which means that we do not have an admissible order.

Example 9
By curiosity, we try to string rewrite the generators of I, g_1 and g_2, such that $\mathrm{tip}(g_1') = yz$ and $\mathrm{tip}(g_2') = zy$. These tips are not considered in [Ra] because they are obtained by violating the rules of admissibility.

It seems like the simplest way to obtain the monomial orderings, $yz > xzy$ and $zy > yzx$, on an admissible form is to rewrite $xz = v$ where $z > y > x > v$ under the length-lexicographic order. This order is settled due to the demand of the monomial ordering $zy > yzx > yz > vy$ where we also know that $yzx > zx$ by the rules of admissibility.

Now we have the reduced basis $G' = \{g_1', g_2', g_3'\}$ where $g_1' = yz + vy$, $g_2' = zy + vyx$ and $g_3' = xz - v$. We find the overlap relations:

$$O(g_1' \cdot y, y \cdot g_2') = yvyx - vyy = g_4'$$
$$O(g_2' \cdot z, z \cdot g_1') = vyxz - zvy \xrightarrow{g_3'} zvy - vyv = g_5'$$
$$O(g_3' \cdot y, x \cdot g_2') = xvyx + vy = g_6'$$

$$O(g_2' \cdot vyx, z \cdot g_4') = vyxvyx + zvyy \xrightarrow{g_6', g_5'} vyvy - vyvy = 0$$
$$O(g_4' \cdot z, yvy \cdot g_3') = yvyv - vyyz \xrightarrow{g_1'} yvyv + vyvy = g_7'$$
$$O(g_3' \cdot vy, x \cdot g_5') = xvyv - vvy = g_8'$$
$$O(g_1' \cdot vy, y \cdot g_5') = yvyv + vyvy \xrightarrow{g_7'} 0$$
$$O(g_5' \cdot z, zv \cdot g_1') = zvvy + vyvz = g_9'$$
$$O(g_6' \cdot z, xvy \cdot g_3') = xvyv + vyz \xrightarrow{g_8', g_1'} vvy - vvy = 0$$
$$O(g_4' \cdot vyx, yvy \cdot g_6') = vyyvyx + yvyvy \xrightarrow{g_4', g_7'} vyvyy - vyvyy = 0$$
$$O(g_6' \cdot vyx, xvy \cdot g_6') = xvyvy - vyvyx \xrightarrow{g_8', g_4'} vvyy - vvyy = 0$$

$$O(g_5' \cdot vyv, zv \cdot g_7') = zvvyvy + vyvvyv \xrightarrow{g_9', g_5'} vyvvyv - vyvvyv = 0$$
$$O(g_7' \cdot yx, yv \cdot g_4') = vyvyyx + yvvyy = g_{10}'$$
$$O(g_2' \cdot vyv, z \cdot g_7') = vyxvyv - zvyvy \xrightarrow{g_8', g_5'} vyvvy - vyvvy = 0$$
$$O(g_8' \cdot yv, xv \cdot g_7') = xvvyvy + vvyyv = g_{11}'$$
$$O(g_6' \cdot vyv, xvy \cdot g_8') = xvyvvy + vyvyv \xrightarrow{g_8', g_7'} vvyvy - vvyvy = 0$$
$$O(g_4' \cdot vyv, yvy \cdot g_8') = yvyvvy - vyyvyv \xrightarrow{g_7'} vyvyvy - vyvyvy = 0$$
$$O(g_8' \cdot yx, xv \cdot g_4') = xvvyy - vvyyx = g_{12}'$$

basis for the ideal I' using the string rewrite system, but when it comes to I, he just claims that it is not possible without explaining why. There are obvious differences between the ideals, but exactly what makes I so special he does not say.

$O(g'_9 \cdot z, zvv \cdot g'_1) = zvvvy - vyvzz = g'_{13}$
$O(g'_9 \cdot vyv, zvv \cdot g'_7) = zvvvyvy - vyvzvyv \xrightarrow{g_{13}, g'_5} vyvvyvv - vyvvyvv = 0$
$O(g'_9 \cdot vyx, zvv \cdot g'_4) = vyvzvyx + zvvvyy \xrightarrow{g_{13}, g'_5, g'_2} vyvvyvx - vyvvyvx = 0$
$O(g'_1 \cdot vvy, y \cdot g'_9) = yvyvz - vyvvy \xrightarrow{g'_7, g'_1} vyvvy - vyvvy = 0$
$O(g'_3 \cdot vvy, x \cdot g'_9) = xvyvz + vvvy \xrightarrow{g'_8, g'_1} vvvy - vvy = 0$

We stop computing overlap relations and observe that g'_5, g'_9 and g'_{13} are on the form $p_i = zv^iy + (-1)^i vyvz^{i-1}$, generated by g'_1 such that $O(p_i \cdot z, zv^i \cdot g'_1) = p_{i+1}$. At this point, other patterns in the infinite Gröbner basis are not easy to see, but the number and the complexity of the pattern seems to grow, even though there are many overlap relations which are reduced to zero.

Another way of rewriting strings can of course also be done, but it seems infeasible to do this in a way that would give a finite Gröbner basis. We see that the generating polynomials of I have several overlaps, but no self-overlaps, for any combination of tips. Thus, the conclusion in [Ra] that ideals on the form of I are infeasible to realize a finite Gröbner basis for, seems reasonable.

An ideal generated by the set in (1) can be generalized by exchanging the variables, x, y, z, with monomials, X, Y, Z. As with the variables, the set of monomials, $\{X, Y, Z\}$ can not have any overlaps. The monomials must consist of $n \geq 5$ variables, $x_1, x_2, .., x_n$, such that

$$X = x_1 \cdot \rho_1 \left(\prod_{i=2}^{n-1} x_i \right) \cdot x_n$$
$$Y = x_1 \cdot \rho_2 \left(\prod_{i=2}^{n-1} x_i \right) \cdot x_n$$
$$Z = x_1 \cdot \rho_3 \left(\prod_{i=2}^{n-1} x_i \right) \cdot x_n$$

where ρ_1, ρ_2, ρ_3 are distinct permutations of the variables $\{x_2, x_3, .., x_{n-1}\}$. Notice if $n < 5$, it is impossible to obtain three different permutations. We now have a generalized version of the generating set in (1):

$$G = \{g_1 = XZY + YZ, g_2 = YZX + ZY\} \tag{2}$$

Realizing finite Gröbner bases for principal ideals

Lemma 4.2 gives us an idea of how we easily can realize a finite Gröbner basis for some principal ideals. One thing to have in mind, is that a principal ideal can only have

a possible infinite reduced Gröbner basis if the tip of the generating polynomial has self-overlaps.

Notice that both example 8 and lemma 4.2 do not make explicitly use of the rule of admissibility (i). So the arguments of why the ordering $yz > xzy$ is admissible are incomplete. However, the following lemma gives the full proof of why $yz > xzy$ is an admissible monomial ordering.

Lemma 4.3. *Let $f = c_1\vec{\mathbf{x}}_1 + c_2\vec{\mathbf{x}}_2 + .. + c_r\vec{\mathbf{x}}_r$ be a polynomial which generates a principal ideal $I = \langle f \rangle$. We compare some monomial, $\vec{\mathbf{x}}_k$, with all the other monomials $\{\vec{\mathbf{x}}_i\}_{i=1,i\neq k}^{r}$ by setting $\vec{\mathbf{x}}_k = A_i \cdot X \cdot B_i$ and $\vec{\mathbf{x}}_i = A_i \cdot Y_i \cdot B_i$ where A_i is the largest common substring from the left, and B_i is the largest common substring form the right between $\vec{\mathbf{x}}_k$ and $\vec{\mathbf{x}}_i$.*

If there exists a monomial, $\vec{\mathbf{x}}_k$, in f with no self-overlaps and where $X \neq 1$ does not divide any Y_i in comparison to all $\{\vec{\mathbf{x}}_i\}_{i=1,i\neq k}^{r}$, then we can find a finite Gröbner basis for the ideal for some admissible order where $\mathrm{tip}(f) = \vec{\mathbf{x}}_k$.

Proof. Let $\vec{\mathbf{x}}_1 = A \cdot X \cdot B$ and $\vec{\mathbf{x}}_2 = A \cdot Y \cdot B$ be two monomials where X and Y do not divide each other, and where $\vec{\mathbf{x}}_2 > \vec{\mathbf{x}}_1$ by lemma 4.1 for some monomials A, B (possibly equal 1) and $X, Y \neq 1$. Still, we can show that the ordering $\vec{\mathbf{x}}_1 > \vec{\mathbf{x}}_2$ does not violate the rules of admissibility presented at the beginning of chapter 3. We denote $S_k(\vec{\mathbf{x}}_i)$ as the set of substrings of $\vec{\mathbf{x}}_i$ with length k. This means that every element in $S_k(\vec{\mathbf{x}}_i)$ divides $\vec{\mathbf{x}}_i$.

From lemma 4.1 we know that $l(\vec{\mathbf{x}}_2) > l(\vec{\mathbf{x}}_1) = k$, and by using the rule of admissibility (ii), we see that $\vec{\mathbf{x}}_2 > e$ for all $e \in S_k(\vec{\mathbf{x}}_2)$. Since $\vec{\mathbf{x}}_1$ does not divide $\vec{\mathbf{x}}_2$, we know that $\vec{\mathbf{x}}_1 \notin S_k(\vec{\mathbf{x}}_2)$. If follows that the ordering $\vec{\mathbf{x}}_1 > \vec{\mathbf{x}}_2$ does not violate condition (ii). Further, if $\vec{\mathbf{x}}_1 > \vec{\mathbf{x}}_2$, we see that also $X > Y$ by the rule of admissibility (i). Since X does not divide Y, we can use the same arguments as above and thus, the ordering $\vec{\mathbf{x}}_1 > \vec{\mathbf{x}}_2$ must be admissible.

If a polynomial, f, contains such a monomial described in lemma 4.3, the monomial can be set as the tip without violating the rules of admissibility. Since it got no self-overlaps, a principal ideal generated by this polynomial will of course have f as a finite Gröbner basis. □

Remark: Exactly how to define an admissible monomial order which do not follow lemma 4.1, seems to be hard. One may use one of the given monomial orders from chapter 3 and add exceptions, but this can be inconvenient to operate with.

Example 10
This example gives a generalization of the monomial ordering $yz > xzy$, which is admissible by lemma 4.3, but do not follow the given monomial orders in chapter 3. If we write $\vec{\mathbf{x}}_1 = A \cdot X \cdot B = yz$, and $\vec{\mathbf{x}}_2 = A \cdot Y \cdot B = xzy$, we see that $A = B = 1$ and $X = yz$ does not divide $Y = xzy$.

By the rule of admissibility (i), we multiply with z^{i-1} from the right on each side of $yz > xzy$ and get $yz^i > xzyz^{i-1}$. We now see that $xzyz^{i-1}$ can be reduced $i-1$ times by $yz - xzy$ such that $xzyz^{i-1} \xrightarrow{yz-xzy} = (xz)^i y$. It follows that $yz^i > (xz)^i y$ for $i \in \mathbb{N} - \{0\}$,

which has to be considered in the attempt of creating a suitable monomial order [4].

Lemma 4.3 is the first approach you should consider in the attempt to realize a finite Gröbner basis for some principal ideal. If all monomials with no self-overlaps do divide another monomial, the next step is to try the string rewrite system, as we did successfully with the principal ideal $\langle xyx - xy \rangle$. In general, we can realize a finite Gröbner basis for any principal ideal generated by a polynomial on the form

$$ABA - a \cdot AB$$

where the monomial AB has no self-overlaps for some monomials A and B. By symmetric arguments, we could do the same for the polynomial $ABA - BA$. We see that lemma 4.3 does not help us, so we rewrite the string $AB = z$ and get a finite reduced Gröbner basis on the form:

$$zA - a \cdot z$$
$$AB - z$$
$$zB - a^{-1} \cdot zz, \ (B > z)$$

Since z is not a part of A or B, there are no self-overlaps. We see that AB is represented as a tip, which by lemma 4.2, is impossible to obtain by only changing the (admissible) order.

Remark: There is one case where this set is not tip-reduced, and that is if zA divides zB, which means $B = A^k \cdot C$ for some monomial C. This division can only be done k times due to the lack of z's in B. We see that we get $zB \xrightarrow{zA-az} zC - a^{-k-1} \cdot zz$. $C \neq 1$ because else the monomial AB will have overlaps, so the finite reduced Gröbner basis holds for $C > z$.

At the end of section 4.2, we left the proof for realizing a finite Gröbner basis on hold. The principal ideal that was considered in that section is on the form $I = \langle xTx - a \cdot xW \rangle$ where $a \in \mathbb{F}_q - \{0\}$ and $T \geq W$ are monomials with no overlaps which does not start or end with x.

The similarity to the polynomial presented above ($ABA - AB$) is big, but we see that there are used stronger conditions with $A = x$, and that both T and W can not start or end with x or have overlaps. We see that T and W corresponds to the B's which may have self-overlaps.

The equal conditions is that neither xW or xT have self-overlaps. This is due to the fact that W and T can not have self-overlaps. As we know, the corresponding AB's has also no self-overlaps.

The property gained by increasing the strictness of the conditions, is the possibility that $T > W$, so which of these principal ideals is the most generalized, is hard to say.

[4] y may be viewed as an operator where z goes in on the right, and xz comes out on the left.

Proof.
If $W = T$, we see that the generating polynomial, $xTx - a \cdot xT$, is on the same form as above, which we found a finite Gröbner basis for.

One way of realizing a finite Gröbner basis is to get xW as the tip for some admissible order. The use of lemma 4.3 will help us find out if the monomial ordering $xW > xTx$ is admissible. If it is, we can get a finite Gröbner basis because xW does not have self-overlaps.

Further, if xW does not divide xTx, we can use the rewrite $xT = z$ and get a Gröbner basis for the ordering $xW > zx$:

$$\begin{aligned} xT - z \\ xW - a^{-1} \cdot zx \end{aligned}$$

Since this Gröbner basis is tip-reduced and the set, $\{T, W\}$, does not have any overlaps, it is indeed a finite reduced Gröbner basis.

It is clear that we now can continue the proof with considering $T > W$ and $xT = u \cdot xW \cdot v$ for some monomials u and v. Observe that if $u \neq 1$, then it starts with x. We use the rewrite $uxWv = z$ and get the polynomials

$$\begin{aligned} p = uxWv - z \xrightarrow{p_2} & \qquad uzxv - a \cdot z = p_1 \\ xTx - a \cdot xW \Rightarrow uxWvx - a \cdot xW \xrightarrow{p} & \qquad xW - a^{-1} \cdot zx = p_2 \end{aligned}$$

for $x > z$. Now there are four cases depending on the values of u and v.

(i) $u = v = 1$ gives $W = T$

(ii) $u = 1, v \neq 1$: $p_1 = zxv - a \cdot z$ and $p_2 = xW - a^{-1} \cdot zx$.
z is not represented in v or W, so the only possible overlap relation is that the x on the left in p_2 overlaps with some x in v or the x next to z. By $u = 1$, we see that $T = W \cdot v$ which means that v can not end with x This forces the left part of W to overlap with parts of (or the whole) v, which contradicts the fact that W and T has no overlaps.

If v is on the form $W^k \cdot v'$, then p_1 is reduced to $z^{k+1}xv' - a^{k+1} \cdot z = p_1'$. Because $T = Wv$ has no overlaps, v can not end with W. It follows that p_1 can not be reduced by p_2 such that $\text{tip}(p_1')$ ends with x. Thus, an overlap relation between p_1 and p_2 will never occur for any "legal" form of v.

(iii) $u, v \neq 1$: $p_1 = xu'zxv - a \cdot z$ and $p_2 = xW - a^{-1} \cdot zx$
As mentioned, if $u \neq 1$, we can rewrite $u = xu'$ where $T = u'xWv$. Since W and v do not end with x, the only possible overlaps is obtained by violating the rule that T and W can not have overlaps.

If xW divides xv or xu' from the left, we have the same case as in part (ii), so that is already considered. But now the v can end with xW, and then we get $p_1 = xu'zxv - a \cdot z = xu'zx \cdot v'xW - a \cdot z \xrightarrow{p_2} xu'zxv'zx - a^2 \cdot z := p_1$ (updated). This means

that we get two overlaps from the set $\{p_1, p_2\}$:

$$O(p_1 \cdot u'zxv'zx, xu'zxv'z \cdot p_1) = xu'zxv'zz - zu'zxv'zx = p_3$$
$$O(p_1 \cdot W, xu'zxv'z \cdot p_2) = zW - a^{-2} \cdot xu'zxv'zzx \xrightarrow{p_3} zW - a^{-2} \cdot zu'zxv'zxx = p_4$$

The monomial ordering in p_3 is set due to $x > z$, and by using lemma 4.3 we can get $zW > xu'zxv'zzx$ since W do not start or end with x. Now notice that $T = u'xWv'xW$, and because it can not have any self-overlaps, we can not have $v' = M \cdot u'$ for any monomial M. This eliminates the possibility of p_1 to have any self-overlaps, we will only have two more overlap relations:

$O(p_3 \cdot W, xu'zxv'z \cdot p_4)zu'zxv'zxW - a^{-2} \cdot xu'zxv'zzu'zxv'zxx \xrightarrow{p_3} zu'zxv'zxW - a^{-2} \cdot zu'zxv'zxu'zxv'zxx \xrightarrow{p_1} zu'zxv'zxW - a^{-1} \cdot zu'zxv'zzx \xrightarrow{p_2} a^{-1} \cdot zu'zxv'zzx - a^{-1} \cdot zu'zxv'zzx = 0$

$O(p_1 \cdot u'zxv'zz, xu'zxv'z \cdot p_3) = a \cdot zu'zxv'zz - xu'zxv'zzu'zxv'zx \xrightarrow{p_3} a \cdot zu'zxv'zz - zu'zxv'zxu'zxv'zx \xrightarrow{p_1} a \cdot zu'zxv'zz - a \cdot zu'zxv'zz = 0$

All overlap relations are reduced to zero, and this means that we have realized a finite Gröbner basis, $G = \{p_1, p_2, p_3, p_4\}$, where

$$p_1 = xu'zxv'zx - a \cdot z$$
$$p_2 = xW - a^{-1} \cdot zx$$
$$p_3 = xu'zxv'zz - zu'zxv'zx$$
$$p_4 = zW - a^{-2} \cdot zu'zxv'zxx$$

If $\text{tip}(p_2)$ divides $\text{tip}(p_1)$ or $\text{tip}(p_3)$ by u' or v' it will not affect the overlap relations since u', v' or W do not contain z.

(iv) $v = 1$[5], $u \neq 1$: $p_1 = xu'zx - a \cdot z$ and $p_2 = xW - a^{-1} \cdot zx$
By the same method as in part (iii), we get the finite Gröbner basis:

$$p_1 = xu'zx - a \cdot z$$
$$p_2 = xW - a^{-1} \cdot zx$$
$$p_3 = xu'zz - zu'zx$$
$$p_4 = zW - a^{-2} \cdot zu'zxx$$

□

We have now proven two principal ideals on a general form to have finite Gröbner bases, but not all principal ideals seem to have a finite Gröbner basis, even if we use string rewrite system. However, in the following we do not consider principal ideals,

[5]The case where $v = 1$ gives us obviously an overlap relation between p_1 and p_2, but in [Ra, page 29], it is stated that there are no overlaps, even though the monomials u and v are considered to possibly be 1.

because in general, principal ideals tend to have less complex (infinite) Gröbner bases than non-principal ideals.

Other ideals where it is infeasible to realize a finite Gröbner basis

In addition to the ideal given earlier, [Ra] presents a conjecture as a result of research with *Opal*. The conjecture is based on probability, and there exists exceptional cases for specific values of the coefficients.

Conjecture 4.4. *Let $G = \{g_1, g_2, .., g_s\}$ be a finite subset of $\mathbb{F}_q\langle x_1, x_2, .., x_n\rangle$ whose elements have the same tip, T, of length $l(T) = \alpha \geq 5$. Let N be the set of distinct strings of length $(\alpha - 1)$ that occur in all the g_i's combined. Then if*

(i) $|N| \approx 2 \cdot s$
(ii) $|N| \approx \frac{1}{3}n^{\alpha-1}$
(iii) $|N| < \frac{1}{2}n^{\alpha-1}$

there is a high probability that the reduced Gröbner basis of $\langle G\rangle$ is infinite.

Remarks
(i) $n^{\alpha-1}$ is the number of all possible strings of n variables of length $\alpha - 1$
(ii) After reduction of G, there will at most be just one polynomial with the tip T.
(iii) The coefficients of the monomials in the g_i's has to be chosen randomly.

Example 11
We now give an example of such an ideal in $\mathbb{F}_q\langle \mathbf{x}\rangle = \mathbb{Z}_{331}\langle x, y\rangle$. The generating polynomials, $G = \{g_1, g_2, g_3, g_4, g_5\}$, are all on the form

$g_i = xxxyyy + a_1 \cdot xyxyyy + a_2 \cdot yxxyyy + a_3 \cdot xxxyx + a_4 \cdot xxxyy + a_5 \cdot xxyyy + a_6 \cdot xyxyx + a_7 \cdot xyxyy + a_8 \cdot xyyyy + a_9 \cdot yxxyx + a_{10} \cdot yxxyy + a_{11} \cdot yxyyy +$ lower terms.

We see that $n = 2$, $s = 5$, $\alpha = 6$ and $|N| = 9$ (because all g_i's consist of the same monomials). This means that the conditions from conjecture 4.4 are fulfilled by

$$|N| \approx 10 = 2 \cdot s,\ |N| \approx 10,66 = \tfrac{1}{3}n^{\alpha-1},\ |N| < 16 = \tfrac{1}{2}n^{\alpha-1}$$

The set, G, can be reduced such that the five first monomials of any g_i are represented as the tips of the five generating polynomials. Then we get
$\text{tip}(G) = \{xxxyyy, xyxyyy, yxxyyy, xxxyx, xxxyy\}$. We see that $xxxyy$ divides $xxxyyy$ and therefore the tips of a tip-reduced G are $\text{tip}(G) = \{xyxyyy, yxxyyy, xxxyx, xxxyy, xxyyy\}$ where the monomials do not divide each other. [6] Notice that the original common tip, $xxxyyy$, is not presented in $\text{tip}(G)$.

In general, if the first k monomials do not divide each other for k generating polynomials with equal monomials, the tips of the reduced set equals those k monomials.

[6] The tip-reduced set presented in example 3.5.2 in [Ra], has the same tips.

Remark: If all coefficients but one in a polynomial, g_i $(2 \leq i \leq 5)$, equals the corresponding coefficients in g_1, the reduction by g_1 will only leave us a single monomial. This makes it easy to find examples of ideals based on conjecture 4.4 with finite Gröbner bases by "choosing" tip(G) to have no overlaps.

The lack of techniques of proving that an ideal do not have a finite Gröbner basis, leaves us no proof for conjecture 4.4. This could be something to do research on, with the goal of increasing the probability that no finite Gröbner basis can be realized. Nevertheless, we can do some observations which supports the conjecture. We know that one-third of all possible strings of length $\alpha - 1$ occurs in the g_i's, and since the number of polynomials is half of $|N|$, the tip-reduced set of G will most likely have tips of length approximately to $\alpha - 1$. This gives us a high probability that the elements of tip(G) have overlaps. There is also a high probability that some of these overlaps do not reduce to zero because the number of strings of length $\alpha - 1$ in G are not too many by condition (iii).

Using lemma 4.3 in order to get tips with no overlaps seems hopeless due to the magnitude of this system. One could think that increasing the number of variables could help minimizing the number of overlaps, but as we see of the conjecture, there are conditions which provides a certain relationship between the size of the system and the number of variables.

To summarize this chapter, we observe that principal ideals are left out from further use in a cryptographic setting. The reason for this is that principal ideals with infinite Gröbner bases tend to generate predictable sequences of polynomials. If g_1 generates polynomials $g_2, g_3, g_4, ..$, we have $\text{tip}(g_1) < \text{tip}(g_2) < \text{tip}(g_3) < \text{tip}(g_4) < ...$ In general, sequences generated by self-overlaps, seem to be predictable.

Ideals generated by polynomials on the form of (1) and (2), are together with conjecture 4.4 considered further in Polly Cracker cryptosystems, because it seems infeasible to obtain a finite Gröbner basis for these ideals. In addition, their infinite reduced Gröbner basis have complex patterns which are hard to predict.

5 Noncommutative Polly Cracker cryptosystems

The noncommutative Polly Cracker cryptosystem can be described in a similar way as in the commutative case. The main difference is that finding a Gröbner basis from the polynomials in the public key is infeasible because the basis is infinite. One should be aware that the ideals considered in the previous chapter are used as the public key. The private key is a finite Gröbner basis for a larger ideal. From the ideal generated by the polynomials in the public key, it may be possible to obtain partial Gröbner bases which can be used by a cryptanalyst. What [Ra] means with a "partial" Gröbner basis is hard to understand. It is stated that: "*Opal* was unable to compute a partial Gröbner basis.", but does that mean that no extra element in the infinite Gröbner basis was computed at all? My interpretation is that a partial Gröbner basis is a vague description, but which deals with a significant amount of computed elements in a infinite Gröbner basis.

Further, we present a summarize of the noncommutative Polly Cracker cryptosystem.

Private key: The reduced finite Gröbner basis, $G = \{g_1, g_2, .., g_t\}$, for a two-sided ideal, $I \in \mathbb{F}_q\langle \mathbf{x} \rangle$.
Public key: A set of polynomials, $B = \{q_r\}_{r=1}^{s}$, where $q_r = \sum_{i=1}^{t} \sum_{j=1}^{d_{ir}} f_{rij} g_i h_{rij}$, such that computing a Gröbner basis for $\langle B \rangle \subset I$ is infeasible.
Message space: Monomials in $M \subseteq \mathrm{NonTip}(I)$.

Encryption: $c = p + m$, where $m \in M$, and $p = \sum_{i=1}^{s} \sum_{j=1}^{k_i} F_{ij} q_i H_{ij}$[7] is a polynomial in $J = \langle B \rangle \subsetneq I$.
Decryption: Reduction of c modulo G, $c \xrightarrow{G} m$.

In general, we use a single polynomial, g, as the private key where its tip has no overlaps for some admissible order. This is because we want to minimize the computational time when decrypting, and also because it is easier to create public keys with no finite Gröbner bases. For completeness, we present again the two goals of achievement when attacking this kind of cryptosystem:

(i) You examine the polynomials in the public key to find weakness in their construction. This can help you reveal the private key. If you succeed, every future message can be read.

(ii) You obtain a encrypted message, c, and compare it to the polynomials in the public key. If c is poorly constructed, it can be sorted out which terms of c is the message, m. If you succeed here, you only get to read this one message.

Remark: Examples of attacks based on part (ii), are the linear algebra attack and reduction of c by the public key or a partial Gröbner basis.

[7] In [Ra], the k_i in the construction of p are denoted k_{ir}. In comparision to the construction of the q_r's, the function of r in d_{ir} is to sort out which q_r are being constructed. Since we only construct one p, the r in k_{ir} is useless and can be dropped.

One of the security aspects of this system, is the infeasibility of realizing a finite Gröbner basis for the ideal $J = \langle B \rangle$. Seemingly suitable ideals to use were presented in the previous chapter. There are also problems with constructing the ciphertext such that attacks of the type in part (ii) will not succeed. In general, the linear algebra attack is hoped to be useless because the ciphertext is generated by a multiple on each side of the q_r's (due to the noncommutativity), which gives the cryptanalyst a non-linear system of equations to solve.

Further we consider Polly Cracker cryptosystems where the public key is on the form presented in (1), (2) and conjecture 4.4.

5.1 Cryptosystems with the public key: $xzy + yz$ and $yzx + zy$

As mentioned at the end of the last chapter, the infinite Gröbner basis of the ideal $I = \langle q_1, q_2 \rangle$ where $q_1 = xzy + yz$ and $q_2 = yzx + zy$, has many sequences where the patterns are hard to predict compared to some principal ideal. In fact, [Ra] found the number of such sequences to grow as the magnitude of the computation of partial Gröbner bases were increased, using *Opal*. Another important discovery is that the tip of a computed element will often have shorter length than the element computed in the previous step. This helps the unpredictability of the infinite Gröbner basis, and we present some of the sequences:

$u_{1,n} = x(zx)^{n-1}z^{n+1}y - yz^n x(zx)^{n-1}$, $n \geq 1$
$u_{2,n} = zyzy^n x + y^{n-1}yzzy$, $n \geq 1$
$u_{3,n} = xz^{n+2}yzy + yzzyz^{n+1}$, $n \geq 1$
$u_{4,n} = zyzzyzy^n x + yzy^n x + yzy^{n+1}zzy$, $n \geq 1$
$u_{5,n} = zyzy^{n+1}zzx + y^n zzyzzy$, $n \geq 2$

Remark: For every $n \geq 1$, the tip of a polynomial in the second sequence divides a tip of a polynomial in the fourth sequence. This is not mentioned in [Ra], but it should not effect the promising properties of this ideal.

When constructing a public key on the form $B = \{q_1, q_2\}$, we can replace the variables x, z, y by polynomials f, g, h with the same properties. This means that the tips are on the form $\text{tip}(f) = x$, $\text{tip}(g) = z$ and $\text{tip}(h) = y$ for any admissible order. It follows from lemma 4.3 that the only possible additional terms in these polynomials are constants, a, b, c. Thus, the public key, B, is on the form $q_1 = fgh + hg$, $q_2 = hgf + gh$ where $f = x - a$, $h = y - b$ and the private key: $g = z - c$. This gives us:

$$q_1 = xzy - azy - cxy - bxz + yz + bcx + (ac - c)y + (ab - b)z + (bc - abc)$$
$$q_2 = yzx - bzx - cyx - ayz + zy + bcx + (ac - c)z + (ab - b)y + (bc - abc)$$

We know that the ideal, $J = \langle q_1, q_2 \rangle$, has an infinite reduced Gröbner basis for all admissible orders, as proved for ideals on the form of (1) from section 4.3. By comparison

to the general form of the polynomials in the public key, $q_r = \sum_{i=1}^{t} \sum_{j=1}^{d_{ir}} f_{rij} g_i h_{rij}$, we see that $t = 1$ and $d_{11} = d_{12} = 2$. The private key, g, seems infeasible to obtain due to the fact that $J \subsetneq I = \langle g \rangle$. Notice that $\text{tip}(g) = z$, so the message space, $M \subseteq \text{NonTip}(I)$, consists of all monomials without z. We can write $M = \mathbb{F}_q\langle x, y \rangle$.

The encryption of a message, m, is presented in [Ra, page 42] as $c = F_1 q_1 H_1 + F_2 q_2 H_2 + m$ for polynomials $F_1, F_2, H_1, H_2 \in R$. In comparison to $c = \sum_{i=1}^{s} \sum_{j=1}^{k_i} F_{ij} q_i H_{ij} + m$, we see that $s = 2$ and $k_1 = k_2 = 1$. This way of encrypting has obvious flaws because there are many instances where a sum over k_i monomials, F_{ij} and H_{ij}, multiplied with q_i, can not be rewritten as a multiple of q_i and two polynomials, F_1, H_1. That is why, in general, encryption should be done by using some chosen number of k_i monomials. This issue is brought to attention in example 1.

After studying attacks of the type in part (ii), [Ra] states that unless $\text{tip}(c)$ is large, you can correctly reduce the ciphertext, c, using the public key, $\{q_1, q_2\}$. Still, if $\text{tip}(c)$ is large, a partial Gröbner basis can be used to correctly reduce c, so we now present some techniques to construct the ciphertext such that reduction by the public key does not give the message, m, as outcome.

a) We try to construct c, such that $\text{tip}(c)$ is on the form $x(zyzx)^n(zy)^k$ or $(yz)^k(xzyz)^n x$ for some $n \in \mathbb{N}$ and $k \in \{1, 2\}$. This will ensure that the public key, $\{q_1, q_2\}$, does not correctly reduce c. However, if the cryptanalyst has computed a partial Gröbner basis, he will most likely manage to reduce c correctly, and thus, this approach do not provide sufficient security.

b) Another way of ensuring that the public key does not correctly reduce the ciphertext, is to construct c such that the $\text{tip}(q_i)$'s do not appear in c. For this method to work, we need the polynomials in the public key to have "overlapping overlaps". Notice that if they would only have overlaps, a partial Gröbner basis could most likely reduce c correctly, because it is constructed using such overlaps [8].

In order to have "overlapping overlaps", we need at least three polynomials, q_i, q_j and q_k, where $\text{tip}(q_i) = W_1 W_2 W_3$, $\text{tip}(q_j) = W_2 W_3 W_4$ and $\text{tip}(q_k) = W_3 W_4 W_5$ for some monomials $\{W_i\}_{i=1}^{5}$. It follows that the public key has to be extended to a finite set of polynomials, $Q = \{q_i\}_{i=1}^{s}$ for $s \geq 3$, where the q_i's are carefully selected from a partial Gröbner basis. Even though q_1 and q_2 do not contribute to the "overlapping overlaps", they must be a part of the public key to ensure that $\langle Q \rangle = \langle B \rangle = J$, which do not have a finite Gröbner basis[9].

Considering the sequences, $u_{1,n}$, $u_{3,n}$ and $u_{5,n}$, presented at the beginning of this section, we see that the tips have "overlapping overlaps" for some values of n. We find the the suitable tips of the polynomials in these sequences, where $k \geq 2$:

[8] This is my way of explaining an issue that is not brought to attention in [Ra].
[9] This is also not brought to attention in [Ra], and needs therefore an explanation.

$$
\begin{aligned}
u_{1,n} &= x(zx)^{k-1}z^{k+1}y = x(zx)^{k-2}z \cdot xz^k \cdot zy && = W_1 \cdot W_2 \cdot W_3 \\
u_{3,n} &= xz^{k+1}yzy = xz^k \cdot zy \cdot zy && = W_2 \cdot W_3 \cdot W_4 \\
u_{5,n} &= zyzy^{k+1}zzx = zy \cdot zy \cdot y^k zzx && = W_3 \cdot W_4 \cdot W_5
\end{aligned}
$$

In general, the encryption is done by choosing arbitrary constants, $\alpha, \beta \in \mathbb{F}_q$, such that $c = \alpha q_i W_4 W_5 + \beta W_1 q_j W_5 - (\alpha+\beta) W_1 W_2 q_k + m$, which ensures that multiples of $\mathrm{tip}(q_i)$, $\mathrm{tip}(g_j)$ and $\mathrm{tip}(q_k)$ are canceled in the entire sum.

It follows from this way of encrypting that the cryptanalyst will not find the message, m, by reduction of c using the public key, $\{q_i\}_{i=1}^s$. However, if he can compute a partial Gröbner basis, this system will not provide sufficient security even if the number of "overlapping overlaps" is large.

c) This next approach is also based on the idea that the tips of the polynomials in the public key should be subtracted off in c. The difference is that the extended public key, Q, consists of polynomials with the same tip, T. For s polynomials in $Q = \{F_i\}_{i=1}^s$, we need to find s arbitrary constants, α_i, such that $\sum_{i=1}^s \alpha_i = 0$. The encryption is done by setting $c = \sum_{i=1}^s \alpha_i F_i + m$, and it is easy to see that T does not appear as a monomial in c. Notice that the F_i's are generated by q_1 and q_2, and then of course, also generated by the private key, g.

As in approach b), the reduction of c by Q does not reveal the message, m, but if the cryptanalyst is able to compute a partial Gröbner basis or tip-reduce Q, this technique is vulnerable in the same way as in the previous settings. However, this approach led to the conjecture 4.4, which is discussed in a cryptographic setting in section 5.3.

Remark: This approach needs a method for how we can create the set $Q = \{F_1, F_2, .., F_s\}$ out of q_1, q_2 as the public key, and still maintain $\langle Q \rangle = J$.

Security evaluation
We evaluate the security based on the type of attacks presented at the beginning of this chapter.

(i) The private key can not be revealed by computation of a Gröbner basis for the ideal generated by the public key. Nevertheless, this system is so small and predictable (the tips of f, g, h are always x, z, y) that the cryptanalyst would know how the public key and its coefficients are created. Then he can equate the coefficients of the monomials in c with the constants given in the general form of q_1 and q_2, and thus, find the private key, $g = z - c$.

(ii) It is clear that if the cryptanalyst can compute a partial Gröbner basis, he can possibly decrypt any encrypted message, c, based on the techniques presented above. We assume that the small size of this system makes any computational obstacle feasible to overcome.

5.2 Cryptosystems with the public key: $XZY + YZ$ and $YZX + ZY$

We now consider a cryptosystem with the public key $B = \{q_1, q_2\}$ where $q_1 = XZY + YZ$ and $q_2 = YZX + ZY$. We choose the number of variables arbitrarily to be $n = 6$, and observe that the forms of X, Y and Z are

$$X = x_1 \cdot \rho_1 \left(\prod_{i=2}^{5} x_i \right) \cdot x_6, \; Y = x_1 \cdot \rho_2 \left(\prod_{i=2}^{5} x_i \right) \cdot x_6, \; Z = x_1 \cdot \rho_3 \left(\prod_{i=2}^{5} x_i \right) \cdot x_6$$

for some distinct permutations $\{\rho_i\}_{i=1}^{3}$ of $x_2x_3x_4x_5$. The ideal, $I = \langle B \rangle$, should have the same properties as the one presented in previous section, and in general, the same cryptographic evaluations can be done.

When constructing a public key on the form $B = \{q_1, q_2\}$, we can replace the monomials X, Z, Y by polynomials f, g, h with the same properties. This means that the tips are on the form $\text{tip}(f) = X$, $\text{tip}(g) = Z$ and $\text{tip}(h) = Y$ for any admissible order. It follows from lemma 4.3 that the only possible additional terms in these polynomials are $a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_6x_6$, since these are the only terms which are guaranteed to divide X, Y and Z [10]. Thus, the public key, B, is on the form $q_1 = fgh + hg$, $q_2 = hgf + gh$ where $f = X + \sum_{i=1}^{6} a_ix_i + a_0$, $h = Y + \sum_{i=1}^{6} b_ix_i + b_0$ and the private key: $g = Z + \sum_{i=1}^{6} c_ix_i + c_0$. Notice that since $\text{tip}(g) = Z$, the message space, $M \subseteq \text{NonTip}(I)$, can consist of all one variable polynomials. An explicit presentation of q_1 and q_2 is avoided due to their large size.

The encryption can be done in the same way as any of the three techniques from the previous section, but in this setting, *Opal* is not able to compute a finite Gröbner basis after running for at least 24 hours, according to [Ra].

Remarks
(i) A partial Gröbner basis of the ideal generated by the public key, will have tips on the same form as in the previous section, but the size of this system seems to increase the computational time of the partial Gröbner basis enormously.
(ii) The approach of encryption, b), from the previous section needs a partial Gröbner basis to be viable, so in this case it may not work.

Security evaluation
We evaluate the security based on the type of attacks presented at the beginning of this chapter.

(i) In this cryptosystem the private key, g, might seem to be better concealed, because its tip can consist of any permutation, ρ_i, with the following terms c_ix_i with arbitrary nonzero coefficients, c_i. In addition, this public key is much larger and complex than the one presented in the previous section. Still, the form of the public key is predictable,

[10] In [Ra] there are no arguments for why f, g and h have to be on the given form. Nevertheless, by seeing things in connection with lemma 4.3, I express my understanding here.

and a cryptanalyst with knowledge of the construction can easily reveal the private key, $g = Z + \sum_{i=1}^{6} c_i x_i + c_o$. For example, the cryptanalyst would know that $\text{tip}(q_1) = XZY$ and $\text{tip}(q_2) = YZX$, which makes it easy to spot out $\text{tip}(g) = Z$.

(*ii*) The public key of this system is so large that *Opal* is unable to compute a partial Gröbner basis for the ideal generated by the public key. This makes the techniques of encryption from the previous section, a) and c), promising if the ciphertext, c, is constructed such that the public key does not reduce c correctly. This cryptosystem are suggested to be worth more research, but is not evaluated further in [Ra]. However, the idea of similar tips of the polynomials in the public key from c), is investigated further in the next section.

5.3 Cryptosystems with the public key based on conjecture 4.4

This section describes a cryptosystem which can possibly be made secure. We present two techniques of encryptions, a) and b), where only technique b) seems to provide sufficient security against the considered attacks. It may be look like a generalization of the cryptosystem presented in the previous section, but the public key is not on the form, $q_1 = fgh - hg$ and $q_2 = hgf - gh$, and the inability to find finite Gröbner bases is based on probability as mentioned in the conjecture.

We start with considering the private key, $g = W + \sum_{i=1}^{n} a_i x_i + a_0$ [11], where $\{a_i\}_{i=0}^{n} \in \mathbb{F}_q - \{0\}$ in a free algebra, $\mathbb{F}_q\langle \mathbf{x} \rangle$, of n variables. The $\text{tip}(g) = W$ has to contain all the variables and have no self-overlaps. This makes the message space $M \subseteq \text{NonTip}(\langle g \rangle)$ very large and can consist of all monomials which do not contain all the variables.

The private key is of course used in the construction of the public key, $B = \{q_i\}_{i=1}^{s}$. We use the idea from the previous section that all the s polynomials should have equal tips, T. The polynomials q_i are constructed such that for all $1 \leq i \leq s$ and some monomials W_F and W_H, we have:

(i) $q_i = f_i \cdot g \cdot h_i$
(ii) $\text{tip}(f_i) = W_F$
(iii) $\text{tip}(h_i) = W_H$
(iv) f_i and h_i contain a proper number of words of length $l(W_F)$ and $l(W_F) - 1$, $l(W_H)$ and $l(W_H) - 1$ respectively, such that the conditions in conjecture 4.4 are fulfilled.

It follows that the common tip of all q_i's is $T = W_F \cdot W \cdot W_H$. Further we present two techniques of encryption.

a) Choose arbitrary constants, $\{\beta_i\}_{i=1}^{s} \in \mathbb{F}_q$, such that $\sum_{i=1}^{s} \beta_i = 0$. The encryption is done in the same way as technique c) in section 5.1, namely by setting $c = \sum_{i=1}^{s} \beta_i q_i + m$, and observe that $\text{tip}(c) < T$. As in section 5.1, the set of polynomials in the public key,

[11] This form makes $\text{tip}(g)$ have no overlaps for any admissible order. However, the only thing of importance is that g is a finite Gröbner basis for some chosen order. This means that g can be chosen arbitrarily as long as $\text{tip}(g)$ has no overlaps.

B, is easy to tip-reduce if the set is relatively small. It follows that the reduced set, B', is often able to reduce the ciphertext, c, correctly, and thus, this method of encryption is considered insecure.

Remark: I comparison to the technique c) of encryption described in section 5.1, we see a lot of resemblances, but one difference is that technique c) has to make use of polynomials on the form, $q_1 = fgh + hg$ and $q_2 = hgf + gh$, to generate the polynomial set with equal tips in the public key. In this case, the polynomials are created directly by the private key, g, without losing its secrecy. Still, the possibility of correctly reducing the polynomials in the public key, makes any of these techniques insecure.

b) This technique of encryption appears to be secure, and it is doing the opposite of a), namely extending the tip(c) compared to T. We choose arbitrary monomials, F_{ij} and H_{ij}, and construct the ciphertext, $c = \sum_{i=1}^{s} \sum_{j=1}^{k_i} F_{ij} q_i H_{ij} + m$, where for every q_k, there is at least one pair of F_{kj}, H_{kj} that satisfies $F_{kj} \cdot H_{kj} \geq T$.

The intention is not to avoid that c is divisible by any q_i, but make the outcome of any reduction of c, have more terms than before the reduction. Using the polynomials in the public key or partial Gröbner basis in reduction of c, will thus not reveal the message, m. According to [Ra], if the number of variables are $n = 3$, there are instances where *Opal* is not able to compute partial Gröbner bases after running for at least 24 hours. That is not crucial here, because the security is based on other aspects, but it can make it easier to do some precautions regarding the encryption.

If we compare these cryptosystems with the ones in section 5.1 and 5.2, we see that the main difference is that we do not use public keys on the form $q_1 = fgh + hg$ and $q_2 = hgf + gh$. Because of this, it is possible to find a finite Gröbner basis for J if we choose the right coefficients. However, ideals based on conjecture 4.4 are most likely not to have a finite Gröbner basis as long as we randomize the coefficients. The reason why we use conjecture 4.4 is because of the promising technique of encryption by using polynomials in the public key with the same tips. As a result, we can have many $q_i = f_i g h_i$'s in the public key with the same tip, not generated by $fgh + hg$ and $hgf + gh$, but only by g.

The benefits of this system compared to the earlier ones, is that it is far less predictable, due to all the possible forms of the f_i's, h_i's and g and the arbitrary number of elements in the public key, B.

Security evaluation
We evaluate the security based on the type of attacks presented at the beginning of this chapter.

(i) It is clear that $\langle B \rangle = J \subsetneq I = \langle g \rangle$, provided that J does have an infinite Gröbner basis for all admissible orders. It follows that $g \notin J$, and g is concealed from any attack based on computing Gröbner bases. Another aspect is that the size and the unpre-

dictability of this system makes any attempt of revealing the private key seem infeasible. The only knowledge a cryptanalyst has about this system, is that the tip of the private key is presented somewhere in the middle of the tip of every polynomial in the public key. Further concerns about this issue is presented in the next chapter.

(*ii*) We see that canceling the tips when encrypting, as proposed in technique a), will not ensure that a correct reduction of c is infeasible. This may have do to with that tip-reduction of $B = \{q_i\}_{i=1}^{s}$ and computation of a partial Gröbner basis are also achieved by canceling tips.

However, the technique b) of this section is promising. Encryption is done by adding big multiples of the q_i's to the message, m, such that reduction of c by these q_i's adds a big amount of terms to the reduced ciphertext, $c' \in \mathrm{NonTip}(\langle \mathrm{tip}(B) \rangle)$[12]. A cryptanalyst will after the reduction of c have terms spread in the sets $M \subseteq \mathrm{NonTip}(I) \subsetneq \mathrm{NonTip}(J) \subsetneq \mathrm{NonTip}(\langle \mathrm{tip}(B) \rangle)$, where the monomials in $\mathrm{NonTip}(\langle \mathrm{tip}(B) \rangle) \setminus \mathrm{NonTip}(J)$ can not be reduced without the infinite Gröbner basis of J. As long as g is safe, there is no way he can reduce a polynomial $p \in J$ to zero. It follows that a reduced ciphertext can be written $c' = p' + m$, and the cryptanalyst will not be able to sort out which terms in c' are components of the original message, m. Further concerns and the linear algebra attack is considered later.

Remark: If $c' \in \mathrm{NonTip}(J)$, no terms in c' can be divided by any polynomial in the infinite Gröbner basis of J. This means that c has been correctly reduced and $c' = m \in M \subseteq \mathrm{NonTip}(I)$.

[12] In [Ra, example 4.2], the reduction of some $p \in J$ by B or a partial Gröbner basis of $\langle B \rangle$, gives a remainder, p', containing much more terms than p.

6 Some security aspects in the Polly Cracker cryptosystem

This chapter consider the works of T. S. Rai and S. Bulygin [RaBu] in addition to [Ra], and there will be presented suggestions of how to resist certain attacks on a Polly Cracker cryptosystem based on conjecture 4.4. The security of such a cryptosystem has two issues. One is how the private key can be concealed safely, and the other is how to find a method of encryption which is resistant to linear algebra attacks.

The reader may have observed that the issue of concealing the private key now differs from its original objective (from the commutative case). As long as the polynomials in the public key generates an infinite Gröbner basis for an ideal which is a subset of the ideal generated by the private key, the cryptanalyst is unable to reveal the private key using the noncommutative version of Buchberger's algorithm. Instead, the problem is how to generate the public key such that not too much information about the private key is exposed.

The problem of encryption seems only to be about the linear algebra attack. Correctly reduction of c by the public key or a partial Gröbner basis seems infeasible if you use the technique of encryption b) described in section 5.3.

6.1 Chosen-ciphertext attacks

In this section we assume that the cryptanalyst has temporary access to the decryption key (also called the decryption black box) without knowing the details of it. This can be exploited with a chosen-ciphertext attack. The cryptanalyst carefully chooses certain ciphertexts to decrypt, in order to get as much information about the private key as possible.

The unpredictability of the private key, G, and the public key, B, in a cryptosystem based on conjecture 4.4, makes it hard for a cryptanalyst to explore the properties of G by looking at B. The only thing that is for certain, is that the largest tip, T, that occurs in $B = \{q_i\}_{i=1}^{s}$, is in $\langle \text{tip}(G) \rangle$. We know that every $q_i \xrightarrow{G} 0$, so of course, any $\text{tip}(q_i)$ has to be divisible by some monomial in $\text{tip}(G)$. Also, if it is publicly known that $M = \text{NonTip}(I)$, one could study the message space in order to get information about $\text{tip}(G)$.

In the two cryptosystems considered in section 5.1 and 5.2, it will be fairly easy to spot out the monomials of the private key because the form of the polynomials are publicly known. In order to prevent that further information about the private key is revealed to the cryptanalyst, [Ra, page 51] presents the use of non-monic polynomials as f, g and h such that the all the monomials in the public key have coefficients which are a multiple of three unknown constants. (When using monic polynomials, we know that three of the constants equals 1.)

However, despite the infeasibility of solving this system of cubic equations, one can find an own version of the private key which would work just as good as the original, by using a chosen-ciphertext attack. The reader should be aware that the exact values of the coefficients in a private key, g, is not necessary to obtain for a cryptanalyst. What is crucial to find, is the *relations* between the coefficients. For instance, if $g = \alpha \cdot xy + \beta \cdot x - \gamma$,

one could find another version $g' = xy + \alpha^{-1} \cdot \beta \cdot x - \alpha^{-1} \cdot \gamma$ where $\langle g' \rangle = \langle g \rangle$. This is similar to the commutative case where the private key is not a specific Gröbner basis, but just some Gröbner basis for the ideal $\langle G \rangle = I$.

A successful chosen-ciphertext attack

Consider a Polly Cracker cryptosystem where the following conditions are satisfied:

(*i*) The private key is a finite (not necessarily reduced) Gröbner basis, $G = \{g_1, g_2, .., g_t\}$.
(*ii*) $\text{tip}(G)$ can be easily determined from the public key.
(*iii*) The cryptanalyst has temporary access to the decryption black box.

Then the chosen-ciphertext attack can be used to find a Gröbner basis for $\langle G \rangle = I$.

The attack is using a fake ciphertext with some $\text{tip}(g_i) \notin M$ as the "message". The cryptanalyst uses now his temporary access to the decryption black box to decrypt $c = \sum_{i=1}^{s} \sum_{j=1}^{k_i} F_{ij} q_i H_{ij} + \text{tip}(g_i)$. It is clear that since $\sum_{i=1}^{s} \sum_{j=1}^{k_i} F_{ij} q_i H_{ij} \xrightarrow{G} 0$, the only thing that is returned from the black box is $N(\text{tip}(g_i))$, by $\text{tip}(g_i) \xrightarrow{G} N(\text{tip}(g_i))$. This is done for all elements $\text{tip}(g_i) \in \text{tip}(G)$, and we obtain a set of polynomials $G' = \{g_i'\}_{i=1}^{t}$ where all $g_i' = \text{tip}(g_i) - N(\text{tip}\, g_i)$.

As earlier mentioned, any polynomial $f \in \mathbb{F}_q\langle \mathbf{x} \rangle / I$ can by uniquely written as $f = i_f + N(f)$, where $i_f \in I$. It follows that $f - N(f) \in I$. If f is substituted with $\text{tip}(g_i)$, we see that every $g_i' = \text{tip}(g_i) - N(\text{tip}\, g_i) \in I$, so we have that $\langle G' \rangle \subset \langle G \rangle$. Since $\text{tip}(G') = \text{tip}(G)$, we know that G' is a Gröbner basis for $\langle G' \rangle = \langle G \rangle = I$.

Notice that G' is alternative version of the private key, G. When reducing $\text{tip}(g_i)$ by g_i, we get $\text{tip}(g_i) - \text{Ctip}(g_i)^{-1} \cdot g_i = \text{Ctip}(g_i)^{-1} \cdot \text{tail}(g_i)$. $\text{tail}(g_i)$ may be reduced futher if G is not a reduced Gröbner basis. It follows that $g_i' = \text{Ctip}(g_i)^{-1} \cdot g_i$, and it is easy to see that $\text{Ctip}(g_i) = 1$ and $G = G'$ if and only if G is a reduced Gröbner basis.

A chosen-ciphertext attack without any knowledge of $\text{tip}(G)$

Now we make an assumption that the tips can not be found from the public key, but the admissible order used in the cryptosystem are publicly known.

A cryptanalyst can now only determine the largest tip in the public key, $T = \max_{1 \le i \le s}(\text{tip}(q_i))$, which we know has to be divisible by some monomial in $\text{tip}(G)$. He then constructs a fake ciphertext with T as the "message" and gets $N(T)$ in return from the decryption black box. Out of this he creates a polynomial, $g_1' = T - N(T) \in G' \subset I$. Further he considers every polynomial $W < T$, and creates in a similar way a polynomial, $g_i' \in G'$ after using W as the "message" in a fake ciphertext, giving $N(W)$ as outcome. The cryptanalyst has now obtained $G' \subset I$ in the same way as above. Notice that if $W = N(W)$, then W is not divisible by any monomial in $\text{tip}(G)$ and the polynomial $g_k = W - N(W) = 0$ must be discarded.

We now know that $\langle G' \rangle \subset I$, but also that $\text{tip}(G) \subset \text{tip}(G')$ [13], so it follows that

[13] This is stated in [RaBu], but the following example seems to give a contradiction.

$\langle G' \rangle = \langle G \rangle = I$. Thus, G' is an alternative version of the private key.

Remark: By [RaBu], this attack does not operate in polynomial time, so one should construct the public key such that the number of elements $\{W | W < T\}$ is sufficiently large.

Example 12
Let $T = \max_{1 \leq i \leq s}(\text{tip}(q_i))$ be the largest tip in the public key, $B = \{q_i\}_{i=1}^{s}$ which is generated by a reduced Gröbner basis, G. We know that some monomial in $\text{tip}(G)$ divides T, but another question is if there exists a monomial $W \in \text{tip}(G)$ where $W > T$? If this is true, the chosen-ciphertext attack described above will not succeed in finding a Gröbner basis.

We recall the finite reduced Gröbner basis presented in example 2, $G = \{g_1, g_2, g_3, g_4, g_5\}$, where

$$
\begin{aligned}
g_1 &= yxz - yz \\
g_2 &= zy - x \\
g_3 &= yxx - yx \\
g_4 &= xxz - xz \\
g_5 &= xxx - xx
\end{aligned}
$$

for a length-lexicographic order with $x > y > z$. Because $\text{tip}(g_k)$ has greater length than any monomial in $\text{tail}(g_k)$, the order of x, y, z could be chosen arbitrarily without affecting the Gröbner basis. Notice by example that neither $\{g_5\}$, $\{g_3, g_4\}$ or $\{g_2, g_3\}$ generates the ideal, $\langle G \rangle = I$. The smallest set of the g_i's that generates I is $\{g_1, g_2\}$, and it can be shown that if g_1 or g_2 are left out from G, the set would generate another ideal $I' \subsetneq I$.

Further we present a way of constructing polynomials in the public key with the same tips. The point here is not to exactly follow the conjecture 4.4, but to compare the largest tip, T, of the q_i's with tips in $\text{tip}(G)$. We construct every q_i in a similar way, but with different coefficients:

$$
\begin{aligned}
q_i &= az \cdot g_1 + g_2 \cdot (-axz + d) + by \cdot g_5 + g_5 \cdot cz + g_3 \cdot (-bx + e) + (-cx + f) \cdot g_4 = \\
&azyxz - azyz - azyxz + axxz + byxxx - byxx + cxxxz - cxxz - byxxx + byxx - cxxxz + \\
&cxxz + dzy - dx + eyxx - eyx + fxxz + fxz \\
&= dzy - dx + eyxx - eyx + (a + f)xxz + fxz - azyz
\end{aligned}
$$

In the construction, we made use of the overlap relations between the g_i's. Notice that this would be the outcome for any order on x, y, z. Further we consider two cases.

a) The order is $x > y > z$
The polynomials in the public key are on the form $q_i = (a + f)xxz + eyxx - azyz +$

$fxz - eyx + dzy - dx$, and $\text{tip}(q_i) = xxz = T$.

If a cryptanalyst uses a chosen-ciphertext attack on this system, we see that the monomials $\{W|W < T\}$, are not including $\text{tip}(g_5) = xxx$. This means that the cryptanalyst does not have a Gröbner basis at the end of the attack, contradicting the proof in [RaBu, page 14] where they used that $\text{tip}(G) \subset \text{tip}(G') \subset \{W|W < T\}$.

However, we know that the cryptanalyst will obtain g_1 and g_2 ($T > \text{tip}(g_1), \text{tip}(g_2)$), which are the generators of the finite Gröbner basis, so he can then easily compute G.

b) The order is $y > z > x$
The polynomials in the public key are on the form $q_i = +eyxx - azyz(a+f)xxz - eyx + dzy + fxz - dx$, and we see that $\text{tip}(q_i) = yxx = T$.

In the same way as in a), we see that $\text{tip}(g_1) = yxz > yxx = T$, and thus, the cryptanalyst will not find a Gröbner basis for $\langle G \rangle = I$ using the chosen-ciphertext attack. Nevertheless, it can be shown that $G' = \{g_2, g_3, g_4, g_5\}$ is in itself a reduced Gröbner basis for another ideal, say $I' \subsetneq I$. It follows that the cryptanalyst is unable to compute g_1 and will believe that he has obtained the private key, G.

In fact, both the Gröbner bases, G and G', can be used to decrypt correctly. We see that $z \cdot g_1$ is used in the making of the polynomials in the public key, but it may be reduced to zero by g_2 and g_4, such that we have $z \cdot g_1 = g_2 \cdot (xz - z) + g_4$. This means that the Gröbner basis G' can be used as a version of the private key.

By this example we have shown that the chosen-ciphertext attack is not as straightforward as presented in [RaBu]. However, the threat seems to be unchanged.

Concealing the private key from chosen-ciphertext attacks

In the previous section, we tried to make a cryptosystem secure from a chosen-ciphertext attack by reducing the publicly known information and increasing the size of the largest tip, T, of the polynomials in the public key. Still the system is not sufficiently secured.

Now we try to consider the protection against this attack from a different angle, by setting some conditions for the message space and the decryption black box. This is proven to provide resistance from any chosen-ciphertext attack. In fact, the private key can be a single polynomial, g, with no overlaps and where $\text{tip}(g)$ is publicly known. A person who wants to communicate with r other persons without fearing the chosen-ciphertext attack, needs to do the following:

(*i*) Create r disjoint message spaces, M_i, such that $\bigcup_{i=1}^{r} M_i = M \subsetneq \text{NonTip}(I)$ and $\bigcap_{i=1}^{r} M_i = \emptyset$

(*ii*) Assign one of the message spaces, $\{M_i\}_{i=1}^{r}$, to each of the r persons.

(*iii*) Ensure that every $g_i \in G$ contains a monomial, $\vec{\mathbf{x}}_i \in \text{NonTip}(I) \setminus M$, such that $u \cdot \vec{\mathbf{x}}_i \cdot v \notin M$ for any monomials u and v.

(*iv*) If person k wants to have a ciphertext, c, decrypted, the decryption algorithm (or black box) must return c unreduced if some term in c is not in "his" message space, M_k.

If a cryptanalyst tries to decrypt a fake ciphertext, c, with the decryption black box using some $\text{tip}(g_i)$ as the "message", we see that $\text{tip}(g_i) \xrightarrow{g_i} \text{Ctip}(g_i)^{-1}\,\text{tail}(g_i)$ which we know contains some monomial $\vec{\mathbf{x}} \notin M$. This makes the black box return the unreduced c. Since any $M_i \subsetneq M$, every one of the r trusted persons will not be able to decrypt fake ciphertexts.

However, by introducing several message spaces we do not get a "proper" public key cryptosystem. That is why we have to modify this cryptosystem so that M is not divided into smaller message spaces. This provides that we can have an arbitrary number of users of the same message space. The modified cryptosystem is as follows:

(*i*) The message space is $M \subsetneq \text{NonTip}(I)$

(*ii*) Ensure that every $g_i \in G$ contains at least one monomial, $\vec{\mathbf{x}}_i \in \text{NonTip}(I) \setminus M$, such that $u \cdot \vec{\mathbf{x}}_i \cdot v \notin M$ for any monomials u and v.

(*iii*) If c' is the reduced ciphertext, $c \xrightarrow{G} c'$, we program the decryption algorithm (or black box) to return c unreduced if any term in c' is in $\text{NonTip}(I) \setminus M$.

In this cryptosystem the private key remains safe from the chosen-ciphertext attacks presented earlier, but now the system is vulnerable to another version of the attack:

Say person A and person B communicates with person C through this cryptosystem created by person C. If an encrypted message, m, is sent from person A to person C, one can see the ciphertext, c.

If person B wants to find out what person A sent, he can disguise the ciphertext, c, as his own by constructing $c' = p + c + m'$ where $p = \sum_{i=1}^{s} \sum_{j=1}^{k_i} F_{ij} q_i H_{ij}$. Further he sends c' to person A but gives a fake excuse for having person A return the message. Then person B will get back the reduced $c' \xrightarrow{G} m + m'$ where it is easy to find m by subtracting off the arbitrarily chosen $m' \in M$.

Remark: The decryption black box is just a picture of ways to obtain the reduction of a ciphertext you sent yourself.

One last thing to consider is if the same term $a \cdot \vec{\mathbf{x}} \in \text{NonTip}(I) \setminus M$ appears in every $g_i \in G$. If a cryptanalyst tries to decrypt $c = p - \text{tip}(g_{k_1}) + \text{tip}(g_{k_2})$, the term $\vec{\mathbf{x}}$ vanishes and he gets the reduction of c in return. He can continue systematically by choosing proper pairs of the g_i's in order to find out which terms belongs to which polynomial. However, he would not know the form of $\vec{\mathbf{x}}$, and of course, this attack will not work if one just randomizes and increases the terms $a_i \cdot \vec{\mathbf{x}}_i \in \text{NonTip}(I) \setminus M$ in each g_i.

6.2 A method of encryption which is resistant to linear algebra attacks

This section describes two different linear algebra attacks and ways of encountering them. The first attack can only be used if the cryptanalyst knows the form of F_{ij} and H_{ij} in the construction of the ciphertext. The other compares the ciphertext and the public key, and can always be used.

Linear algebra attack

Here we consider the standard linear algebra attack presented in [Ko] and a way to make the cryptosystem resistant to it. We recall the encryption of a message, m, using the public key $B = \{q_i\}_{i=1}^{s}$ which is $c = \sum_{i=1}^{s} \sum_{j=1}^{k_i} F_{ij} q_i H_{ij} + m$. By looking at the coefficients in c and the q_i's, the cryptanalyst tries to find the unknown coefficients of F_{ij}, H_{ij} and m. This gives a quadratic set of equations for which there exists no known method for solving.

However, a simple adjustment can give us a set of linear equations. The trick is to view the multiple of the coefficients of some monomials F_{ij} and H_{ij} as one unknown coefficient, δ. Of course, the cryptanalyst is only interested in the coefficients in the message. The coefficient, ω_l, of every monomial, $W_l \in M$ in c, gives a linear equation on the following form:

$$\omega_l = \sum_r \delta_r \cdot \lambda_r + m_{W_l} \tag{3}$$

If $W_l \notin M$, we get:

$$\omega_l = \sum_r \delta_r \cdot \lambda_r \tag{4}$$

λ_r is a known coefficient of some monomial in one of the q_i's, δ_r is the unknown multiple of the coefficients of some F_{ij} and H_{ij} and m_{W_l} is the coefficient of W_l in m. This attack will be successful it the number of such equations exceeds (or equals) the number of variables, δ_r and m_{W_l}.

Remark: The form (the monomials) of F_{ij} and H_{ij} has to be known in order to set up these equations[14]. This means that a cryptanalyst needs only to find the coefficients in order to obtain the message. In general, we see that the construction of c may bring several equal monomials. This is why the coefficient of some W_l may be written as a sum of several coefficients on the form $\delta_r \cdot \lambda_r$.

To ensure that the cryptosystem is resistant to this attack, one should increase the number, k_i, of the terms F_{ij} and H_{ij}. If one increases the number k_i with one for all $1 \leq i \leq s$, the linear system of coefficients will contain s more variables. The key concern when doing this is not to increase the number of equations. Of course, if the number of variables and equations increases with the same amount, this method will not help.

[14] This is not specified in [Ra].

We illustrate this by giving an example where the public key is constructed as in conjecture 4.4.

Example 13
Let $R = \mathbb{Z}_{331}\langle x, y\rangle$ where we use the length-lexicographic order with $x > y$. The private key is a single polynomial, g, where $\text{tip}(g) = xy$. We have a public key consisting of five ($s = 5$) polynomials on the same form:

$$q_i = \lambda_{1i} \cdot xxxyyy + \lambda_{2i} \cdot xyxyyy + \lambda_{3i} \cdot yxxyyy + \lambda_{4i} \cdot xxxyx + \lambda_{5i} \cdot xxxyy + \lambda_{6i} \cdot xxyyy + \lambda_{7i} \cdot xyxyx + \lambda_{8i} \cdot xyxyy + \lambda_{9i} \cdot xyyyy + \lambda_{10i} \cdot yxxyx + \lambda_{11i} \cdot yxxyy + \lambda_{12i} \cdot yxyyy + \lambda_{13i} \cdot xxxx + \lambda_{14i} \cdot xxxy + \lambda_{15i} \cdot xxyx + \lambda_{16i} \cdot xxyy + \lambda_{17i} \cdot xyxx + \lambda_{18i} \cdot xyxy + \lambda_{19i} \cdot xyyx + \lambda_{20i} \cdot xyyy + \lambda_{21i} \cdot yxxx + \lambda_{22i} \cdot yxxy + \lambda_{23i} \cdot yxyx + \lambda_{24i} \cdot yxyy + \lambda_{25i} \cdot yyyy + \lambda_{26i} \cdot xxx + \lambda_{27i} \cdot xxy + \lambda_{28i} \cdot xyx + \lambda_{29i} \cdot xyy + \lambda_{30i} \cdot yxx + \lambda_{31i} \cdot yxy + \lambda_{32i} \cdot yyx + \lambda_{33i} \cdot yyy + \lambda_{34i} \cdot xx + \lambda_{35i} \cdot xy + \lambda_{36i} \cdot yx + \lambda_{37i} \cdot yy + \lambda_{38i} \cdot x + \lambda_{39i} \cdot y + \lambda_{40i}$$

We construct the ciphertext by using the following forms of F_{ij} and H_{ij}

$$\begin{aligned} F_{i1} &= \alpha_{i1} \cdot xxx & \beta_{i1} \cdot yyy &= H_{i1} \\ F_{i2} &= \alpha_{i2} \cdot xxx & \beta_{i2} &= H_{i2} \\ F_{i3} &= \alpha_{i3} & \beta_{i3} \cdot yyy &= H_{i3} \\ F_{i4} &= \alpha_{i4} & \beta_{i4} &= H_{i4} \end{aligned} \tag{5}$$

where α_{ij} and β_{ij} are arbitrary constants and $k_i = 4$ for all $1 \leq i \leq 5$. Notice that $\max_{1\leq j\leq 4}(F_{ij} \cdot H_{ij}) \geq T = xxxyyy$, as in the condition of section 5.3. We encrypt a message, m, by setting $c = p + m$ where $p = \sum_{i=1}^{5}\sum_{j=1}^{4} F_{ij}q_iH_{ij}$. The number of terms in p are 130, but will not be explicitly given due to its large size.

If a cryptanalyst uses a linear algebra attack on this system, he can see that the number of variables on the form $\delta_r = \alpha_{ij} \cdot \beta_{ij}$ are $s \cdot k_i = 5 \cdot 4 = 20$. In addition he finds the monomials of c which are in the message space, M. There are 18 such monomials which are not divisible by xy:

$\{x^i\}_{i=1}^{7}$, $\{y^i\}_{i=1}^{7}$, yxx, yyx, yx and a constant term [15].

The cryptanalyst sets up 18 linear equations for the coefficients of these monomials on the form (3). The remaining monomials in c are not in the message space, M, and lead to 112 equations on the form (4). Now we have 38 variables but 130 equations. A way of increasing the number of variables while the number of equations remains the same, is to increase k_i but keep the additional F_{ij}'s and H_{ij}'s on the same form as in (5)[16]. This will keep the number of terms in p on 130, and one could assume that if $k_i \geq 23$, we would have $s \cdot k_i + 18 = 5 \cdot 23 + 18 = 133$ variables which would secure this system.

[15] Since the cryptanalyst knows the construction of p, these are the only monomials one should use in the message space.

[16] In [Ra, page 55] it is wrongfully stated that this system is secure from a linear algebra attack if $k_i \geq 7$ (in this setting: $k_i \geq 28$). Even the additional unknown constants m_{W_l}, has not been considered.

However, the new variables are not interesting for a cryptanalyst to find. We see that some polynomial, q_i, are then multiplied with several terms, F_{ij} and H_{ij}, which are on the same form.

Let λ be the coefficient of some monomial d in q_k and $F_{kj_1} \cdot H_{kj_1}$ be on the same form as $F_{kj_2} \cdot H_{kj_2}$ with the coefficients, $\alpha_{kj_1} \cdot \beta_{kj_1} = \delta_1$ and $\alpha_{kj_2} \cdot \beta_{kj_2} = \delta_2$, respectively. We consider W_l in c as the monomial we get from $F_{kj_1} \cdot \lambda \cdot d \cdot H_{kj_1}$ and $F_{kj_2} \cdot \lambda \cdot d \cdot H_{kj_2}$. The coefficient ω_l of W_l is now given by

$$\omega_l = \sum_{r=1}^{2} \delta_r \cdot \lambda_r = \lambda \cdot \sum_{r=1}^{2} \delta_r = \lambda(\delta_1 + \delta_2)$$

where $\lambda_1 = \lambda_2 = \lambda$ is the known coefficient of d in q_k. The sum over the unknown δ_r's, can then be viewed as one variable in the system of linear equations, so as long as you do not find new forms of the F_{ij}'s and H_{ij}'s, there is no point of increasing the k_i.

In general, we can see that in order to not get too many equations, one should keep the size of all the monomials, F_{ij} and H_{ij}, pretty close to each other. When increasing the number of F_{ij}'s and H_{ij}'s from this example, it seems best if they are constructed such that $1 < F_{ij} \cdot H_{ij} < xxxyyy$, and of course, not equal to each other. In addition, the number of equations will be minimized if all the q_i's are on the same form. In the contrary, if there are giant leaps in the size of these polynomials and monomials, the number of terms in the ciphertext will increase and lead to more equations.

Further, if the form of the monomials, F_{ij} and H_{ij}, are not known, the cryptanalyst is unable to set up equations. He could then try an attack which is described as the intelligent linear algebra attack from the commutative case[17].

The intelligent linear algebra attack

This attack is the same as presented in [Ko] for the commutative case and needs no knowledge of the encryption. The cryptanalyst compares the public key to the ciphertext in search for weakness in the construction.

This attack starts with the assumption that a multiple of some polynomial in the public key, $F_{ij} \cdot q_i \cdot H_{ij}$, will most likely be presented as terms in c. An easy way to go against this attack is to make sure that all the polynomials, $\{q_i\}_{i=1}^{s}$, are on the exact same form[18]. Then the cryptanalyst has no chance of knowing which terms in c are a multiple of which polynomial q_i.

However, if every F_{ij} and H_{ij} are chosen arbitrarily, the cryptanalyst can find a multiple, $F_{ij} \cdot q_k \cdot H_{ij}$, in c and go through the coefficients of all q_i's in order to find the proper q_k. This may work because the coefficients of some F_{ij} and H_{ij} can be seen upon as one unknown multiple of the known coefficients in q_k[19].

The way to provide security against this attack is to ensure that for every $1 \leq i \leq s$, $\{F_{ij} \cdot H_{ij}\}_{i=1}^{k}$ are on the same form. Now every q_i is multiplied with the same monomials

[17]This attack is not considered in [Ra].

[18]This is impossible for a public key on the form presented in section 5.2.

[19]This method is presented as a remark at the end of chapter 2.

but different unknown coefficients, so the coefficients of all monomials in c are a sum of at least s unknowns.

Example 14
We will give a small example of how arbitrarily chosen F_{ij}'s and H_{ij}'s can make a system vulnerable to an intelligent linear algebra attack. The order is length-lexicographic with $x > y$, and encryption is achieved by using the following elements:

$$\begin{aligned} q_1 &= a_1 \cdot xxy + a_2 \cdot yy + a_3 \\ q_2 &= a_4 \cdot xxy + a_5 \cdot yy + a_6 \\ F_{11} &= b_1 \cdot yy & b_2 &= H_{11} \\ F_{12} &= b_3 & b_4 &= H_{12} \\ F_{21} &= b_5 \cdot xx & b_6 \cdot y &= H_{21} \\ F_{22} &= b_7 \cdot xxx & b_8 \cdot xyy &= H_{22} \end{aligned}$$

Now we let $b_1 \cdot b_2 = \beta_1$, $b_3 \cdot b_4 = \beta_2$, $b_5 \cdot b_6 = \beta_3$ and $b_7 \cdot b_8 = \beta_4$ and compute $p = \sum_{i=1}^{2} \sum_{j=1}^{2} F_{ij} q_i H_{ij}$:

$$p = \beta_4 a_4 \cdot xxxxxyxyy + \beta_4 a_5 \cdot xxxyyxyy + (\beta_4 a_6 + \beta_3 a_4) \cdot xxxxyy + \beta_3 a_5 \cdot xxyyy + \beta_1 a_1 \cdot yyxxy + \beta_1 a_2 \cdot yyyy + (\beta_3 a_6 + \beta_2 a_1) \cdot xxy + (\beta_1 a_3 + \beta_2 a_2) \cdot yy + \beta_2 a_3$$

Say we send an empty message $c = p$, then the cryptanalyst can easily observe that the terms $\beta_4 a_4 \cdot xxxxxyxyy + \beta_4 a_5 \cdot xxxyyxyy + (\beta_4 a_6 + \beta_3 a_4) \cdot xxxxyy$ are a multiple of either q_1 and q_2. Further he checks if the coefficients in q_1 or q_2 can help him. If he tries q_2, he computes $\mathrm{Ctip}(q_2)^{-1} \cdot \mathrm{Ctip}(p) = a_4^{-1} \cdot \beta_4 a_4 = \beta_4$. Now let $q_2' = \mathrm{tail}(q_2)$ and $p' = \mathrm{tail}(p)$. The cryptanalyst continues by computing $\mathrm{Ctip}(q_2')^{-1} \cdot \mathrm{Ctip}(p') = a_5^{-1} \cdot \beta_4 a_5 = \beta_4$.

Notice that the cryptanalyst will not get β_4 as the multiple if he proceeds with this technique. However, since two of three coefficients in q_2 had the same multiple, he can with high probability conclude that $\beta_4 \cdot xxx \cdot q_2 \cdot xyy$ is used in the construction of c. Due to the poor construction of c, a cryptanalyst will find all the unknown β_i's by this method.

7 Security and weakness

In this chapter we evaluate some security aspects in a noncommutative Polly Cracker cryptosystem. We summarize the results of the previous sections and try make the necessary precautions such that any of the considered attacks will not be successful. The reader is referred to the beginning of chapter 5 for an overview of the cryptosystem.

7.1 Concealing the private key

By computation of Gröbner bases, it is infeasible for a cryptanalyst to obtain the private key as long as the polynomials in the public key, $B = \{q_i\}_{i=1}^{s}$, generates an ideal with no finite Gröbner basis. Chapter 4 gave examples of such ideals, but it is essential to not use monic polynomials. Even if the cryptanalyst knows the exact form of the private key, he can not find the coefficients because he will have to solve a system of cubic equations, which there are no known solutions for.

However, if a cryptanalyst succeeds with the chosen-ciphertext attack, he can find his own variant of the private key where the relations between the coefficients is the same as in the original private key. As we saw in chapter 6, this attack can be defeated, so we now give some necessary precautions in order to conceal the private key sufficiently:

- Realizing a finite Gröbner basis for $\langle B \rangle = I$ must be infeasible.
- Use non-monic polynomials with arbitrary coefficients when creating the public key.
- Design the message space and the decryption algorithm as proposed in section 6.1.

7.2 Encryption

The predictable form of the public keys given in section 5.1 and 5.2 (or (1) and (2)), makes any encrypted message by such keys vulnerable to some attack by a cryptanalyst. The general form of an ideal from conjecture 4.4 makes any attack more difficult, and that is why we will focus on such ideals in the following.

Decryption by reduction

The main concern regarding encryption of a message is to avoid that the cryptanalyst can correctly reduce the ciphertext using the public key and a partial Gröbner basis[20]. In section 5.3 we saw that an attempt of canceling the $\{\text{tip}(q_i) = T\}_{i=1}^{s}$ in the ciphertext by using coefficients with the property $\sum_{i=1}^{s} \beta_i = 0$, did not provide sufficient security from a "reduction-attack". Instead there was presented a promising method of encryption without elements of tip-reduction:

$c = \sum_{i=1}^{s} \sum_{j=1}^{k_i} F_{ij} q_i H_{ij} + m$
$\max_{1 \leq j \leq k_i}(F_{ij} \cdot H_{ij}) \geq T$ for $1 \leq i \leq s$

[20] This is not avoided by using a public key on the form of (1) from section 5.1.

The lack of a Gröbner basis for $\langle B \rangle = J$ makes it infeasible for a cryptanalyst to correctly reduce the ciphertext, because of the large terms in c, provided by $F_{ij} \cdot H_{ij} \geq T$. A reduced ciphertext, c', by a partial Gröbner basis for J, will have many more terms than the original c. Notice that if $c' \in \text{NonTip}(J)$, then $c' = m$, so we have terms in c' that are not divisible by any polynomial in the partial Gröbner basis, but are also not in $\text{NonTip}(J)$.

If a cryptanalyst is unable to compute a partial Gröbner basis (n must be large), we could reduce the set $\{q_i\}_{i=1}^s \rightarrow \{q_i'\}_{i=1}^s$ and then find $\{\text{tip}(q_i')\}_{i=1}^s$. Then we encrypt a message such that some $F_{ik} \cdot H_{ik} \notin \text{NonTip}(J)$ is not divisible by, or have no overlaps with any of the $\text{tip}(q_i')$'s. Since a cryptanalyst only can reduce c by $\{q_i'\}_{i=1}^s$, it follows that F_{ik} and H_{ik} will maintain, and we are guaranteed that $c' \neq m$. However, this technique seems hard to realize in practice.

Linear algebra attack

As stated in the last chapter, we can construct a ciphertext which is resistant to this attack if we ensure that the number of unknowns exceed the number of equations. Further we present some observations which can be used to regulate this:

- For $1 \leq i \leq s$, if $F_{il} \cdot H_{il} \neq F_{ij} \neq H_{ij}$ for all $1 \leq l \neq j \leq k_i$, then we increase the number of unknowns by k_i and get $s \cdot k_i$ unknowns.

- The number of equations will be at the lowest if all polynomials $\{q_i\}_{i=1}^s$ have the same form. Different q_i's will give different terms in c, and thus, more equations.

- Say $\min_{i,j}(F_{ij} \cdot H_{ij})$ is a constant, then the number of equations increases with $\max_{i,j}(F_{ij} \cdot H_{ij})$. To introduce more unknowns without increasing the number of equations too much, we chose monomials $F_{ij} \cdot H_{ij} < \max_{i,j}(F_{ij} \cdot H_{ij})$.

It is important to ensure that a sufficient amount of terms of $\sum_{i=1}^s \sum_{j=1}^{k_i} F_{ij} q_i H_{ij}$ are in the message space, such that m can not easily be determined from c.

This attack needs precise knowledge of the construction of the ciphertext to be used. If the F_{ij}'s and H_{ij}'s are chosen arbitrarily, the attack is useless.

The intelligent linear algebra attack

It seems clear from the previous chapter that protection against such an attack is provided if we do the following:

- Let all elements in the public key, $B = \{q_i\}^i$, be on the same form.
- For every $1 \leq i \leq s$, let $\{F_{ij} \cdot H_{ij}\}_{j=1}^k$ be on the same form.

7.3 A proposal of a Polly Cracker cryptosystem

Now we present a way to create a Polly Cracker cryptosystem which is hopefully resistant to all the considered attacks.

- Choose an arbitrary polynomial, g, as the private key where $\text{tip}(g)$ has no overlaps. Make also sure that $M \subsetneq \text{NonTip}(\langle g \rangle)$ such that at least one term in g are in $\text{NonTip}(\langle g \rangle) \setminus M$.

- Construct the public key, $B = \{q_i\}_{i=1}^{s}$, such that $\langle B \rangle = J$ is an ideal on the form of conjecture 4.4, and let all the q_i's be on the same form, but with different arbitrary coefficients.

- Choose $2k$ arbitrary monomials, u, where we have $\{u_{2i-1} \cdot u_{2i}\}_{i=1}^{k}$ such that $u_{2i-1} \cdot u_{2i} \neq u_{2j-1} \cdot u_{2j}$ for all $1 \leq i \neq j \leq k$. In addition, we must ensure that $\text{tip}(q_i) = T \leq \max_{1 \leq i \leq k}(u_{2i-1} \cdot u_{2i})$.

- Let $k_i = k$ such that for $1 \leq i \leq s$, we set $\{F_{ij} \cdot H_{ij}\}_{j=1}^{k} = \{a_{2i-1}u_{2i-1} \cdot a_{2i}u_{2i}\}_{i=1}^{k}$ for some arbitrary constants $\{a_j\}_{i=1}^{2k}$.

- If the decryption algorithm reduces the ciphertext $c \xrightarrow{g} c'$, where at least one term in c' is in $\text{NonTip}(\langle g \rangle) \setminus M$, then c is returned without reduction.

In general, a cryptosystem on this form seems promising, but the different parameters need have a certain magnitude. Also, it is important to design B and $\{u_j\}_{j=1}^{2k}$ such that the encryption $c = p + m = \sum_{i=1}^{s} \sum_{j=1}^{k_i} F_{ij} q_i H_{ij} + m$, provides p to have a significant number of terms in the message space, M.

7.4 Conclusion

We know that the cryptosystem presented above is protected against the linear algebra attacks presented in section 6.2, and the private key seems infeasible to obtain by a cryptanalyst, even if he finds $\text{tip}(g)$. Nevertheless, there are three weakness's in this system that we must consider:

1. Conjecture 4.4 is just on experimental level, and there exists no unconditional guarantee that an ideal generated by polynomials on the same form as in the conjecture will have an infinite Gröbner basis.

2. The encryption method, b), presented in section 5.3 is also on an experimental level. If the cryptanalyst is unable to compute a partial Gröbner basis and we have several monomials $F_{ij} \cdot H_{ij} > T$, correctly reduction by the public key seems certain to fail. However, a proof is infeasible to give at this moment.

3. A public key cryptosystem needs a common message space, M, for an arbitrary number of users. That leaves this system vulnerable to an attack described in section 6.1, where one can decrypt a ciphertext constructed by another person by disguising it as your own ciphertext, using the decryption black box.

It seems hard to prove that a cryptosystem can be secured from all these weakness's. However, there are several promising aspects behind these points, and this should provide that there will be done more research in order to make this cryptosystem secure.

References

[AdLo] W. Adams and P. Loustaunau: An Introduction to Gröbner Bases. Amer. Math. Soc., Providence, 1994.

[Be] G. Bergman: The diamond lemma for ring theory, Adv. Math. 29, 1978, pp 178-218.

[Bu] B. Buchberger: Ein Algorithmus zum Auffinden der Basiselemente des Restklasserings nach einem nulldimensionalen Polynomideal, Ph.D. Thesis, Inst. University of Innsbruck, Innsbruck, Austria, 1965.

[Bu2] B. Buchberger: Gröbner bases: an algorithmic method in polynomial ideal theory. In N. K. Bose, editor, Multidimensional Systems Theory, Mathematics and its Applications, pages 184-232. D. Reidel Publishing Company, Dordrecht, Holland, 1985.

[FeKo] M. Fellows and N. Koblitz: Combinatorial cryptosystems galore! Contemporary Math. 168, 1994, 51-61.

[GHK] E. Green, L. S. Heath and B. J. Keller: Opal: A system for computing noncommutative Gröbner bases (system description). Eighth International Conference on Rewriting Techniques and Applications (RTA-97), 1997, pp. 331-334.

[Ko] N. Koblitz: Algebraic aspects of cryptography, Algorithms and computations in Math., vol. 3, Springer, 1997.

[Mo] T.Mora et al: [pseudonyms Boo Barkee, Deh Cac Can, Julia Ecks, Theo Moriarty, R.F. Ree] Why you cannot even hope to use Gröbner bases in public key cryptology: An open letter to scientist who failed and a challenge to those who have not yet failed, Jour. Symb. Comput., vol. 18, 1994, pp 497-501.

[Ra] T. S. Rai: Infinite Gröbner Bases And Noncommutative Polly Cracker Cryptosystems. Ph.D. Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 2004.

[RaBu] T. S. Rai and S. Bulygin: Noncommutative Polly Cracker-type cryptosystems and chosen-ciphertext security, 2008.