Semirings and Semigroup Actions in Public-Key Cryptography

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Christopher J. Monico, B.S., M.S.

Joachim Rosenthal, Director

Department of Mathematics
Notre Dame, Indiana
April 2002

Semirings and Semigroup Actions in Public-Key Cryptography

Abstract

by

Christopher J. Monico

In this dissertation, several generalizations of cryptographic protocols based on the Discrete Logarithm Problem (DLP) are examined.

It is well known that the Pohlig-Hellman algorithm can reduce the computation of a DLP in an abelian group to the computation of DLPs in simple abelian groups. As an example of this, we consider the DLP in rings of the form $\mathbb{F}_q[\underline{x}]/I$, where I is a zero-dimensional ideal. This example culminates in an interesting primary decomposition algorithm for zero-dimensional ideals (over $\mathbb{Q}$).

In the next chapter, we consider the possible difficulty of the DLP in semirings. Since more general versions of the Pohlig-Hellman algorithm may apply, an extended discussion of finite, additively commutative, congruence-free (i.e., simple) semirings follows. We classify such semirings, except for the additively idempotent ones.

Finally, a generalization of the DLP itself is discussed. It is shown that every semigroup action on a finite set gives rise to a Diffie-Hellman type protocol. A Pollard-rho type algorithm is given for solving instances of the group action problem. A particular semigroup action of $\mathrm{Mat}_n(\mathbb{Z})$ on H^n, where H is an abelian semigroup, is discussed as an example where the semigroup action problem may be hard enough to build a cryptosystem on it.

To my parents for a lifetime of love and support, and to my soulmate, Joy.

CONTENTS

ACKNOWLEDGEMENTS

I would be greatly remiss not to acknowledge those who have made this work possible. Certainly, I am indebted to the University of Notre Dame, the Department of Mathematics and the Center for Applied Mathematics of Notre Dame (CAM). In addition to providing generous financial support, the department and CAM have provided a wonderful environment fostering intellectual stimulation, friendships and general academic enrichment. Simply being here has been a truly rewarding experience in its own right.

I owe many thanks to my advisor and friend, Joachim Rosenthal. Without his help, support, and confidence in me, I would have been lost.

The staff of the mathematics department have been more than helpful; neverminding their high level of competence in performing their regular duties, they do an extraordinary job of providing smiles and brightening up my day, which has not gone unnoticed or unappreciated.

There are many individuals to whom I am specifically indebted. Professor Alex Hahn was integral in the completion of my education. It was Steve Peterson that saw the mathematician within me and inspired me to pursue it; but it was the entire Department of Mathematics at Monmouth University that made it possible, including Barbara Lynn Bodner, Tom Smith, Judy Tobin, and Boyd Swartz.

I am also grateful to Karen Chandler, Mara Neusel, and Andrew Sommese for their valuable feedback and suggestions, which greatly improved the quality of this work.

CHAPTER 1

INTRODUCTION

Cryptology is the study of methods of secure communication. The general situation is that two parties, often called Alice and Bob, wish to communicate securely with each other. The problem for Alice and Bob is that there may be one or more eavesdroppers present. For simplicity, we assume that there is one eavesdropper and call her Eve.

The information that Alice wishes to communicate to Bob (or vice-versa) is called *plaintext*. Since most information that Alice and Bob could wish to exchange can be represented digitally, we assume that the plaintext is an element of a finite set (i.e., perhaps $\mathbb{Z}_{256}$ or $(\mathbb{Z}_{256})^n$). Generally, they will achieve secure communication by disguising the plaintext so that it is meaningless to Eve. Such a disguised version of plaintext is called *ciphertext*, and can also be considered as an element of a finite set. The challenge for Alice and Bob is to produce ciphertext satisfying three requirements:

- Two distinct elements of plaintext result in two distinct elements of ciphertext.
- If Alice receives an element of ciphertext from Bob, she can recover the corresponding plaintext (or vice-versa).
- For an arbitrary element of ciphertext, Eve cannot feasibly determine the corresponding plaintext.

The first two requirements insure reliable communication between Alice and Bob. The final requirement serves to assure Alice and Bob that they are communicating securely.

The purpose of this dissertation is to study some possible methods by which Alice and Bob can produce ciphertext satisfying these criteria.

1.1 Background

There are three models of secure communication, summarized in Figures 1.1, 1.2, and 1.3, which cover most modern study of cryptology. They fall into two more general classes: *symmetric ciphers* and *asymmetric ciphers.*

Suppose Alice wishes to securely communicate some plaintext to Bob. She generally accomplishes this by applying an encryption function F_1 to the plaintext, obtaining ciphertext. Of course an arbitrary such function will not do; Bob must have the inverse function F_2, and it should not be easy for an eavesdropper to recover the plaintext from the ciphertext.

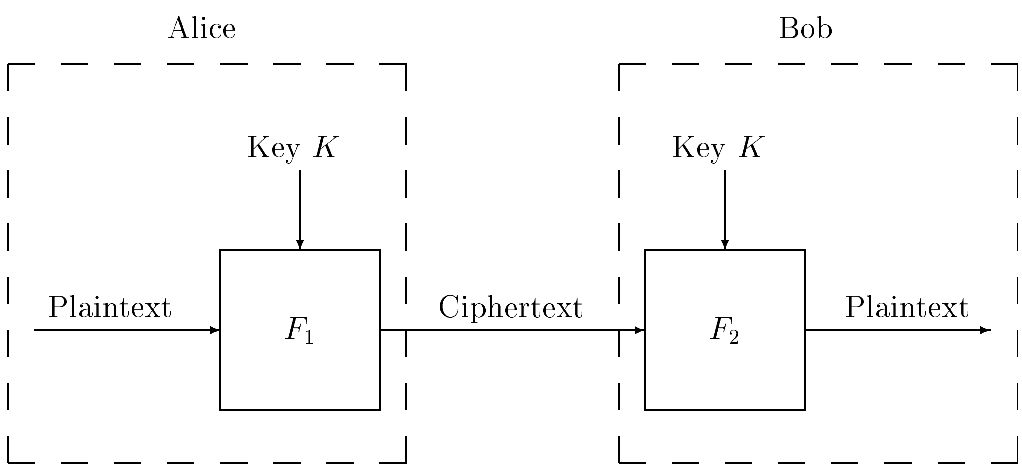

Figure 1.1. Symmetric cipher model

To facilitate implementation, F_1 and F_2 are generally chosen from some widely known class of functions. To prevent a malicious eavesdropper from effectively guessing the functions and recovering the plaintext, these functions generally accept two inputs: the plaintext and a *key.* Herein lies the difference between the symmetric model and the asymmetric model. In the symmetric model, Alice and Bob must a priori have the same key K. However, in the asymmetric model this is not the case.

Figure 1.2 shows an asymmetric model in which Alice and Bob first go through a negotiation phase to agree upon a shared key (which an eavesdropper can presumably not determine due to the nature of the negotiation). Figure 1.3 shows an asymmetric model in which there are two distinct keys: the *encryption key* and the *decryption key.* This enables Bob to freely distribute his encryption key, confident that he is the only person with the decryption key.

Until the 1970's the only model that was studied was the symmetric cipher

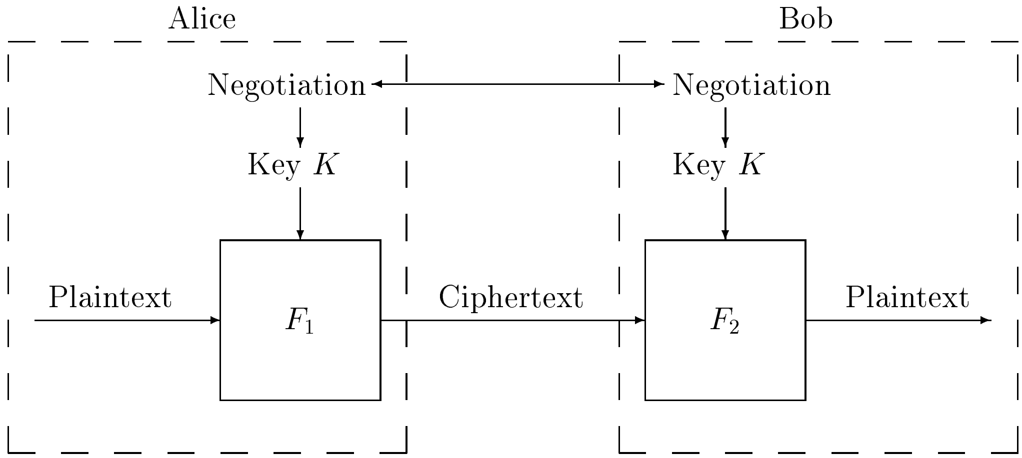

Figure 1.2. One asymmetric cipher model

model. Claude Shannon had already, in the 1950's, begun to study cryptography as a mathematical discipline. He succeeded in showing that the *one-time pad* is an "unconditionally-secure" implementation of a symmetric cipher [43]. However, it was still believed that in order for Alice and Bob to communicate securely with each other, they must already share a secret key.

It was the observation of Whitfield Diffie and Martin Hellman that the secret key they must share can be arbitrary in some sense. Neither Alice nor Bob must necessarily choose the key, so long as they can agree on it without an eavesdropper being able to determine it. This observation, in the 1970's, lead to the *Diffie-Hellman key exchange* [12, 13].

Protocol 1.1 (Diffie-Hellman Key Exchange)

1. Alice and Bob agree on some finite group G and an element $g \in G$.

2. Alice privately chooses an integer a and computes $\alpha = g^a$. She sends α to Bob.

3. Bob privately chooses an integer b and computes $\beta = g^b$. He sends β to Alice.

4. Alice and Bob can both compute

$$k = g^{ab} = \beta^a = \alpha^b.$$

Observe that this, together with a symmetric cipher accepting the key k, is an implementation of the asymmetric model in Figure 1.2. Suppose now that an

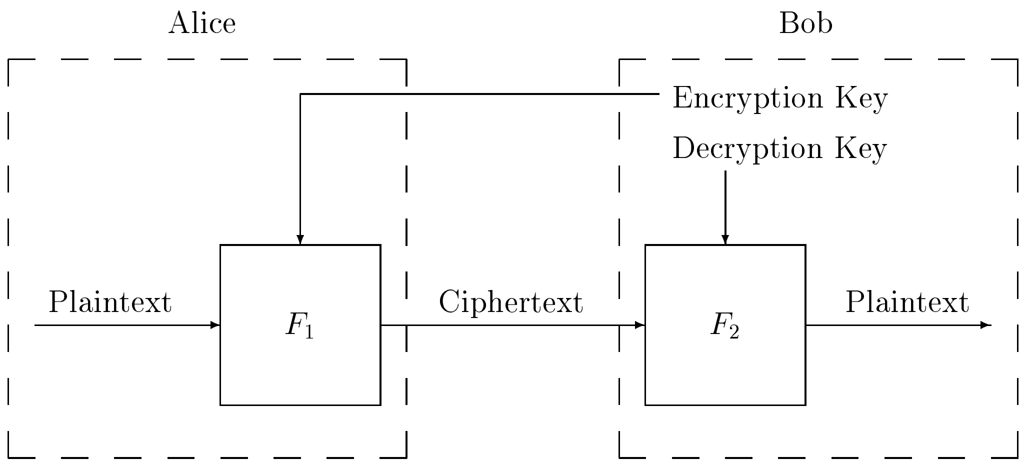

Figure 1.3. Another asymmetric cipher model

eavesdropper Eve wishes to find k. Eve certainly knows G, g, α, and β. However, there is no obvious way for her to find k without knowing either a or b. So, if the problem of finding n given g and g^n is hard, the problem of finding k in this scenario is possibly hard as well. Thus Alice and Bob can probably rest assured that if they do this in a smart way, Eve will not be able to find k. They are then free to use k as the key for their favorite cipher and communicate securely with each other.

One popular implementation of the asymmetric model in Figure 1.3 is the RSA public key cryptosystem [32], named after its founders Ron Rivest, Adi Shamir, and Leonard Adleman. Bob can efficiently generate two large random primes p and q, and compute $n = pq$ [21]. Knowing the factorization of n, Bob can easily compute $\phi(n) = (p-1)(q-1)$, and choose an integer $e > 1$ with $(e, \phi(n)) = 1$. He then freely distributes his encryption key (n, e). If Alice wishes to send Bob a message $m \in \mathbb{Z}_n$, she applies the encryption function $m \mapsto m^e \pmod{n}$.

This map appears to satisfy all of the required criteria so long as n is hard to factor. That is, since Bob knows $\phi(n)$, he may find an integer d such that $ed \equiv 1 \pmod{\phi(n)}$ and decrypt received message with the map $x \mapsto x^d \pmod{n}$.

Notice that there is an important distinction between these two methods; using the Diffie-Hellman key exchange requires bi-directional communication (negotiation) while the RSA method does not. However, with a little extra effort one can do away with the negotiation step in the Diffie-Hellman key-exchange and obtain the ElGamal public key cryptosystem [15].

Protocol 1.2 (ElGamal public key cryptosystem)

1. Alice chooses a finite abelian group G and $g \in G$. She also chooses an integer a and computes $\alpha = g^a$. She publishes her public key (g, α).

2. Bob wishes to send Alice the message $m \in G$. He first obtains her public key (g, α).

3. Bob chooses a random integer b and computes $\beta = g^b$ and $\mu = m\alpha^b$. He sends the pair (β, μ) to Alice.

4. Alice recovers m by computing

$$\mu\beta^{-a} = \mu g^{-ab} = m(g^a)^b g^{-ab} = m.$$

There is little difference between ElGamal encryption and Diffie-Hellman key exchange except the observation that, Alice may publish the pair (g, α) as a public key, since she can presumably compute inverses in the group G.

This dissertation will study asymmetric ciphers only. Symmetric ciphers, for which a shared secret key is already available, generally have the luxury of concentrating simultaneously on efficiency and security. As such, they are often constructed with specific hardware in mind using some complicated series of elementary operations such as bitwise exclusive-or and bitwise permutations. Such ciphers generally do not lend themselves to nice algebraic descriptions. However, symmetric ciphers are still heavily studied and used. In practice, asymmetric techniques are often used only to agree upon a key for use in a symmetric cipher. This is primarily because implementations of symmetric ciphers are faster and consume less resources than their asymmetric counterparts.

There is one central concept of cryptography that is difficult to define precisely. Specifically, we often require that the solution to some particular problem be "hard" to compute. Of course, one could make this well-defined by adopting the conventions in complexity theory, and calling a problem hard if it is NP-complete or NP-hard. However, this turns out to be inconvenient, since the complexity classes of many problems central to cryptography are unknown. Furthermore, since it is unknown whether P and NP are equal, such a definition would be truly meaningless from an application perspective.

Instead, we adopt the convention of calling a problem *hard* if there is no known polynomial-time algorithm for solving it. Such a definition, while not rigorous, allows for the possibility that all NP problems are hard, while also allowing for the existence of hard problems even if P=NP. Of course this is far from rigorous, and requires some intuition and judgment in usage.

This apparent lack of rigor does not hinder cryptography in its mathematical study, however. Being careful to qualify statements with "We believe X is hard" or "If X is hard, ...", one can still make precise statements around these imprecise concepts.

1.2 Formal introduction

We provide now the formal introduction of the concepts and terms that will be necessary in the sequel.

Problem 1.3 Let G be a finite cyclic group with generator $x \in G$ and an element $y \in G$. The *Discrete Logarithm Problem* (DLP) asks for an integer k such that $x^k = y$. The least non-negative integer k with this property is denoted by $\log_x y$.

Observe that this definition easily extends to an arbitrary group, so long as $y \in \langle x \rangle$. Compare this definition with the following:

Problem 1.4 Let G be a finite cyclic group with generator $g \in G$ and elements $\alpha, \beta \in G$. The *Diffie-Hellman Problem* (DHP) asks for $\gamma \in G$ such that $\gamma = g^{(\log_g \alpha)(\log_g \beta)}$.

Certainly a polynomial-time algorithm for solving the DLP in a particular group extends to a polynomial-time algorithm for solving the DHP. It is unknown if these two problems are equivalent in general, though some progress has recently been made toward showing an equivalence with some additional assumptions (see, e.g. [32, Fact 3.77][29]). In any case, if one wishes to use the DHP in a particular group as the basis of a cryptosystem, it is necessary that the DLP be hard in that group.

Many groups have been proposed for which the Diffie-Hellman problem may be hard and used securely. However, in practice there two that are most often used. One is the multiplicative group $(\mathbb{F}_q)^*$ of a finite field of order q. It is a slight modification of this that is employed in the Digital Signature Algorithm (DSA) [32]. There is, however, a subexponential-time algorithm for solving the DLP in this group, called the index-calculus algorithm. In the case of prime fields, there is a more advanced version of the index-calculus known as the number field sieve, to solve the DLP with expected time [18]

$$O(exp((1.923 + o(1))(\ln q)^{1/3}(\ln \ln q)^{2/3})).$$

If $q = 2^m$, there is a variation of the index-calculus known as Coppersmith's algorithm, to solve the DLP with expected time [10]

$$O(exp((c + o(1))(\ln q)^{1/3}(\ln \ln q)^{2/3}))$$

for some $c < 1.587$. Note that the DLP is still considered hard in these groups because the runtimes of these algorithms are not bounded above by any polynomial in $\ln q$. However, the existence of these subexponential-time algorithms means that one must use larger key sizes than if only exponential-time attacks were known. For example, it is recommended that for prime fields, p should have at least 1024 bits [32, Note 8.24].

Another group that has received much attention in this context is the group of rational points on an elliptic curve over a finite field [22, 23, 8]. The best known algorithm for solving the DLP in such a group is the Pollard-rho algorithm for discrete logs, which works in any finite group. It has expected runtime $O(\sqrt{n})$, where n is the order of the group. If n has just 150 bits and the elliptic curve is chosen judicially, the DLP is already very formidable. With at least 200 bits and a good curve, it is considered secure by the cryptographic community. The drawback of this system, however, is that the group operations are rather expensive relative to the operations of addition and multiplication in a finite field. This is primarily because an addition of two points requires the inversion of an element from the underlying field.

These two examples provide the motivation for much of this dissertation. Ideally, one would like to have a group where the DLP is hard with small key sizes, but the group operations themselves are very inexpensive; i.e., smaller keys than with $(\mathbb{F}_q)^*$ and faster than with elliptic curve groups. More generally, one would prefer to have any asymmetric system with small key sizes where the underlying (arithmetic) operations are inexpensive.

Finally, we give some notations that are used throughout this dissertation. For a finite set X, we denote the cardinality of X by $|X|$. If s and t are integers, we denote the greatest common divisor of s and t by (s, t). For any ring R with 1, we will let R^* be the group of multiplicatively invertible elements of R.

In the next two sections we present two important algorithms that generalize to many settings. In searching for generalizations of known DLP settings, they may be viewed as the first tools one should use to determine potentially secure parameters.

1.3 Pollard-rho for discrete logarithms

First we present the Pollard-rho algorithm for solving the DLP [41]. It is important because generalized versions of it apply in many scenarios. We will see one such generalization in Chapter 4, Algorithm 4.4.

Algorithm 1.5 Pollard-rho for DLP
Input: A finite cyclic group G with order $n = |G|$, generator α, and an element $\beta \in G$.
Output: $\log_\alpha \beta$.

1. Choose a random partition $G = S_1 \cup S_2 \cup S_3$ of roughly equal size. For $(\zeta, a, b) \in G \times \mathbb{Z}_n \times \mathbb{Z}_n$ define

$$F(\zeta, a, b) = \begin{cases} (\zeta\alpha, a+1, b) & \text{if } \zeta \in S_1, \\ (\zeta\beta, a, b+1) & \text{if } \zeta \in S_2, \\ (\zeta^2, 2a, 2b) & \text{if } \zeta \in S_3. \end{cases}$$

2. Choose two random elements $a_1, b_1 \in \mathbb{Z}_n$. Set

$$\zeta_1 \leftarrow \alpha^{a_1}\beta^{b_1} \qquad \text{and} \qquad (\zeta_2, a_2, b_2) \leftarrow F(\zeta_1, a_1, b_1).$$

   Set $i \leftarrow 1$.

3. If $\zeta_i = \zeta_{2i}$ and $b_{2i} - b_i$ is invertible modulo n, output $(a_i - a_{2i})/(b_{2i} - b_i)$ and terminate. If $\zeta_i = \zeta_{2i}$ and $b_{2i} - b_i$ is not invertible, goto 2.

4. Set

$$\begin{aligned} (\zeta_{i+1}, a_{i+1}, b_{i+1}) &\leftarrow F(\zeta_i, a_i, b_i), \\ (\zeta_{2i+1}, a_{2i+1}, b_{2i+1}) &\leftarrow F(\zeta_{2i}, a_{2i}, b_{2i}), \\ (\zeta_{2i+2}, a_{2i+2}, b_{2i+2}) &\leftarrow F(\zeta_{2i+1}, a_{2i+1}, b_{2i+1}), \end{aligned}$$

   and $i \leftarrow i + 1$. Goto 3.

Note that in step 3, if $\zeta_i = \zeta_{2i}$ and n has only large prime factors, then $b_{2i} - b_i$ is invertible with high probability. We will see in the next section that in fact n may be assumed to be prime.

The success of the Pollard-rho algorithm is due to the so-called birthday problem. If the partition chosen is sufficiently random, one may consider the sequence $\zeta_1, \zeta_2, \ldots$ to be a pseudo-random sequence from G, in which each element of the sequence depends only on the previous element. In this case, one expects that after approximately the first $\sqrt{n\pi/2}$ elements the sequence begins to repeat. When the sequence begins to repeat, this will be detected shortly after in step 3.

Some work has been done by Teske[45] and Wiener and Zuccherato[50] that suggests using a partition into more subsets may reduce the constant in the expected runtime. The work of Wiener and Zuccherato also provides some additional improvement in the constant in the case where the underlying group is the rational points on an elliptic curve over a finite field.

A consequence of the Pollard-rho algorithm is that there is a minimum key size we may hope for when using the DLP. One should probably have at least 150 bits for a system to have any chance at being secure.

1.4 The Pohlig-Hellman algorithm

We present here the Pohlig-Hellman algorithm for computing discrete logarithms [40]. If G is not simple, it will reduce a DLP in G to several DLPs in smaller groups. It thus restricts the possible choices for groups in which the DLP is maximally hard. The idea is to take advantage of many-to-one homomorphic images in which the DLP is easier to solve. This will be important in the sequel because generalizations suggest that, even if one considers structures other than groups for the DLP, one should consider using simple structures. Specifically, using simple structures is the most reliable way to avoid similar attacks. Our presentation follows that from [32].

Algorithm 1.6 Pohlig-Hellman
Input: A finite cyclic group G with order $n = |G|$, generator α, and an element $\beta \in G$. The prime factorization $n = p_1^{e_1} \cdots p_k^{e_k}$.
Output: $\log_\alpha \beta$.

1. For i from 1 to k do the following:
   (a) *(Simplify notation)* Set $q \leftarrow p_i$ and $e \leftarrow e_i$.
   (b) Set $\gamma \leftarrow 1$ and $l_{-1} \leftarrow 0$.
   (c) Compute $\bar{\alpha} \leftarrow \alpha^{n/q}$.
   (d) For j from 0 to $e-1$ do the following:
      - Compute $\gamma \leftarrow \gamma\alpha^{l_{j-1}q^{j-1}}$ and $\bar{\beta} \leftarrow (\beta\gamma^{-1})^{n/q^{j+1}}$.
      - Compute $l_j \leftarrow \log_{\bar{\alpha}} \bar{\beta}$ (For example, using the Pollard-rho algorithm).
   (e) Set $x_i \leftarrow l_0 + l_1 q + \cdots + l_{e-1} q^{e-1}$.

2. Compute the least nonnegative integer, x, such that $x \equiv x_i \pmod{p_i^{e_i}}$ for $1 \leq i \leq k$.

3. Output x.

Note that there are several well-known algorithms for performing step 2 in polynomial time.

The idea is that in step 1 one can find $x_i = \log_\alpha \beta \pmod{p_i^{e_i}}$. This is further reduced by finding the base-p expansion of x_i one digit at a time. Observe that each time step 1(d) is reached, $\bar{\alpha}$ has order p_i. Thus, one need only compute discrete logarithms in (sub)groups of order p_i. Unless there is some trivial way to solve the DLP in G, this is more efficient than computing one discrete logarithm in the full group with order n. It is thus desirable that one should choose G with prime order so this algorithm yields no reduction at all.

More generally, observe that if G is not simple, there exist groups G_i and homomorphisms of the form

$$f_i : G \longrightarrow G_i$$

with nontrivial kernels. One may then solve the corresponding DLP in each homomorphic image. Furthermore, if there exist such G_i such that

$$\begin{aligned} f : G &\longrightarrow G_1 \times \cdots \times G_k \\ g &\longmapsto (f_1(g), \cdots, f_k(g)) \end{aligned}$$

is a monomorphism, then solving the DLP in each G_i solves the DLP in G up to an application of the Chinese remainder theorem.

1.5 Pohlig-Hellman type reductions for rings

The Pohlig-Hellman algorithm implies that if G is a finite, non-simple group, one can solve the discrete logarithm problem in G using fewer than $O(\sqrt{|G|})$ operations. The goal of this section is to motivate an extension of this to the case of rings. Specifically, if R is a non-simple ring for which one can compute at least one nontrivial homomorphism $f : R \longrightarrow R/I$, then f can be used to simplify the computation of discrete logarithms in R. At the very least, we show that to avoid attacks needing fewer than $O(\sqrt{|R|})$ operations, one must have $|R| = p^k$ for some prime p. Even in this situation, if R is not simple one would be relying on the difficulty of not only the DLP, but also the difficulty of finding efficiently computable homomorphisms with non-trivial kernels.

Of course, since $(R, \cdot)$ is not a group (unless R is a field), the reduction is not quite as easy as with Pohlig-Hellman. There is, however, one case which is nearly as easy. If $|R| = st$ with $(s,t) = 1$, then one may find integers u and v such that $us + vt = 1$. It is not hard to see that $f(r) = (us)r$ is ring homomorphism whose image is isomorphic to R/I_s, where $I_s = \{x \in R \,|\, sx = 0\}$ is a proper ideal. This observation will give rise to Algorithm 1.9, a variant of Pohlig-Hellman that shows one should only consider rings with prime power order. Some technical details are in order before we can present this algorithm.

Lemma 1.7 *Let R be a finite ring, $x \in R$ and $0 < a < b$ integers such that $x^a = x^b$. Then there exists $k \geq 1$ such that for all $n \geq a$, $x^a = x^n$ if and only if $n \equiv a \pmod{k}$.*

Proof: Let $\beta > a$ be the smallest integer such that $x^a = x^\beta$. Set $k = \beta - a$. Then $x^{a+k} = x^{a+\beta-a} = x^\beta = r^a$. By induction, $x^{a+lk} = x^a$ for all $l \geq 0$. Thus, $n \equiv a \pmod{k}$ and $n \geq a$ imply $x^n = x^a$.

The converse is clearly true if $k = 1$, so assume $k > 1$ and suppose that $x^a = x^n$ and $n \geq a$. Then there exists $l \geq 0$ such that

$$a + lk \leq n < a + (l+1)k.$$

Let $n = a + lk + j$. We will show that $j = 0$. Certainly $x^a = x^n = x^{lk+a+j} = x^{a+j}$. But $lk + a + j < a + lk + k$ is equivalent to $j < k$. By assumption, k is the least positive integer such that $x^a = x^{a+k}$, whence $j = 0$. $\square$

The integer k in Lemma 1.7 is called the *cycle length* of x, and denoted by L_x. The least positive integer s_x such that

$$x^{s_x} = x^{s_x + L_x}$$

is called the *cycle start* of x. It's easy to see for all $0 \neq x \in R$, the cycle length and cycle start of x are well-defined. That is, for all $x \in R$ there exist least positive integers, L_x, s_x, so that for all $c, d \geq s_x$ the following holds:

$$x^c = x^d \Leftrightarrow c \equiv d \pmod{L_x}.$$

Furthermore, if $|R| = N$ and $x \in R$, one can compute L_x with at most $O(\sqrt{N} \ln \sqrt{N})$ operations by the following algorithm.

Algorithm 1.8 (Baby-step giant-step for cycle length)
Input: A finite ring R with $|R| = N$ and an element $x \in R$.
Output: The cycle length of x.

1. Set $m \leftarrow \left\lceil \sqrt{N} \right\rceil$. Choose a prime $q > N$.

2. For $0 \leq i \leq m$, compute and store in a table the pairs (i, x^{q+im}). Sort the table by the second component.

3. Find the least positive integer b_1 such that x^{q+b_1} is in the table: $x^{q+b_1} = x^{q+a_1 m}$. *(Note:* $0 < b_1 < m$*).*

4. Find the least positive integer b_2 such that x^{2q+b_2} is in the table: $x^{2q+b_2} = x^{q+a_2 m}$. *(Again,* $0 < b_2 < m$*).*

5. Compute $g \leftarrow (a_1 m - b_1, a_2 m - b_2 - q)$.

6. For each divisor d of g below some bound B, do the following:

   - If $x^{N+g/d} = x^N$, set $g \leftarrow g/d$.

7. Output $L_x = g$ and stop.

The exact bound, B, is not easy to calculate. However, it is certainly below $\sqrt{a_1 m - b_1}$. We conjecture that the probability of the output being correct is at least

$$1 - \sum_{\substack{p|g \\ p>B}} \frac{1}{p},$$

which would certainly suffice for practical purposes, with even a fixed bound of around 10^6. The algorithm could also be modified to find several such a_i, b_i and compute the gcd over all matches, further increasing the probability of success.

Also observe that in step 1, m is chosen to be $\left\lceil \sqrt{N} \right\rceil$ since N is an upper bound on the cycle length. If a better upper bound is known, one can use it to further increase the efficiency.

We also remark that given L_x, one may compute s_x with a simple binary search using $O(\ln N)$ operations. Then, given L_x and s_x, the discrete logarithm problem $y = x^e$ can be solved using a straightforward variation of the Pollard-rho algorithm.

We are now in the position to give a first reduction algorithm for the DLP in R when $|R| = st$ with $(s,t) = 1$.

Algorithm 1.9 (Split Pohlig-Hellman)
Input: A finite ring R with $|R| = st$ and $(s,t) = 1$. Elements $x, y \in R$ such that $y = x^e$ for some $e > 0$.
Output: A positive integer $\hat{e}$ such that $y = x^{\hat{e}}$.

1. Use the extended Euclidean algorithm to find integers u and v such that $us + vt = 1$. For $r \in R$, let $f_s(r) = (us)r$ and $f_t(r) = (vt)r$.

2. Use Algorithm 1.8 to compute $l_1 \leftarrow L_{f_s(x)}$. Use Pollard-rho to find the least positive integer e_1 such that $f_s(y) = f_s(x)^{e_1}$. If $y = x^{e_1}$, output e_1 and stop.

3. Use Algorithm 1.8 to compute $l_2 \leftarrow L_{f_t(x)}$. Use Pollard-rho to find the least positive integer e_2 such that $f_t(y) = f_t(x)^{e_2}$. If $y = x^{e_2}$, output e_2 and stop.

4. Use the Chinese remainder theorem to find the smallest integer $\hat{e} \geq \max\{e_1, e_2\}$ such that $\hat{e} \equiv e_1 \pmod{l_1}$ and $\hat{e} \equiv e_2 \pmod{l_2}$. Output $\hat{e}$ and stop.

Remark 1.10 The correctness of the algorithm follows from the fact that $f_s(R) \cong R/I_s$, $f_t(R) \cong R/I_t$ and $R \cong R/I_s \times R/I_t$. Furthermore, step 2 is accomplished with $O(\sqrt{t} \ln \sqrt{t})$ operations and step 3 is accomplished with $O(\sqrt{s} \ln \sqrt{s})$ operations. This is certainly less than $O(\sqrt{N})$, so this is, indeed, a simplification of the problem. Furthermore, this algorithm can easily be called recursively (or adapted) to handle multiple coprime factors. So, if one wishes to find a finite ring, R, where the best known solution of the DLP needs $O(\sqrt{|R|})$ operations, it must be the case that $|R| = p^k$, for some prime, p.

Example 1.11 For a small example, suppose we wish to compute $e = \log_{11} 726$ in $\mathbb{Z}_{1829}$. Since $1829 = 31 * 59$ and $10 * 59 - 19 * 31 = 1$, our "splitting homomorphisms" are $f_{31}(r) = (-19 * 31)r = 1240r$ and $f_{59}(r) = (10 * 59)r = 590r$. We then have

$$f_{31}(726) = f_{31}(11)^e \iff 372 = 837^e.$$

It is easily verified that the cycle length of 837 is 58 in $\mathbb{Z}_{1829}$, so with about $\sqrt{58} \approx 8$ iterations, Pollard-rho will discover that $e = 11 + 58k_1$, for some k_1. Similarly,

$$f_{59}(726) = f_{59}(11)^e \iff 354 = 1003^e.$$

The cycle length of 1003 in $\mathbb{Z}_{1829}$ is 30, so with about $\sqrt{30} \approx 6$ iterations, Pollard-rho will discover that $e = 7 + 30k_2$ for some k_2.

Finally, the least positive integer $\hat{e}$ satisfying

$$\begin{cases} \hat{e} & \equiv & 11 \pmod{58} \\ \hat{e} & \equiv & 7 \pmod{30} \end{cases}$$

is 127. We verify that indeed $11^{127} \equiv 726 \pmod{1829}$. Furthermore, this was found with approximately 14 iterations. The cycle length of 11 in $\mathbb{Z}_{1829}$ is 870, so we would have needed about $\sqrt{870} \approx 30$ iterations without the split.

Suppose now that $y, x \in R$ with $y = x^e$ for some e. Further suppose that e_1 and k are known integers such that $e = e_1 + sk$ for some $s \geq 0$. We will show how this additional information about the exponent can be used to simplify the computation of $\log_x y$.

Algorithm 1.12 (Modified baby-step/giant-step)
Input: A finite ring R with $N = |R|$. Elements $y, x \in R$, an integer $e_1 \geq 0$, and an integer $k > 0$ such that $y = x^{e_1+sk}$ for some $s \geq 0$.
Output: A non-negative integer s such that $y = x^{e_1+sk}$.

1. If $y = x^{e_1}$, output $s = 0$ and stop.
2. Set $m \leftarrow \left\lceil \sqrt{\frac{N-e_1}{k}} \right\rceil$.
3. For $1 \leq i \leq m$ compute and store in a table the pairs (i, x^{e_1+imk}).
4. Sort the table with the second component as the primary key, first component as the secondary key.
5. Set $j \leftarrow 0$.
6. Compute x^{jk}. If $y = x^{jk}x^{e_1}$, output j and stop.
7. Compute yx^{jk}, and test to see if it is in the lookup table. If it is not goto step 9.
8. Let i_0 denote the first occurrence of yx^{jk} in the lookup table, and i_1 the second (if there are multiple occurrences). If $y = x^{e_1+(i_0m-j)k}$, output $s = i_0m - j$ and stop. If there were multiple occurrences and $y = x^{e_1+(i_1m-j)k}$, output $s = i_1m - j$ and stop.
9. Set $j \leftarrow j + 1$ and goto step 6.

Proof: (of correctness) By assumption, there exists s such that $y = x^{e_1+sk}$. Since we may assume $e_1 + sk < N$, it follows that $s < \frac{N-e_1}{k}$. Thus, there exist $0 \leq i, j < m$ such that $s = im + j$. Assume that s is the least non-negative integer such that $y = x^{e_1+sk}$ and $0 \leq i, j < m$ are the unique integers such that

$$s = im + j.$$

If $i = 0$, the solution will be found at the j-th iteration of step 6, since $y = x^{e_1+jk}$. We may thus assume that $i \geq 1$. Then $y = x^{e_1+imk+jk}$, so

$$yx^{(m-j)k} = x^{e_1+imk+jk+mk-jk} = x^{e_1+(i+1)mk}$$

and $yx^{(m-j)k}$ is in the lookup table because $1 \leq \hat{j} < m$, where $\hat{j} = m - j$. We will now show that if there are multiple occurrences of $yx^{\hat{j}}$ in the lookup table, the solution will be found with one of the first two occurrences.

Let $i_0 < i_1 < \ldots < i_t$ be all the occurrences of $yx^{\hat{j}}$ in the table so that

$$yx^{\hat{j}} = x^{e_1+i_0mk} = x^{e_1+i_1mk} = \cdots = x^{e_1+i_tmk}.$$

Let $\alpha = \min\{i_{z+1} - i_z\}$. We will first show that $i_n = i_0 + n\alpha$. To see this, let $z \geq 0$ such that $i_{z+1} - i_z = \alpha$. Then

$$\begin{aligned} x^{e_1+i_0mk+\alpha mk} &= x^{e_1+i_0mk+i_{z+1}mk-i_zmk} \\ &= x^{e_1+i_zmk+i_{z+1}mk-i_zmk} \\ &= x^{e_1+i_{z+1}mk} \\ &= x^{e_1+i_0mk}. \end{aligned}$$

By induction, $x^{e_1+(i_0+n\alpha)mk} = x^{e_1+i_0mk}$ for all $n \geq 0$. But if $i_{n+1} < i_0 + (n+1)\alpha$ for some n, we would have $i_{n+1} - i_n < (n+1)\alpha - n\alpha = \alpha$, a contradiction. Thus, $i_n = i_0 + n\alpha$.

We know that some $i_n = i$ will give the solution. If $n = 0$ or $n = 1$, this will be one of the first two occurrences of $yx^{\hat{j}}$ in the table, and it will be found. Suppose that $n > 1$. We will show that i_1 also gives the solution. Observe that

$$\begin{aligned} y &= x^{e_1+(i_nm-\hat{j})k} \\ &= x^{e_1+(i_0mk+n\alpha mk)-\hat{j}k} \\ &= \left(x^{e_1+i_0mk+(n-1)\alpha mk}\right)\left(x^{\alpha mk-\hat{j}k}\right) \\ &= \left(x^{e_1+i_0mk}\right)\left(x^{\alpha mk-\hat{j}k}\right) \\ &= x^{e_1+i_1mk-\hat{j}k} \\ &= x^{e_1+(i_1m-\hat{j})k}. \end{aligned}$$

Thus, if some i_n gives the solution, then i_0 or i_1 will also give the solution, proving the correctness of the algorithm. $\square$

Remark 1.13 Since the sort is the most expensive step, the runtime of this algorithm is certainly $O(m \ln m)$. However, it may be possible to lower this using conventional hashing techniques instead of a sort.

Explicitly, here is the algorithm for simplifying a DLP computation when an efficiently computable homomorphism is known.

Algorithm 1.14 (Local Pohlig-Hellman)
Input:

- A ring R with $N = |R|$ and an ideal $I \subset R$ with $n = |I|$.
- $f : R \longrightarrow R/I$ such that the decision problem $f(x) \stackrel{?}{=} f(y)$ is efficiently decidable.
- Elements $x, y \in R$ with $y = x^e$ for some $e > 0$.

Output: A positive integer $\hat{e}$ such that $y = x^{\hat{e}}$.

1. Use Algorithm 1.8 to compute $k \leftarrow L_{f(x)}$. Use Pollard-rho to find e_1 such that $f(y) = f(x)^{e_1}$.
2. Use Algorithm 1.12 to find $s \geq 0$ such that $y = x^{e_1+sk}$. Output $\hat{e} = e_1 + sk$ and stop.

Remark 1.15 Step 1 needs $O(t \ln t)$ operations, where $t = \sqrt{N/n}$. Step 2 needs $O(m \ln m)$ operations, where $m = \sqrt{(N - e_1)/k}$, with e_1 and k as in step 1. Thus, the total number of operations is less than $O(\sqrt{N})$, so long as $k > 1$.

To summarize, if one builds an asymmetric cipher based on the difficulty of the DLP in a ring R, there are several criteria that should be followed. The ring should have prime power order. It should also be hard to find or compute endomorphisms of R with nontrivial kernels. The most reliable way to accomplish this is to choose R as a simple ring.

If R is a simple ring the Wedderburn-Artin Theorem implies that $R \cong \mathrm{Mat}_n(\mathbb{F}_q)$ for some $n \geq 1$ and some finite field $\mathbb{F}_q$. Furthermore, given $N = |R|$, it is easy to determine such n and q. Since the reduction of Menezes and Wu [34, 33] can be adapted to $\mathrm{Mat}_n(\mathbb{F}_q)$ in a straightforward way, it should be the case that the isomorphism

$$\phi : R \longrightarrow \mathrm{Mat}_n(\mathbb{F}_q)$$

is hard to compute. Some work has been done by Hendrik Lenstra in computing isomorphisms between finite fields [27] which could potentially extend to this case. Further research is needed to determine if such an extension of Lenstra's result is possible. It would, however, be unwise to build an asymmetric cipher on the DLP in a simple ring without good reason to believe that such an extension is not possible.

The conclusion is that there are many restrictions on rings for which the DLP is possibly as hard as the DLP in already known settings independent of the integer factorization problem.

1.6 Overview of this dissertation

In this dissertation we study generalizations of existing asymmetric cryptosystems. An ambitious goal would be that this work eventually leads to a secure new system that is more efficient than existing systems. Perhaps more importantly though, such work is needed in the event that existing systems become insecure due to some new discovery. For example, the successful implementation of a quantum computer would compromise both the RSA system and systems based on the difficulty of computing discrete logarithms[44]. Many alternative systems have been proposed already (e.g., those in [31, 38, 3, 2, 23, 37]). However, many of the alternatives do suffer from drawbacks. For example, some cannot be implemented efficiently and some do not have a clear underlying problem that should be hard to solve. In addition, some systems simply have not received enough attention from the cryptographic community to be reasonably sure that they are secure. In any case, the search for more alternative systems is certainly worthwhile.

In Chapter 2 we will study the DLP in a special class of finite rings. We will observe that simple rings do maximize the difficulty of the DLP in this case. This observation will lead to an interesting primary decomposition algorithm for zero-dimensional ideals; a result of interest in its own right.

Chapter 3 considers the discrete logarithm problem in finite semirings. Since most well-known constructions of large semirings from smaller ones require commutative addition (excepting, of course, direct products and sums), we restrict our study to finite, additively commutative semirings. Since simplicity is again a consideration, we derive some structure information about such 'simple' semirings.

In the fourth chapter, we consider a natural generalization of the DLP itself as the building block of an asymmetric cipher. The generalized version considers arbitrary semigroup actions on finite sets. We will show that a variation of the Pollard-rho algorithm can be used to solve a group action problem. Finally we will present a particular semigroup action which may eventually lead to an interesting cryptosystem.

CHAPTER 2

AN INTERESTING EXAMPLE

Here we will examine the DLP in the ring $\mathbb{F}_q[\underline{x}]/I$, where I is a zero-dimensional ideal. While Section 2.1 will show that the DLP in this ring can be reduced to a known case in polynomial-time, it leads to an interesting algorithm for computing the primary-decomposition of a zero-dimensional ideal[36]; a result which is interesting in its own right. The main result is that, given a Gröbner basis for a zero-dimensional ideal, its primary decomposition may be computed without computing any additional Gröbner basis. It is also a good example of how the study of cryptography can lead to interesting results in other fields of mathematics.

There are already several efficient algorithms known to compute the primary decomposition of an ideal. We would like to mention the papers by Eisenbud, Huneke and Vasconcelos [14] and by Gianni, Trager and Zacharias [17]. The algorithms in these papers first reduce the general problem of primary decomposition to primary decomposition of zero-dimensional ideals. Work has also been done on explicitly computing the solutions of zero-dimensional ideals in 'nice' forms [25, 1], as well as for computing Gröbner bases of zero-dimensional ideals themselves [16].

After presenting our algorithm, we will see that it is closely related to a special case given by Eisenbud, Huneke and Vasconcelos for computing the primary decomposition of a radical zero-dimensional ideal [14].

The algorithm presented here has complexity that is far easier to measure than most existing ones, as it requires no intermediate Gröbner basis computations. Furthermore, this algorithm does not require that the ideal I be radical while some others do [14] nor does it rely on a normal position computation as in [6].

2.1 Motivation

Suppose one wishes to consider a cryptosystem based on the DLP in $R = \mathbb{F}_q[\underline{x}]/I$, where I is a zero-dimensional ideal with a Gröbner basis given. In the case of one variable and I a prime ideal, this is precisely the standard construction of an extension field. So in considering generalizations of existing DLP settings, it is natural to consider R^*. However, recall from Section 1.5 that simple rings are generally desirable to avoid Pohlig-Hellman type attacks. It is easy to see that R is simple iff I is prime, in which case $R \cong \mathbb{F}_{q^n}$, for some n. This already suggests that the DLP might not be maximally difficult in this setting, but we would like to see how the general considerations mentioned in Section 1.5 apply in this case.

Suppose I is not a primary ideal, so that there exists a nontrivial primary decomposition $I = Q_1 \cap \cdots \cap Q_c$ with $c > 1$. Consider the homomorphisms

$$f_i : R \longrightarrow \mathbb{F}_q[\underline{x}]/Q_i. \tag{2.1}$$

Certainly if one can efficiently compute the primary decomposition of I one may use these homomorphisms to reduce the DLP in R^* to DLPs in smaller rings. However, with minimal effort we may reduce the DLP in another way entirely. Suppose $r \in R^*$ and consider the multiplication map

$$\begin{array}{rcl} m_r : R & \longrightarrow & R \\ x & \longmapsto & rx. \end{array}$$

Considering R as an $\mathbb{F}_q$-vectorspace, m_r is a linear map. If we fix a basis for R, say the standard monomials [6], then each m_r can be identified with a matrix M_r relative to this basis. If $n = \dim_{\mathbb{F}_q} R$, then

$$\begin{array}{rcl} M : R^* & \longrightarrow & \mathrm{GL}_n(\mathbb{F}_q) \\ r & \longmapsto & M_r \end{array}$$

is a group monomorphism. One may use the result of Menezes and Wu [33] to reduce the DLP in $\mathrm{GL}_n(\mathbb{F}_q)$ to the DLP in some small extension fields of $\mathbb{F}_q$ in probabilistic polynomial-time. Thus, one may likewise reduce the DLP in R^*. The question then arises of how this reduction compares to our original idea for a reduction using the homomorphisms given in Equation 2.1. We will see that, in fact, these reductions are very closely related.

A closer inspection of the reduction of Menezes and Wu [33] is now in order. Given $A \in \mathrm{GL}_n(\mathbb{F}_q)$ one may, of course, compute the Jordan decomposition $A = UJU^{-1}$. Their main result, not at all obvious, was that this can be done in probabilistic polynomial time. To compute the discrete logarithm $\log_A B$ one may then compute it on one Jordan block at a time, which polynomial-time reduces to computing DLPs in small extension fields of $\mathbb{F}_q$ in an obvious way. Stated differently, one may polynomial-time reduce the computation of $\log_A B$ to a DLP on each invariant factor of A.

Let $S = \mathbb{F}_q[\underline{x}]$ and suppose that $I = Q_1 \cap \cdots Q_c$ is a reduced primary decomposition [20, Definition VII.2.12]. One then has that

$$\phi : R \longrightarrow S/Q_1 \times \cdots \times S/Q_c$$

is actually an isomorphism [20, Corollary III.2.26]. For $r \in R^*$, let μ_r denote the induced endomorphism of $S/Q_1 \times \cdots \times S/Q_c$ corresponding to m_r and consider the following diagram.

$$\begin{tikzcd}
R \arrow[r, "m_r"] \arrow[d, "\phi"'] & R \\
S/Q_1 \times \cdots \times S/Q_c \arrow[r, "\mu_r"'] & S/Q_1 \times \cdots \times S/Q_c \arrow[u, "\phi^{-1}"']
\end{tikzcd}$$

From this diagram, it is easy to see that each component, S/Q_i, corresponds to an invariant factor of the endomorphism m_r. This endomorphism and its relationship with the primary components of I has been studied in the case where the underlying field is algebraically closed in Cox, Little and O'Shea [11, Ch. 4.2].

For the remainder of this chapter, we turn our attention away from cryptography specifically, and examine the more general implications of the DLP reductions discussed in the previous section.

2.2 Notation and background

Let K be a perfect field that admits efficient operations and factorization of polynomials in $K[t]$. Computationally we are considering the rationals $K = \mathbb{Q}$ and the Galois field $\mathbb{F}_q$ with q elements. Let $S = K[x_1, \ldots, x_s]$ and let $I \subseteq S$ be a zero-dimensional ideal. Set $R := S/I$ and $n := \dim_K R$. For $r \in R$,

$$\begin{array}{rcl} m_r : R & \longrightarrow & R \\ x & \mapsto & rx \end{array}$$

is the vector space endomorphism induced by multiplication by r. M_r is the matrix associated with m_r with respect to the basis given by some fixed ordering of the standard monomials. $p_r(t) \in K[t]$ is the characteristic polynomial of M_r. We also call $p_r(t)$ the *characteristic polynomial of* r since it is independent of the choice of basis used to determine M_r. That is, since similar matrices have the same characteristic polynomial, any matrix representing m_r with respect to a different basis will give rise to the same $p_r(t)$. For $r \in R$ we will let $\tilde{r}$ denote a lift of r to S.

If $I \subseteq S$ is an ideal and $f \in S$ then $\langle I, f\rangle$ denotes the ideal $I + \langle f\rangle$. When we refer to the variety $\mathcal{V}(I)$ of an ideal with $I \subseteq K[x_1, \ldots, x_s]$, we are considering it as a subset of the algebraic closure

$$\mathcal{V}(I) \subseteq \overline{K}^s$$

since K is not assumed algebraically closed.

In addition, we rely heavily on standard results about the decomposition of a linear transformation. For background, the reader is referred to [20] .

Let $K(u) \supseteq K$ be a finite dimensional algebraic extension and $\alpha, \beta \in K(u)$. Recall that α and β are said to be *conjugates* if there exists a monic irreducible polynomial $f(t) \in K[t]$ such that $f(\alpha) = f(\beta) = 0$. Elements α and β are conjugates if and only if $\alpha = \sigma(\beta)$ for some $\sigma \in \mathrm{Aut}_K K(u)$.

2.3 Decomposition of m_r

Suppose $I = Q_1 \cap \cdots \cap Q_c$ is a reduced primary decomposition of I. Set $R_i := S/Q_i$ and $n_i := \dim_K R_i$ and consider

$$\begin{aligned} \delta : R &\longrightarrow R_1 \times \cdots \times R_c \\ s + I &\mapsto (s + Q_1, \ldots, s + Q_c). \end{aligned}$$

Then δ is an isomorphism. Since each R_i is an n_i-dimensional K-vector space, let $\mathcal{B}_i = \{e_{i1}, \ldots, e_{in_i}\}$ denote a basis. Then

$$\mathcal{B} = \bigcup_{i=1}^{c} \{(0_{R_1}, \ldots, 0_{R_{i-1}}, x, 0_{R_{i+1}}, \ldots, 0_{R_c}) \mid x \in \mathcal{B}_i\}$$

is a basis for $R_1 \times \cdots \times R_c$ as a K-vector space. When considering R as a K-vector space, we will always take some fixed ordering of the standard monomials as a basis. Then since δ is an isomorphism, there is a change of basis matrix $C \in \mathrm{GL}_n(K)$ to translate from the standard monomial basis to $\mathcal{B}$. That is, if M is the matrix representation of an endomorphism of R relative to the standard monomial basis, CMC^{-1} is the matrix representing the same endomorphism with respect to $\mathcal{B}$.

For $r \in R$, $m_r \in \mathrm{End}(R)$ and $m'_r := \delta m_r \delta^{-1} \in \mathrm{End}(R_1 \times \cdots \times R_c)$. Furthermore, notice that m'_r is given by:

$$\begin{aligned} m'_r : R_1 \times \cdots \times R_c &\longrightarrow R_1 \times \cdots \times R_c \\ (s + Q_1, \ldots, s + Q_c) &\mapsto (s\tilde{r} + Q_1, \ldots, s\tilde{r} + Q_c) \end{aligned}$$

where $\tilde{r}$ is a lift of r to S. In particular, for every

$$(0, \ldots, 0, s + Q_i, 0, \ldots, 0) \in R_1 \times \cdots \times R_c,$$

and $r \in R^*$, one has that

$$m'_r(0, \ldots, 0, s + Q_i, 0, \ldots, 0) = (0, \ldots, 0, s\tilde{r} + Q_i, 0, \ldots, 0)$$

whence R_i is an m'_r-invariant subspace. Thus, if $M'_{i,r}$ is the matrix of $m'_r|_{R_i}$ relative to some fixed basis, there is a basis of R relative to which m'_r has the matrix

$$M'_r = \begin{pmatrix} M'_{1,r} & & & & \\ & M'_{2,r} & & 0 & \\ & & \ddots & & \\ & 0 & & & \\ & & & & M'_{c,r} \end{pmatrix}.$$

For a proof of this last fact see, e.g. [20, Lemma VII 4.5]. Let $p_{i,r}(t) \in K[t]$ be the characteristic polynomial of $M'_{i,r}$ and $p_r(t) \in K[t]$ the characteristic polynomial of M'_r. Then

$$p_r(t) = \prod_{i=1}^{c} p_{i,r}(t).$$

Since similar matrices have the same characteristic polynomial, $p_r(t)$ is also the characteristic polynomial of M_r, hence of r as well. In particular, $p_r(t)$ has at least one irreducible factor for each primary component. This gives the following lemma.

Lemma 2.1 *Let R be as above and $r \in R$. Suppose that $p_r(t) = f_1(t)^{j_1} \cdots f_m(t)^{j_m}$ with the f_i irreducible and $j_i > 0$. Then the number of distinct primary ideals in a reduced primary decomposition of I is at most $\sum_{i=1}^{m} j_i$.*

Now, if $r_1, r_2 \in R$ with $\tilde{r}_1 + Q_i = \tilde{r}_2 + Q_i$, one has $\tilde{r}_1 - \tilde{r}_2 \in Q_i$. Whence,

$$\begin{aligned} m'_{r_1}(0, \ldots, 0, s + Q_i, 0, \ldots, 0) &= (0, \ldots, 0, s\tilde{r}_1 + Q_i, 0, \ldots, 0) \\ &= (0, \ldots, 0, s\tilde{r}_1 - s(\tilde{r}_1 - \tilde{r}_2) + Q_i, 0, \ldots, 0) \\ &= (0, \ldots, 0, s\tilde{r}_2 + Q_i, 0, \ldots, 0) \\ &= m'_{r_2}(0, \ldots, 0, s + Q_i, 0, \ldots, 0). \end{aligned}$$

So, to study $m'_r|_{R_i}$, it suffices to study the linear transformations m_r in the case where I is primary.

Proposition 2.2 *Let $r \in (S/Q)^*$ with Q primary. Then $p_r(t) = f(t)^k$ for some irreducible $f \in K[t]$ and some $k > 0$.*

Proof: Let $n = \dim_K S/Q$. Since Q is zero-dimensional, $\mathcal{V}(Q) = \{z_1, \ldots, z_j\} \subseteq \overline{K}^s$. Also, observe that evaluation of r at z_i is well-defined since $f(z_i) = 0$ for all $f \in Q$. Furthermore, if $\tilde{g} \in S$ vanishes on some z_i, it vanishes on $\mathcal{V}(Q)$ (otherwise the corresponding $g \in S/Q$ would be a non-nilpotent zero-divisor. But every zero-divisor in S/Q is necessarily nilpotent since Q is primary).

Let $\alpha = r(z_1) \in \overline{K}$, and $f(t) \in K[t]$ be the minimal polynomial of α. So f is irreducible and $f(\alpha) = 0$. Now, let $\tilde{r}$ be any lift of r to S.

Then $f(\tilde{r})(z_1) = f(\tilde{r}(z_1)) = f(\alpha) = 0$. By the observation above, $f(\tilde{r})$ vanishes on $\mathcal{V}(Q)$ and so $f(\tilde{r}) \in \mathrm{Rad}(Q)$. This implies that $f(\tilde{r})^l \in Q$ for some $l > 0$. By the canonical projection onto S/Q we have $f(r)^l = 0$.

Let $m(t)$ be the minimal polynomial of the linear transformation m_r. Then $m(t) \,|\, p_r(t)$. But since $f(r)^l = 0$ we also have $f(m_r)^l = 0$, whence $m(t) \,|\, f(t)^l$. Then $m(t) \,|\, f(t)^l$ implies $m(t) = f(t)^j$ for some $j \leq l$. An irreducible polynomial divides the characteristic polynomial of a matrix if and only if it divides the minimal polynomial, so the only irreducible divisor of $p_r(t)$ is $f(t)$. It follows that $p_r(t) = f(t)^k$, where $k = n/\deg(f)$. □

2.4 Primary decomposition

Suppose $I = Q_1 \cap \cdots \cap Q_c$ is a reduced primary decomposition as before and $R_i := S/Q_i$. Let

$$\begin{aligned} \pi_i : R &\longrightarrow R_i \\ s + I &\mapsto s + Q_i. \end{aligned}$$

Then for $r \in R^*$ we have, as before

$$p_r(t) = \prod_{i=1}^{c} p_{i,r}(t),$$

where $p_{i,r}(t)$ is the characteristic polynomial of $\pi_i(r) \in R_i$.

Proposition 2.3 *Let I, Q_i and $p_{i,r}(t)$ all as above. Then*

$$I = \langle I, p_{1,r}(\tilde{r}) \rangle \cap \cdots \cap \langle I, p_{c,r}(\tilde{r}) \rangle.$$

Proof: It is clear that $I \subseteq \langle I, p_{1,r}(\tilde{r})\rangle \cap \cdots \cap \langle I, p_{c,r}(\tilde{r})\rangle$, so we will show the other direction. Let $f \in \langle I, p_{1,r}(\tilde{r})\rangle \cap \cdots \cap \langle I, p_{c,r}(\tilde{r})\rangle$. Since $I = Q_1 \cap \cdots \cap Q_c$, we have $I \subseteq Q_i$ for $1 \leq i \leq c$. Furthermore, since $p_{i,r}(\tilde{r}) \in Q_i$, it follows that $f \in \langle I, p_{i,r}(\tilde{r})\rangle \subseteq Q_i \Rightarrow f \in Q_i$ for $1 \leq i \leq c$, whence $f \in Q_1 \cap \cdots \cap Q_c = I$. $\square$

Proposition 2.4 *Suppose* $(p_{i,r}(t), p_{j,r}(t)) = 1$ *for* $i \neq j$. *Then* $Q_i = \langle I, p_{i,r}(\tilde{r})\rangle$.

Proof: Certainly $Q_i \supseteq \langle I, p_{i,r}(\tilde{r})\rangle$, so we will show inclusion in the other direction. Without loss of generality, assume $i = 1$, and let $f \in Q_1$. There exist $h_1, h_2 \in K[t]$ such that

$$h_1(t)p_{1,r}(t) + h_2(t)p_{2,r}(t)p_{3,r}(t)\cdots p_{c,r}(t) = 1$$

whence

$$h_1(\tilde{r})p_{1,r}(\tilde{r}) + h_2(\tilde{r})p_{2,r}(\tilde{r})p_{3,r}(\tilde{r})\cdots p_{c,r}(\tilde{r}) = 1.$$

But then

$$fh_1(\tilde{r})p_{1,r}(\tilde{r}) + fh_2(\tilde{r})p_{2,r}(\tilde{r})p_{3,r}(\tilde{r})\cdots p_{c,r}(\tilde{r}) = f.$$

Since $f \in Q_1$ and $p_{j,r}(\tilde{r}) \in Q_j$, the second term in this sum is in $Q_1Q_2\cdots Q_c \subseteq I$ and the first is in $\langle p_{1,r}(\tilde{r})\rangle$, whence $f \in \langle I, p_{1,r}(\tilde{r})\rangle$. $\square$

Notice the implication of the above proposition: If we can identify the $p_{i,r}(t)$ from the factorization of $p_r(t)$ and they satisfy $(p_{i,r}(t), p_{j,r}(t)) = 1$ for $i \neq j$, we can immediately write down the primary components with no further calculation. Of course, if we wish to do some calculations in the primary ideals, it may be desirable to then compute a Gröbner basis for each $\langle I, p_{i,r}(\tilde{r})\rangle$, but this is not necessary for computing the actual primary decomposition.

2.5 The primary decomposition algorithm

We first show that the condition

$$(p_{i,r}(t), p_{j,r}(t)) = 1 \text{ for all } i \neq j$$

is satisfied when $r \in R^*$ is a generic element. This fact will allow us to derive the primary components of I by Proposition 2.4.

Lemma 2.5 *Suppose* $(p_{i,r}(t), p_{j,r}(t)) \neq 1$ *for some* $i \neq j$. *Then if* $\tilde{r}$ *is a lift of* r *to* S, $\tilde{r}(y)$ *is a conjugate of* $\tilde{r}(z)$ *for all* $y \in \mathcal{V}(Q_i), z \in \mathcal{V}(Q_j)$.

Proof: The assumption $(p_{i,r}(t), p_{j,r}(t)) \neq 1$ and Proposition 2.2 imply that $p_{i,r}(t) = f(t)^{m_i}$ and $p_{j,r}(t) = f(t)^{m_j}$ for some irreducible $f \in K[t]$. Furthermore, since $p_{i,r}(\tilde{r}) \in Q_i$, we have $f(\tilde{r}) \in \mathrm{Rad}(Q_i)$, whence $f(\tilde{r})(y) = 0$ for all $y \in \mathcal{V}(Q_i)$. It follows that $f(\tilde{r})(y) = f(\tilde{r}(y)) = 0$. Similarly, $f(\tilde{r}(z)) = 0$. Since f is irreducible, $\tilde{r}(y)$ and $\tilde{r}(z)$ are conjugates. $\square$

Lemma 2.6 *(Existence) Let c denote the number of components in the reduced primary decomposition of I. If $|K| > c$, there exists $r \in R^*$ such that*

$$(p_{i,r}(t), p_{j,r}(t)) = 1 \textit{ for all } i \neq j.$$

Proof: By the previous lemma, it suffices to show that there exists $r \in R^*$ such that $\tilde{r}(y)$ is not a conjugate of $\tilde{r}(z)$ for all $y \in \mathcal{V}(Q_i)$, $z \in \mathcal{V}(Q_j)$, $i \neq j$. For $1 \leq i \leq c$ there exists $r_i \in R$ such that $\tilde{r}_i(y) = 0$ for all $y \in \mathcal{V}(I) \setminus \mathcal{V}(Q_i)$ and $\tilde{r}_i(z) = 1$ for all $z \in \mathcal{V}(Q_i)$. By assumption there exist nonzero elements, $a_1, \ldots, a_c \in K$ that are pairwise distinct. Take $r = a_1 r_1 + \cdots a_c r_c$. Then evaluation of $\tilde{r}$ at any point in $\mathcal{V}(Q_i)$ is a_i. Furthermore, $\tilde{r}$ does not vanish on any point of $\mathcal{V}(I)$, whence $r \in R^*$. $\square$

Example 2.7 Consider $I = \langle x^2 - 2, y^2 - 2\rangle \subseteq \mathbb{Q}[x, y] = S$. In the next section we will see that the primary decomposition of I is given by $I = Q_1 \cap Q_2$, where

$$\begin{aligned} Q_1 &= (x^2 - 2, y^2 - 2, xy - 2) \\ Q_2 &= (x^2 - 2, y^2 - 2, xy + 2). \end{aligned}$$

From this we can see that

$$\begin{aligned} \mathcal{V}(Q_1) &= \{(\sqrt{2}, \sqrt{2}), (-\sqrt{2}, -\sqrt{2})\} \\ \mathcal{V}(Q_2) &= \{(-\sqrt{2}, \sqrt{2}), (\sqrt{2}, -\sqrt{2})\}. \end{aligned}$$

Set $r_1 = \frac{1}{4}(2 + xy)$ and $r_2 = \frac{1}{4}(2 - xy)$. Then r_1 vanishes on $\mathcal{V}(Q_2)$ and evaluates to 1 on $\mathcal{V}(Q_1)$. Similarly, r_2 vanishes on $\mathcal{V}(Q_1)$ and evaluates to 1 on $\mathcal{V}(Q_2)$. Thus, for any distinct nonzero $a, b \in \mathbb{Q}$, $r = ar_1 + br_2 \in S/I$ will yield a characteristic polynomial $p_r(t)$ that has exactly two coprime irreducible factors. However, these are not the only such r; in computing the primary decomposition of I in the next section, we will we choose $r = 1 + x + y$, which is not of this form. One would hope that there are many r that will work, and in fact this is the case.

Proposition 2.8 *Assume $|K| > c$ and let $I = Q_1 \cap \cdots \cap Q_c$ be a reduced primary decomposition. Then for $z_i \in \mathcal{V}(Q_i)$, $z_j \in \mathcal{V}(Q_j)$,$i \neq j$, $\tilde{r}(z_i)$ and $\tilde{r}(z_j)$ are not conjugates over K for generic $r \in R$.*

Proof: Throughout, when we say "conjugate" we mean conjugate over K. Let $n = \dim_K R$. Then there exists a bijection between elements of R and points in K^n. We wish to show that

$$\{r \in R \,|\, r(z_i), r(z_j) \text{ are conjugates for some } i \neq j\}$$

is an algebraic set. Since $\mathcal{V}(I)$ is finite, it suffices to show that for each fixed i, j with $i \neq j$, and $z_i \in Q_i$, $z_j \in Q_j$

$$\{r \in R \,|\, r(z_i), r(z_j) \text{ are conjugates }\}$$

is algebraic. (Then, the first set is a finite union of sets of this latter form). Since i, j are fixed, we will assume $i = 1, j = 2$.

Fix $z_1 \in \mathcal{V}(Q_1), z_2 \in \mathcal{V}(Q_2)$ and let $\mathbb{F} \supseteq K$ be the smallest field extension such that $z_1, z_2 \in \mathbb{F}^s$. Let $e_1, \ldots, e_n$ be the standard monomials in R. Set

$$\begin{array}{rll} C := \{\overline{a} \in K^n \,| & a_1e_1(z_1) + \cdots + a_ne_n(z_1) & \text{and} \\ & a_1e_1(z_2) + \cdots + a_ne_n(z_1) & \text{are conjugates }\}. \end{array}$$

We now must show that C is an algebraic subset of K^n. First note that, by the previous existence lemma, $C \neq K^n$. Now, let $c_i = e_i(z_1), d_i = e_i(z_2) \in \mathbb{F}$ and

$$\begin{aligned} f_1(x_1, \ldots, x_n) &= c_1x_1 + \cdots + c_nx_n \in \mathbb{F}[x_1, \ldots, x_n], \\ f_2(x_1, \ldots, x_n) &= d_1x_1 + \cdots + d_nx_n \in \mathbb{F}[x_1, \ldots, x_n]. \end{aligned}$$

Then

$$C = \{y \in K^n \,|\, f_1(y), f_2(y) \text{ are conjugates }\}.$$

Recall that $f_1(y)$ and $f_2(y)$ are conjugates if and only if $f_1(y) = \sigma(f_2(y))$ for some $\sigma \in \mathrm{Aut}_K\mathbb{F}$. But since $\mathbb{F} \supseteq K$ is a finite extension, $\mathrm{Aut}_K\mathbb{F}$ is finite and we may write

$$\mathrm{Aut}_K\mathbb{F} = \{\sigma_1, \ldots, \sigma_m\}$$

and

$$C_i = \{y \in K^n \,|\, f_1(y) = \sigma_i(f_2(y))\}.$$

Then $C = C_1 \cup \cdots \cup C_m$. But since σ_i is a field isomorphism, we have

$$\sigma_i(f_2(y)) = \sigma_i(d_1)\sigma_i(y_1) + \cdots + \sigma_i(d_n)\sigma_i(y_n).$$

Thus if we set

$$f_{2,i}(x_1, \ldots, x_n) = \sigma_i(d_1)x_1 + \cdots + \sigma_i(d_n)x_n$$

we find that

$$\sigma_i(f_2(y)) = f_{2,i}(\sigma_i(y_1), \ldots, \sigma_i(y_n)).$$

Since $\sigma_i \in \mathrm{Aut}_K \mathbb{F}$ and $y_j \in K$, we have $\sigma_i(y_j) = y_j$, whence $\sigma_i(f_2(y)) = f_{2,i}(y)$ for all $y \in K^n$. We thus need to show that $C_i = \{y \in K^n \mid f_1(y) - f_{2,i}(y) = 0\}$ is an algebraic subset of K^n. First consider

$$\tilde{C}_i = \{y \in \mathbb{F}^n \mid f_1(y) - f_{2,i}(y) = 0\}.$$

This is an algebraic subset of $\mathbb{F}^n$, and $C_i = \tilde{C}_i \cap K^n$. Furthermore

$$C = C_1 \cup \cdots \cup C_m \subset K^n$$

is a proper subset, and so $C_i \subset K^n$ is a proper subset. Hence, C_i is an algebraic subset of K^n, and so C is an algebraic subset of K^n. □

Remark 2.9 In the above proposition, we did not assume K to be infinite. The result, however, is decidedly weak in the case where K is finite.

We now have our primary decomposition for the case where K is infinite:

Algorithm 2.10 ZD Primary Decomposition:
Input: Gröbner basis for a zero-dimensional ideal $I \subset K[x_1, \ldots, x_n] = S$, with K infinite.
Output: Elements $r_1, \ldots, r_c \in R$ such that $\langle I, \tilde{r}_1 \rangle \cap \cdots \cap \langle I, \tilde{r}_c \rangle$ is a reduced primary decomposition of I.

1. Fix a basis, $\{e_1, \ldots, e_n\}$ consisting of the standard monomials of S/I.

2. Choose a random element, $r \in R$, and calculate $p_r(t)$. If $t \mid p_r(t)$, r is not invertible, so repeat until $t \nmid p_r(t)$. (The generic element is invertible, so this won't happen often).

3. Compute the factorization of $p_r(t) = f_1(t)^{d_1} \cdots f_c(t)^{d_c}$ into irreducible components with $(f_i, f_j) = 1$ for $i \neq j$.

4. Calculate $r_i = f_i(r)^{d_i}$ for $1 \leq i \leq c$, and output the $\tilde{r}_i$.

We now wish to compare this algorithm to the special case algorithm given by Eisenbud, Huneke and Vasconcelos [14]. We give here the description given by Decker, Greuel and Pfister [39]. They first perform a radical computation, and hence, assume I is radical.

Algorithm 2.11 DecompEHV(I)

1. Set $R := S/I$. Choose a generic element $f \in R$ and test whether it is a zero-divisor.
2. If f is a zero-divisor, return $\text{DecompEHV}(I : f) \cup \text{DecompEHV}(\langle I, f \rangle)$.
3. Choose m minimal such that $1, f, f^2, \ldots, f^m$ are linearly dependent and denote by $F \in K[t]$ the corresponding dependence relation.
4. If $m < \dim_K R$, restart the algorithm with another f.
5. If F is irreducible then return I.
6. If F factors as $F = G_1 \cdot G_2$, return $\text{DecompEHV}(\langle I, G_1(f) \rangle) \cup \text{DecompEHV}(\langle I, G_2(f) \rangle)$

The characteristic polynomial, $p_r(t)$, plays the same role in Algorithm 2.10 as the dependence relation, F, in the above algorithm. The only mysterious difference is step 4 of Algorithm 2.11. This condition corresponds to the case where we've chosen an r such that $p_r(z_1)$ and $p_r(z_2)$ are conjugates for some $z_1 \in Q_i$, $z_2 \in Q_j$, $i \neq j$. Although we've shown that this does not happen generically, we could incorporate this test into our algorithm, allowing us to try an $r \in R$ that gives rise to a sparse matrix M_r. This could considerably speed up the computation of the characteristic polynomial, and the added step would explicitly be:

3.5 For $i = 1 \ldots c$, if $d_i > 1$, evaluate $p_r(t)/f_i(t)^{d_i - 1}$ at r. If it evaluates to zero, goto step 2.

This added step also makes sure our output is a correct primary decomposition, even though it would be generically without this step.

However, one major difference between the two algorithms is that DecompEHV will require a Gröbner basis computation each time it is recursively called. That is, although the ideal quotient, $(I : f)$, and the sums, $\langle I, f \rangle, \langle I, G_1(f) \rangle, \langle I, G_2(f) \rangle$ may be computed without using Buchberger's algorithm, DecompEHV(I) requires a Gröbner basis for I. So each time a recursive call is made a new Gröbner basis is computed. The other major difference is that Algorithm 2.10 does not require a radical computation.

Before giving some examples, we wish to briefly mention the situation when K is finite. Suppose $K = \mathbb{F}_{p^N}$. We have shown the existence of a "good" $r \in R$ when $p^N > c$, where c is the number of primary components in a reduced primary decomposition of I. We may choose a sufficiently large prime q and find an extension field, $K' = \mathbb{F}_{p^{Nq}} \supset K$. Then there are no intermediate fields between K' and K. In particular, for sufficiently large q we may assume

1. $p^{Nq} > n$
2. $\mathcal{V}(I) \cap K^n = \mathcal{V}(I) \cap (K')^n$.

Then I generates an ideal $I' \subseteq S' = K'[x_1, \ldots, x_s]$ and this first condition gives the existence of a good $r \in (S'/I')^*$. We may find such an r efficiently by incorporating a version of step 4 from DecompEHV into our algorithm, and hence compute a primary decomposition of I' in S', say $I' = Q'_1 \cap \cdots \cap Q'_{c'}$. Furthermore, the second condition and the fact that there are no intermediate fields gives us that $c' = c$. We thus conjecture that a primary decomposition of I in S may be recovered from a decomposition of I' in S'.

2.6 Examples

Here we present the reader with some examples that can be easily verified.

Example 2.12 Consider $I = \langle x^2 - 2, y^2 - 2\rangle \subset \mathbb{Q}[x, y] = S$. The ring S/I is a 4 dimensional $\mathbb{Q}$-vector space with a basis given by the standard monomials $\{1, x, y, xy\}$. Let $r = 1 + x + y$, and we get the matrix representation of r relative to this basis

$$M_r = \begin{pmatrix} 1 & 2 & 2 & 0 \\ 1 & 1 & 0 & 2 \\ 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 \end{pmatrix}.$$

This matrix has the characteristic polynomial

$$p_r(t) = (t^2 - 2t - 7)(t - 1)^2.$$

Taking $p_{1,r}(t) = (t^2 - 2t - 7)$ and $p_{2,r}(t) = (t - 1)^2$, we get

$$I = \langle I, p_{1,r}(r)\rangle \cap \langle I, p_{2,r}(r)\rangle = \langle x^2 - 2, y^2 - 2, xy - 2\rangle \cap \langle x^2 - 2, y^2 - 2, xy + 2\rangle.$$

Example 2.13 Let $I = \langle x^2 + y + 1, 2xy + y\rangle = \langle x^2 + y + 1, 4y^2 + 5y, 2xy + y\rangle$. Let $r = 1 + x + 2y$ and we get

$$M_r = \left(\begin{array}{ccc} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 2 & -2 & -2 \end{array} \right).$$

Then M_r has characteristic polynomial

$$p_r(t) = (t + 2)(t^2 - 2t + 2),$$

which gives the primary components:

$$\begin{aligned} Q_1 &= \langle I, 3 + x + 2y\rangle = \langle 4y + 5, 2x + 1\rangle, \\ Q_2 &= \langle I, -8y\rangle = \langle x^2 + 1, y\rangle. \end{aligned}$$

Example 2.14 Consider $S = \mathbb{F}_5[x, y, z]$ and the ideal

$$I = \langle x^3 + y + 1, y^3 + z + 1, z^2 + x + 1\rangle \subseteq S.$$

Let $r = x + 1 \in R$. Then the standard monomials,

$$\{1, x, y, z, x^2, y^2, xy, xz, yz, x^2y, x^2z, xy^2, xyz, y^2z, x^2y^2, x^2yz, xy^2z, x^2y^2z\},$$

form a basis for $R := S/I$ as an $\mathbb{F}_5$-vector space. Thus, R is an 18-dimensional $\mathbb{F}_5$-vector space. Relative to this basis the matrix M_r is given by

$$M_r = \left(\begin{array}{cccccccccccccccccc}
1 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 4 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \\
0 & 0 & 1 & 0 & 4 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 4 & 0 & 0 & 0 & 0 & 4 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 4 & 0 & 4 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{array} \right).$$

This matrix has characteristic polynomial

$$p_r(t) = 1+t+t^3+2t^4+4t^5+t^6+4t^7+t^8+3t^9+t^{11}+4t^{12}+2t^{13}+3t^{16}+2t^{17}+t^{18},$$

which factors over $\mathbb{F}_5$ as

$$\begin{aligned} p_r(t) &= (t+1)(t^3+3t^2+4t+1) \\ & \quad (t^{14}+3t^{13}+4t^{12}+3t^{11}+4t^{10}+2t^9+4t^8+3t^7+t^6+t^5+t^4+3t^2+t+1). \end{aligned}$$

Denote the irreducible components of the factorization by

$$\begin{aligned} f_1(t) &= t+1, \\ f_2(t) &= t^3+3t^2+4t+1, \text{ and} \\ f_3(t) &= t^{14}+3t^{13}+4t^{12}+3t^{11}+4t^{10}+2t^9+4t^8+3t^7+t^6+t^5+t^4+3t^2+t+1). \end{aligned}$$

Since there is no chance of ambiguity (i.e., all the proper divisors of $p_r(t)$ are mutually coprime), we know that each irreducible factor corresponds to exactly one invariant factor of the endomorphism m_r. That is, although we have not shown the existence of a "good" r in this case, we can be sure that this is a "good" r. Whence, we can set

$$\begin{aligned} p_{1,r}(t) &:= f_1(t), \\ p_{2,r}(t) &:= f_2(t), \\ p_{3,r}(t) &:= f_3(t). \end{aligned}$$

We then have

$$\begin{aligned} p_{1,r}(r) &= x+2, \\ p_{2,r}(r) &= x^2+3x+4y+3, \\ p_{3,r}(r) &= 4x^2yz+2x^2y+2xy^2+3xyz+2x^2+2xy+ \\ & \quad 2xz+2y^2+yz+2x+2y+3z+2. \end{aligned}$$

This yields the primary decomposition $I = Q_1 \cap Q_2 \cap Q_3$, where

$$\begin{aligned} Q_1 &= \langle I, p_{1,r}(r) \rangle, \\ Q_2 &= \langle I, p_{2,r}(r) \rangle, \\ Q_3 &= \langle I, p_{3,r}(r) \rangle. \end{aligned}$$

Notice the complexity of these computations: We were required to compute a basis for S/I, which requires time roughly linear in n. We then computed the matrix

M_r in time roughly n^2, and the characteristic polynomial $p_r(t)$. This last task can be done using Hessenberg's algorithm, which requires time roughly n^3. The time required to factor $p_r(t)$ is a complicated issue that we do not wish to get deeply involved with except to mention that:

- Lenstra, Lenstra, and Lovász have shown that polynomials in $\mathbb{Q}[t]$ can be factored with a deterministic polynomial time algorithm [28, 26].
- If one assumes the Generalized Riemann Hypothesis (GRH), polynomials in $\mathbb{F}_p[t]$ can be factored with a deterministic polynomial time algorithm [9]. Not assuming GRH, factorization of such polynomials still seems to be very efficient in practice.

2.7 Conclusion

In this chapter, the close examination of a poor cryptosystem lead to an algorithm for computing the primary decomposition of a zero-dimensional ideal over an infinite, computable field admitting efficient univariate polynomial factorization. The algorithm is not deterministic, but since we have shown that it requires only one pass in the generic case, it has expected runtime $O(n^3 + F_K(\chi(t)))$, where $n = \dim_K S/I$, and $F_K(\chi(t))$ is the time required to factor the characteristic polynomial $\chi(t)$ of degree n over K. We also exhibited an example where it works with K finite, and have given some indications how it might be adopted to the finite case in general.

The algorithm given here has been implemented by the University of Kaiserslautern in the latest releases of the computer algebra package *Singular* [46]. The function is named "zdprimdec" and is available in the primary decomposition library distributed with the program.

CHAPTER 3

THE DLP IN SEMIRINGS

While the study of semirings is generally acknowledged to have started in around 1934 by Vandiver[47], nomenclature is only recently becoming standardized in the area. In reading the literature, one should be careful to observe the definition of "semiring" that is being used. For example, in [35] many results are given regarding congruence-simple commutative semirings. However, in that paper, semirings were assumed to have a multiplicative identity; an assumption which is generally no longer used. In computer science textbooks on the subject, for example [24], semirings are often assumed to have both a zero and a one.

Currently, the most accepted definition of a semiring requires only that $(S, +)$ and $(S, \cdot)$ be semigroups with left and right distributivity of $\cdot$ over $+$. With this modern definition there are surprisingly few strong results concerning even special classes of semirings. Indeed, it was only in 2001 that a classification of finitely generated, congruence-simple, commutative semirings was given in [5]. Some general results concerning 0-simple semirings are given in [49], but the notion of 0-simple does not correspond to the more general notion of c-simple used in this dissertation.

Recall that in this dissertation we are generally interested in finding new and efficient asymmetric ciphers. Toward that end, we begin the study of finite semirings (which properly contain the class of finite rings) as objects for which the DLP may serve as the building block of an asymmetric cipher. The motivation for this is the following:

- As per the considerations in Section 1.5, there are many restrictions on rings which may have maximally difficult DLPs (i.e., DLPs for which the best known time of attack is $O(2^{N/2})$, where N is the size of the representation of an element).

- There are easy ways to non-trivially construct larger rings from smaller ones (i.e., other than direct products/sums). For example, matrix rings and quotients of polynomial rings are easy to construct and have efficient arithmetic. The same is true in additively-commutative semirings.

In a group or ring, the easiest way to avoid Pohlig-Hellman-type attacks is to insist on the use of a simple group or ring. In semirings, the corresponding notion of simple that will accomplish the same is congruence-simple. It is thus natural to begin this study with an examination of such semirings.

In this chapter, we give some results concerning finite, additively commutative, congruence-simple semirings. While there is cryptographic motivation to study such objects, there is also intrinsic mathematical value to their study; additively commutative semirings arise naturally as the endomorphisms of commutative semigroups. Furthermore, every such semiring is isomorphic to a sub-semiring of such endomorphisms [19].

3.1 Introduction to semirings

For a general introduction to semirings and a large collection of references, the reader is referred to [24, 19].

Definition 3.1 A *semiring* is a nonempty set S together with two associative operations, $+$ and $\cdot$, such that for all $a, b, c \in S$

$$a \cdot (b + c) = a \cdot b + a \cdot c$$
$$(a + b) \cdot c = a \cdot c + b \cdot c.$$

A semiring is called *additively [multiplicatively] commutative* if $(S, +)$ $[(S, \cdot)]$ is commutative. If both $(S, +)$ and $(S, \cdot)$ are commutative, S is simply called *commutative.*

For the purpose of obtaining larger semirings from smaller ones via polynomial and matrix constructions, it is necessary that S be additively commutative. Thus, throughout the rest of this chapter all semirings are assumed to be additively commutative. One can then easily generate small semirings by computer search and use such constructions to obtain larger ones. Simplicity is still a consideration, however, so we will turn our attention to studying simplicity after some more definitions.

Definition 3.2 An element α of a semiring is called *additively [multiplicatively] absorbing* if $\alpha + x = x + \alpha = \alpha$ $[\alpha \cdot x = x \cdot \alpha = \alpha]$ for all $x \in S$. An element ∞ of a semiring is called an *infinity* if it is both additively and multiplicatively absorbing.

Note that an additive identity in a semiring need not be multiplicatively absorbing. If, however, a semiring has a multiplicatively absorbing additive identity, we call it a *zero*, and denote it by 0.

Definition 3.3 A semiring S with additive identity o is called *zero-sum free* if for all $a, b \in S$, $a + b = o$ implies $a = b = o$.

Definition 3.4 An element a of a semiring S is called *additively [multiplicatively] left cancellative* if for all $b, c \in S$

$$a + b = a + c \Rightarrow b = c \qquad [ab = ac \Rightarrow b = c]\,.$$

If an element is additively [multiplicatively] left and right cancellative, it is said to be *additively [multiplicatively] cancellative.* If every element of a semiring S is additively [left] cancellative, S is called additively [left] cancellative. If every element, except possibly an additive identity, of a semiring S is multiplicatively [left] cancellative, S is said to be *multiplicatively [left] cancellative.*

Definition 3.5 Let S and R be semirings, and $f : R \longrightarrow S$ a function. Then f is called a *semiring homomorphism* if for all $x, y \in S$

$$f(x + y) = f(x) + f(y) \qquad \text{and} \qquad f(xy) = f(x)f(y).$$

A bijective semiring homomorphism is called a *semiring isomorphism.*

When the context is clear, we may omit the qualifier 'semiring' from the previous definition.

Definition 3.6 Let S be a semiring and $I \subseteq S$ a subset. Then I is called a *bi-ideal* of S if for all $i \in I$ and $s \in S$,

$$i + s, s + i \in I \qquad \text{and} \qquad is, si \in I.$$

Definition 3.7 A *congruence relation* on a semiring S is an equivalence relation, $\sim$, that also satisfies

$$x_1 \sim x_2 \Rightarrow \begin{cases} c + x_1 & \sim & c + x_2, \\ x_1 + c & \sim & x_2 + c, \\ cx_1 & \sim & cx_2, \\ x_1 c & \sim & x_2 c, \end{cases}$$

for all $x_1, x_2, c \in S$. A semiring S that admits no congruence relations other than the trivial ones, id_S and $S \times S$, is said to be *congruence-simple*, or *c-simple.*

Note that the trivial semiring of order 1 and every semiring of order 2 are congruence-simple. Also note that if $I \subseteq S$ is a bi-ideal then $\mathrm{id}_S \cup (I \times I)$ is a congruence relation. Thus, if $I \subset S$ is a bi-ideal and S is c-simple, then $|I| = 1$ or $I = S$. Congruence relations are important because of the following lemma.

Lemma 3.8 *Let S be a semiring.*

1. *If $\sim$ is a congruence relation on S, then $S/\sim$ is a semiring, with the induced operations, and $\pi : S \longrightarrow S/\sim$ is a semiring homomorphism.*
2. *If $f : S \longrightarrow T$ is a semiring homomorphism, then*

$$x \sim y \quad \text{if} \quad f(x) = f(y)$$

defines a congruence relation on S.

Proof: We need to show that the operations

$$\begin{aligned} \bar{a} + \bar{b} &= \overline{a+b}, \\ \bar{a} \cdot \bar{b} &= \overline{a \cdot b} \end{aligned}$$

are well defined. For this, suppose $a_1 \sim a_2$ and $b_1 \sim b_2$. Then

$$\begin{aligned} a_1 \sim a_2 &\Rightarrow a_1 + b_2 \sim a_2 + b_2, \\ b_1 \sim b_2 &\Rightarrow a_1 + b_1 \sim a_1 + b_2. \end{aligned}$$

It thus follows from transitivity of $\sim$ that $a_1 + b_1 \sim a_2 + b_2$, so $+$ is well defined on equivalence classes. A similar argument shows that $\cdot$ is well defined. Associativity of both operations and distributivity then follow from associativity and distributivity in S.

The statements about homomorphisms are then obvious. □

Because of Lemma 3.8, it would be natural to call a semiring simple if it were c-simple, but this nomenclature has not been adopted in the literature except when the notions of congruence-simple and ideal-simple coincide (i.e., in general, semirings may admit congruence relations that do not arise from ideals).

The following theorem, due to Bashir, Hurt, Jančařék, and Kepka [5, Theorem 14.1], classifies finite c-simple commutative semirings.

Theorem 3.9 *Let S be a commutative, congruence-simple, finite semiring. Then one of the following holds:*

1. *S is isomorphic to one of the five semirings $T_1, \ldots, T_5$ of order 2 defined in Table 3.1.*

2. *S is a finite field.*

3. *S is a zero-multiplication ring of prime order.*

4. *S is isomorphic to $V(G)$ (defined below), for some finite abelian group G.*

For a multiplicative abelian group G, set $V(G) = G \cup \{\infty\}$. Extend the multiplication of G to $V(G)$ by the rule $x\infty = \infty x = \infty$ for all $x \in V(G)$. Define an addition on $V(G)$ by the rules $x + x = x$, $x + y = \infty$ for all $x, y \in V(G)$ with $x \neq y$.

Table 3.1. ALL COMMUTATIVE SEMIRINGS OF ORDER 2

$$\begin{array}{c|cc} (T_1,+) & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 0 \end{array} \qquad \begin{array}{c|cc} (T_1,\cdot) & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 0 \end{array} \qquad \begin{array}{c|cc} (T_2,+) & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 0 \end{array} \qquad \begin{array}{c|cc} (T_2,\cdot) & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}$$

$$\begin{array}{c|cc} (T_3,+) & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array} \qquad \begin{array}{c|cc} (T_3,\cdot) & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 0 \end{array} \qquad \begin{array}{c|cc} (T_4,+) & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array} \qquad \begin{array}{c|cc} (T_4,\cdot) & 0 & 1 \\ \hline 0 & 1 & 1 \\ 1 & 1 & 1 \end{array}$$

$$\begin{array}{c|cc} (T_5,+) & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array} \qquad \begin{array}{c|cc} (T_5,\cdot) & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 1 \end{array} \qquad \begin{array}{c|cc} (T_6,+) & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array} \qquad \begin{array}{c|cc} (T_6,\cdot) & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}$$

$$\begin{array}{c|cc} (T_7,+) & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array} \qquad \begin{array}{c|cc} (T_7,\cdot) & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 0 \end{array} \qquad \begin{array}{c|cc} (T_8,+) & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array} \qquad \begin{array}{c|cc} (T_8,\cdot) & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}$$

Theorem 3.9 has strong implications for the discrete logarithm problem in commutative, congruence-simple, finite semirings. The DLP in a semiring of order 2 is trivial, as is the DLP in a zero-multiplication ring. The DLP is already well known in finite fields. In $V(G)$, the DLP is simply living in a group, which is also well known. Since none of these semirings lead to a new DLP setting which is hard, so we must further restrict our attention to multiplicatively non-commutative finite semirings. We first note that a complete classification up to isomorphism of such finite c-simple semirings is not possible. To see this, note that $V(G)$ is a finite simple semiring for any finite group G. Furthermore, if G_1 and G_2 are two non-isomorphic groups, then $V(G_1)$ and $V(G_2)$ are non-isomorphic semirings. Thus a classification of finite, additively commutative, c-simple semirings up to isomorphism would require a classification of finite groups up to isomorphism.

3.2 Basic results

The goal of this section is to derive some basic structure information for finite, additively commutative, c-simple semirings.

Lemma 3.10 *Let S be a finite, additively commutative, c-simple semiring. If the multiplication table of S has two identical rows [columns], then one of the following holds.*

1. *There exists $c \in S$ such that $xy = c$ for all $x, y \in S$.*
2. *$|S| = 2$.*

Proof: Observe that the relation $\sim$ defined by

$$x \sim y \quad \text{if} \quad xz = yz \text{ for all } z \in S$$

is a congruence relation. By assumption, there exist $r_1 \neq r_2$ such that $r_1 z = r_2 z$ for all $z \in S$ so $\sim = S \times S$. Thus

$$xz = yz \qquad \text{for all } x, y, z \in S. \tag{3.1}$$

Suppose that $(S, \cdot)$ is not left-cancellative. Then there exist $a, b, c, d \in S$ such that $da = db = c$ and $a \neq b$. But $xa = ya, xb = yb$ for all $x, y \in S$. Hence $da = ya, db = yb$ and so $ya = yb = c$ for all $y \in S$. Consider now the congruence relation $\approx$ defined by

$$x \approx y \quad \text{if} \quad zx = zy \text{ for all } z \in S.$$

Since $a \neq b$ and $a \approx b$, it follows that $\approx = S \times S$, whence $zx = zy$ for all $x, y, z \in S$. Then for all $x, y \in S$ we have $xy = xa = da = c$.

Suppose now that $(S, \cdot)$ is left-cancellative. Fix $x \in S$ and let $z = x^2$. Then $xz = zx$. But $\gamma z = xz$ and $\gamma x = zx$ for all $\gamma \in S$, so $\gamma z = \gamma x$. By left-cancellation,

$x^2 = z = x$, so S is multiplicatively idempotent. Furthermore, for all $w \in S$ $w + w = w^2 + w^2 = (w + w)w = w^2 = w$, so S is additively idempotent. We will now show, by contradiction, that $|S| \leq 2$.

Suppose $|S| = n > 2$. For each nonempty subset $A \subseteq S$ let

$$\sigma_A = \sum_{x \in A} x$$

and $\sigma = \sigma_S$. Suppose that $A \subset S$ with $|A| = n - 1$. Consider the relation $\sim = \mathrm{id}_S \cup \{(\sigma_A, \sigma), (\sigma, \sigma_A)\}$. Clearly $\sim$ is an equivalence relation. Since $(S, \cdot)$ is idempotent, Equation 3.1 implies that for each $c \in S$

$$c\sigma_A = \sigma_A\sigma_A = \sigma_A \qquad \text{and} \qquad c\sigma = \sigma\sigma = \sigma.$$

Thus, $c\sigma_A \sim c\sigma$. Similarly,

$$\sigma_A c = c^2 = c \qquad \text{and} \qquad \sigma c = c^2 = c$$

so that $\sigma_A c \sim \sigma c$. Since $(S, +)$ is idempotent, $\sigma + c = \sigma$ and

$$\sigma_A + c = \begin{cases} \sigma_A, & \text{if } c \in A, \\ \sigma, & \text{otherwise} . \end{cases}$$

Thus $\sim$ is a congruence relation. Since $|S| > 2$, it must be the case that $\sim = \mathrm{id}_S$, so $\sigma_A = \sigma$ for all proper $A \subset S$ with $|A| = n - 1$.

By induction, we will now show that $\sigma_A = \sigma$ for any nonempty subset $A \subseteq S$. Suppose this is known to hold for all A with $|A| = k \geq 2$. Let $A \subset S$ with $|A| = k - 1$ and again consider the relation

$$\sim = \mathrm{id}_S \cup \{(\sigma_A, \sigma), (\sigma, \sigma_A)\}.$$

As above, $\sim$ is a multiplicative equivalence relation. Furthermore

$$\sigma_A + c = \begin{cases} \sigma_A, & \text{if } c \in A, \\ \sigma_{A \cup \{c\}}, & \text{otherwise} . \end{cases}$$

But $c \notin A$ implies $|A \cup \{c\}| = k$, so $\sigma_{A \cup \{c\}} = \sigma$ by the inductive assumption. Thus $\sim$ is again a congruence relation. Since $\sim \neq S \times S$, it follows that $\sim = \mathrm{id}_S$, so $\sigma_A = \sigma$.

In particular, this shows that for each $w \in S$,

$$w = \sigma_{\{w\}} = \sigma,$$

a contradiction. Thus $|S| = 2$.

It only remains to see that the same statement holds if "rows" is replaced by "columns". If S has two identical columns, consider the reciprocal semiring $(S', +, \otimes)$ defined by $(S', +) = (S, +)$ and $x \otimes y = yx$. This semiring is c-simple and has two identical rows so the above argument applies. $\square$

Lemma 3.11 *Let S be an finite, additively commutative, c-simple semiring. Then exactly one of the following holds.*

- *$(S,+)$ is cancellative.*
- *There exists $\alpha \in S$ such that $x + \alpha = \alpha$ for all $x \in S$.*

Proof: Consider the relation $\sim$ defined by

$$x \sim y \quad \text{if} \quad x + t = y + t \text{ for some } t \in S.$$

It is easy to see that $\sim$ is a congruence relation. If $\sim = \mathrm{id}_S$, then $(S,+)$ is cancellative. On the other hand, suppose $\sim = S \times S$. Then for all $x, y \in S$ there exists $t_{x,y} \in S$ such that $x + t_{x,y} = y + t_{x,y}$. Set

$$\sigma = \sum_{x \in S} x \qquad \text{and} \qquad \alpha = \sigma + \sigma.$$

For $x, y \in S$ there exists $\sigma' \in S$ such that $\sigma = t_{x,y} + \sigma'$. Then

$$x + \sigma = x + t_{x,y} + \sigma' = y + t_{x,y} + \sigma' = y + \sigma.$$

In particular, $x + \sigma = \sigma + \sigma$ for all $x \in S$. Thus, for all $x \in S$

$$x + \alpha = x + \sigma + \sigma = (\sigma + \sigma) + \sigma = \sigma + \sigma = \alpha.$$

$\square$

Theorem 3.12 *Let S be a finite, additively commutative, c-simple semiring. Then one of the following holds.*

- *$(S,+)$ is a group with identity $0 \in \mathrm{Center}(S)$.*
- *There exists $\infty \in \mathrm{Center}(S)$.*
- *$(S,+)$ is idempotent.*

Proof: Suppose $(S,+)$ is cancellative. Let $S = \{c_1, \ldots, c_n\}$ and $x \in S$. Then $x + c_1, \ldots, x + c_n$ are all distinct. In particular, $x + c_i = x$ for some $c_i \in S$. Then for all $y \in S$ it follows that

$$y + x + c_i = y + x.$$

But since $(S,+)$ is cancellative, this implies $y+c_i = y$ for all $y \in S$. Thus, c_i is actually an additive identity, o. Again let $x \in S$ and notice that $xo = x(o+o) = xo + xo$. Thus, for all $y \in S$ it follows that

$$y + xo + xo = y + xo,$$

and cancellation implies $y + xo = y$ for all $y \in S$, whence $xo = o$. Similarly, $ox = o$. Since x was chosen arbitrarily, $xo = ox = o$ for all $x \in S$ so o is actually a zero, $0 \in \mathrm{Center}(S)$. To see that $(S,+)$ is actually a group, observe that for $x \in S$, $0 \in \{x + s \mid s \in S\}$ so that every element is invertible.

Suppose now that $(S,+)$ is not cancellative. By Lemma 3.11 there exists $\alpha \in S$ such that $x + \alpha = \alpha$ for all $x \in S$. Consider the relation T defined by

$$xTy \quad \text{if} \quad 2x = 2y.$$

Then T is a congruence relation, whence $T = \mathrm{id}_S$ or $T = S \times S$.

Case I: Suppose $T = S \times S$.

Then for all $x \in S$, $x + x = \alpha + \alpha = \alpha$. Thus, $x\alpha = x(\alpha + \alpha) = x\alpha + x\alpha = \alpha$. Similarly, $\alpha x = \alpha$ so $\alpha \in \mathrm{Center}(S)$ and $\alpha = \infty$.

Case II: Suppose $T = \mathrm{id}_S$.

Consider the congruence relation $\sim$ defined by $x \sim y$ if there exist $u, v \in S \cup \{o\}$ and $i \geq 0$ such that

$$\begin{aligned} 2^i x &= y + u, \\ 2^i y &= x + v. \end{aligned}$$

Then $2(2x) = (x) + 3x$ and $2(x) = (2x) + o$, so $x \sim 2x$ for all $x \in S$. If $\sim = \mathrm{id}_S$, then $x = 2x$ for all $x \in S$, whence $(S,+)$ is idempotent. Suppose now that $\sim = S \times S$ and let $x \in S$. Then $x\alpha \sim \alpha$, so there exists $v \in S \cup \{o\}$ and $i \geq 0$ such that $2^i x\alpha = \alpha + v = \alpha$. Then

$$x\alpha = x(2^i \alpha) = 2^i x\alpha = \alpha,$$

so $x\alpha = \alpha$. Similarly, $\alpha x = \alpha$ so $\alpha \in \mathrm{Center}(S)$ and $\alpha = \infty$. $\square$

3.3 The zero case

The goal of this section is to describe finite, additively commutative, c-simple semirings with zero.

Theorem 3.13 *If S is a finite c-simple semiring with zero then one of the following holds.*

- *$S \cong \mathrm{Mat}_n(\mathbb{F}_q)$ for some $n \geq 1$ and some finite field $\mathbb{F}_q$.*
- *S is a zero-multiplication ring ($S^2 = \{0\}$) of prime order.*
- *S is additively idempotent.*

Proof: Consider the set

$$A = \{x \in S \,|\, x + y = 0 \text{ for some } y \in S\}$$

and the relation $\sim$ defined by

$$x \sim y \quad \text{ if } \quad x + a_1 = y + a_2 \text{ for some } a_1, a_2 \in A.$$

It is not hard to see that $\sim$ is a congruence relation.
Case I: $\sim = S \times S$.

Then for each $x, y \in S$, there exist $a_x, a_y \in A$ such that $x + a_x = y + a_y$. But there also exists $(-a_y) \in A$, so that $a_y + (-a_y) = 0$ and

$$y = x + a_x + (-a_y).$$

Since A is closed under addition, it follows that for all $x, y \in S$ there exists $a_{x,y} \in A$ such that

$$y = x + a_{x,y}.$$

But since S is finite, it then follows that $A = S$. So all elements of S have additive inverses, whence $(S, +)$ is a group.

Since $(S, +)$ is a group and S is finite and c-simple, S is then a finite simple ring. Thus, by the Wedderburn-Artin Theorem, either $S^2 = \{0\}$ or there exists a positive integer n and a finite field $\mathbb{F}_q$ such that $S \cong \mathrm{Mat}_n(\mathbb{F}_q)$. Furthermore, if $S^2 = \{0\}$, it must be the case that $(S, +)$ is actually a simple abelian group, whence $|S| = |(S, +)| = p$ for some prime p.
Case II: $\sim = \mathrm{id}_S$.

Suppose there exists $a_1, a_2 \in A$ with $a_1 \neq a_2$. Then there also exist $(-a_1), (-a_2) \in A$ such that $a_1 + (-a_1) = 0$ and $a_2 + (-a_2) = 0$. But then

$$a_1 + (-a_1) = a_2 + (-a_2).$$

So $a_1 \sim a_2$, contradicting the assumption that $\sim = \mathrm{id}_S$. Thus, $A = \{0\}$ and S is zero-sum free. The remainder of the proof is adopted from the proof of Theorem 3.1 in [5].

Consider the congruence relation T defined by

$$x\,T\,y \quad \text{if} \quad 2x = 2y.$$

If $T = S \times S$, then for all $x \in S$, $x + x = 0 + 0 = 0$. But S is zero-sum free, so it must be the case that $T = \mathrm{id}_S$. Consider now the relation $\approx$ defined by

$$x \approx y \qquad \text{if there exist } u, v \in S \cup \{o\} \text{ and } i \geq 0 \text{ such that } \begin{cases} 2^i x &= y + u, \\ 2^i y &= x + v. \end{cases}$$

Then $\approx$ is a congruence relation on S. Note that for all $x \in S$ one has

$$\begin{aligned} 2(x) &= 2x + o, \\ 2(2x) &= x + 3x. \end{aligned}$$

So $x \approx 2x$ for all $x \in S$. If $\approx = \mathrm{id}_S$, it follows that S is additively idempotent. Suppose that $\approx = S \times S$. Also suppose that there exist $a, b, c \in S$ such that $a + b = a + c$. Then $a \approx b$, so there exists $i \geq 0$ and $w \in S \cup \{o\}$ such that

$$2^i b = a + w.$$

We then have that

$$b + 2^i b = b + a + w = c + a + w = c + 2^i b.$$

Claim: $2b = b + c$.

We will show this claim by induction on i. Certainly, if $i = 0$ then $2b = b + c$. But if $i \geq 1$,

$$2(b + 2^{i-1}b) = 2b + 2^i b = b + c + 2^i b = c + b + 2^i b = c + c + 2^i b = 2(c + 2^{i-1}b).$$

Since $T = \mathrm{id}_S$, it follows that $b + 2^{i-1}b = c + 2^{i-1}b$. So by induction, $2b = b + c$. Similarly, from $a \approx c$ it follows that $2c = b + c$. Then $2b = 2c$, whence $b = c$. Thus, S is additively cancellative. For $x \in S$, consider the set $B_x = \{x + s \mid s \in S\}$. Since S is additively cancellative,

$$x + s_1 = x + s_2 \quad \text{iff} \quad s_1 = s_2,$$

so $|B_x| = |S|$. Thus, $B_x = S$. In particular, $0 \in B_x$ implies that there exists $(-x) \in S$ such that $x + (-x) = 0$. So $(S, +)$ is actually a group, and S is a simple ring as in Case I. $\square$

3.4 The ∞ case

In this section, we show that a finite, additively commutative, c-simple semiring with ∞ is either additively idempotent or has order 2.

Lemma 3.14 *Let S be a finite, additively commutative, c-simple semiring with ∞ and $|S| > 2$. Then S is additively idempotent or $S + S = \{\infty\}$.*

Proof: Consider the congruence relation defined by

$$xTy \quad \text{if} \quad 2x = 2y.$$

Case I: $T = \mathrm{id}_S$.

Then $2x = 2y$ iff $x = y$. Set $x \sim y$ if there exists $i \geq 0$ and $u, v \in S \cup \{o\}$ such that

$$\begin{aligned} 2^i x &= y + u, \\ 2^i y &= x + v. \end{aligned}$$

Then $\sim$ is a congruence relation and $x \sim 2x$ for all $x \in S$. But $x \not\sim \infty$ for $x \neq \infty$, so $\sim \neq S \times S$. Thus, $\sim = \mathrm{id}_S$, and so S is additively idempotent.

Case II: $T = S \times S$.

Then $x + x = \infty$ for all $x \in S$. For $\emptyset \neq A \subseteq S$, let

$$\sigma_A = \sum_{x \in A} x.$$

Let $N = |S|$ and suppose that $|A| = N - 1$. Then for every $c \in S$, $\sigma_A + c = \infty$, since $c \in A$, $c = \infty$, or $\sigma_A = \infty$. Furthermore,

$$c\sigma_A = \sum_{x \in A} cx = \begin{cases} \infty, & \text{if } cx_1 = cx_2 \text{ for some distinct } x_1, x_2 \in A, \\ \sigma_A, & \text{otherwise.} \end{cases}$$

Similarly, $\sigma_A c = \infty$ or $\sigma_A c = \sigma_A$. Thus, $\mathcal{B} = \{\sigma_A \mid A \subset S \text{ with } |A| = N - 1\}$ is a bi-ideal. Furthermore, $\infty \in A$ implies $\sigma_A = \infty$. Thus, $|\mathcal{B}| \leq 2$ and so $\mathcal{B} = S \Rightarrow |S| = 2$, a contradiction. Thus $\mathcal{B} = \{\infty\}$, so $\sigma_A = \infty$ for all $A \subset S$ with $|A| = N - 1$.

By induction, we will show that $\sigma_A = \infty$ for all $A \subset S$ with $|A| = 2$. Assume $\sigma_A = \infty$ for all $A \subset S$ with $|A| = k + 1 > 2$.

Suppose now that $A \subset S$ with $|A| = k \geq 2$. Then for $c \in S$,

$$\sigma_A + c = \begin{cases} \infty, & \text{if } c \in A, \\ \sigma_{A \cup \{c\}}, & \text{otherwise .} \end{cases}$$

By assumption, if $c \notin A$ then $\sigma_{A\cup\{c\}} = \infty$, so $\sigma_A + c = \infty$ for all $c \in S$. Also

$$c\sigma_A = \sum_{x\in A} cx = \begin{cases} \infty, & \text{if } cx_1 = cx_2 \text{ for some distinct } x_1, x_2 \in A, \\ \sigma_B, & \text{for some } |B| = k \text{ otherwise.} \end{cases}$$

The same is easily seen to hold for $\sigma_A c$. Observe that $\sigma_X = \infty$ for some $X \subset S$ with $|X| = k$, so

$$\mathcal{B} = \{\sigma_A \mid A \subset S \text{ with } |A| = k\}$$

is a bi-ideal of S.

Case (i): $\mathcal{B} = \{\infty\}$.

Then $\sigma_A = \infty$ for all $A \subset S$ with $|A| = k$, so we may apply the induction and conclude that $\sigma_A = \infty$ for all $A \subset S$ with $|A| = 2$. Thus, $x + y = \infty$ for all $x, y \in S$.

Case (ii): $\mathcal{B} = S$.

We will show directly that $x + y = \infty$ for all $x, y \in S$. By assumption this holds for $x = y$, so suppose $x \neq y$. Then there exist $A_1, A_2 \subset S$ with $|A_1| = |A_2| = k$ and $\sigma_{A_1} = x$, $\sigma_{A_2} = y$.

$$A_1 \cap A_2 \neq \emptyset \quad \Rightarrow \quad x + y = \sigma_{A_1} + \sigma_{A_2} = \infty.$$
$$A_1 \cap A_2 = \emptyset \quad \Rightarrow \quad x + y = \sigma_{A_1} + \sigma_{A_2} = \sigma_{A_1 \cup A_2}.$$

But $|A_1 \cup A_2| > k$. In particular, either $|A_1 \cup A_2| = k+1$ or there exist $\emptyset \neq B_1, B_2 \subset S$ with $|B_1| = k + 1$, $B_1 \cap B_2 = \emptyset$ and $B_1 \cup B_2 = A_1 \cup A_2$. By assumption, $\sigma_{B_1} = \infty$ and we have

$$x + y = \sigma_{A_1\cup A_2} = \sigma_{B_1\cup B_2} = \sigma_{B_1} + \sigma_{B_2} == \infty + \sigma_{B_2} = \infty.$$

Thus $x + y = \infty$ for all $x, y \in S$. □

Lemma 3.15 *Let S be a finite, additively commutative, c-simple semiring with ∞ and $S + S = \{\infty\}$. Then $|S| = 2$.*

Proof: For $x \in S$

$$\mathcal{B}_x = \{uxv \mid u, v \in S\}$$

is a bi-ideal of S.

Case I: $\mathcal{B}_x = \{\infty\}$ for all $x \in S$.

Suppose there exist $x, y \in S$ with $z = xy \neq \infty$. Then for all $u \in S$, we have that $zu = xyu \in \mathcal{B}_y$, so $zu = \infty$. In particular, $zu = \infty u$ for all $u \in S$. By lemma 3.10, either $|S| = 2$ or $xy = \infty$ for all $x, y \in S$. But S c-simple with

$S + S = SS = \{\infty\} \Rightarrow |S| = 2$.

Case II: $\mathcal{B}_x \neq \{\infty\}$ for some $x \in S$.

Then $\mathcal{B}_x = S$. So if $S = \{c_1, \ldots, c_N\}$, then $c_1 x, \ldots, c_N x$ are necessarily distinct as are $xc_1, \ldots, xc_N$. Thus, there exists $l \in S$ such that $lx = x$. Also, for each $z \in S$ there exists $u \in S$ such that $z = xu$. Thus,

$$lz = lxu = xu = z$$

and so $lz = z$ for all $z \in S$. Similarly, there exists $r \in S$ so that $zr = z$ for all $z \in S$. We then have for all $z \in S$

$$zl = (zl)r = z(lr) = zr = z,$$

whence $l = 1$ is a multiplicative identity. So for all $z \in S$ it follows that $z = 1z1 \in \mathcal{B}_z$. But also $\infty \in \mathcal{B}_z$, so $\mathcal{B}_z = S$ for all $z \in S \setminus \{\infty\}$. It follows easily that $(S \setminus \{\infty\}, \cdot)$ is actually a group. Let $G = S \setminus \{\infty\}$ and consider $\sim = \mathrm{id}_S \cup (G \times G)$. Then $\sim$ is a congruence relation on S. But $\sim = S \times S$ implies $(z, \infty) \in \mathrm{id}_S \cup (G \times G)$ for all $z \in S$, a contradiction. So it must be the case that $\sim = \mathrm{id}_S$, whence $|G| = 1$ and so $|S| = 2$. $\square$

We conclude this section with the following theorem.

Theorem 3.16 *If S is a finite, additively commutative, c-simple semiring with ∞ and $|S| > 2$, then S is additively idempotent.*

Proof: Apply Lemmas 3.14 and 3.15. $\square$

3.5 Some idempotent results

In this section, we derive some results that apply to remaining case where S is additively idempotent. We shall assume throughout the section that S is a finite, additively commutative, c-simple semiring with $|S| > 2$. First we will show that if $\mathrm{Center}(S) \neq \emptyset$, then S has a zero, a one, or an infinity.

For each $c \in \mathrm{Center}(S)$ the relation R_c defined by

$$x \, R_c \, y \quad \text{if} \quad xc = yc$$

is a congruence relation. Since S is c-simple by assumption, for each $c \in \mathrm{Center}(S)$ either $R_c = \mathrm{id}_S$ or $R_c = S \times S$.

Lemma 3.17 *If there exists* $c \in \text{Center}(S)$ *with* $R_c = \text{id}_S$*, then there exists* $k \geq 1$ *such that* c^k *is a multiplicative identity.*

Proof: Let $c \in Center(S)$ with $R_c = \text{id}_S$. Consider the set $\{c, c^2, c^3, \ldots\} \subseteq S$. Since S is finite, there exist integers i, j with $0 < i < i + j$ such that

$$c^i = c^{i+j}.$$

But $R_c = \text{id}_S$ implies $c = c^{1+j}$. Let k be the least positive integer such that $c = c^{1+k}$. Then for all $x \in S$

$$(xc^k)c = xc^{k+1} = xc$$

and $R_c = \text{id}_S \Rightarrow xc^k = x$, whence c^k is a multiplicative identity. □

Lemma 3.18 *If there exist* $d_1, d_2 \in \text{Center}(S)$ *with* $R_{d_1} = R_{d_2} = S \times S$*, then* $xd_1 = yd_2$ *for all* $x, y \in S$*.*

Proof: By assumption, $xd_1 = yd_1$ and $xd_2 = yd_2$ for all $x, y \in S$. Thus, for all $x, y \in S$

$$xd_1 = d_2d_1 = d_1d_2 = yd_2.$$

□

Lemma 3.19 *If there exist distinct elements* $d_1, d_2 \in \text{Center}(S)$ *with*

$$R_{d_1} = R_{d_2} = S \times S,$$

then $R_c = S \times S$ *for all* $c \in \text{Center}(S)$*. In particular,* S *does not have a multiplicative identity.*

Proof: Suppose such $d_1, d_2 \in \text{Center}(S)$ exist, and there is $c \in \text{Center}(S)$ with $R_c = \text{id}_S$. By Lemma 3.17, S has a multiplicative identity, 1_S. Also by Lemma 3.18,

$$d_1 = 1_S d_1 = 1_S d_2 = d_2,$$

contradicting the assumption that d_1 and d_2 are distinct. □

Lemma 3.20 *If there exists* $d \in \text{Center}(S)$ *with* $R_d = S \times S$*, then* d^2 *is either a zero or an infinity.*

Proof: By assumption, $xd^2 = d^2$ for all $x \in S$. It is easy to see that the relation $\sim$ defined by

$$x \sim y \quad \text{if} \quad x + d^2 = y + d^2$$

is a congruence relation. If $\sim = S \times S$, then for all $x \in S$

$$x + d^2 = d^2 + d^2 = (d + d)d = (d)d = d^2,$$

whence $d^2 = \infty$. If $\sim = \mathrm{id}_S$, then $x + d^2 = y + d^2$ iff $x = y$. For $x \in S$ let $z_x = x + d^2$. Then

$$z_x + d^2 = x + d^2 + d^2 = x + d^2,$$

so $z_x = x$. Thus $x + d^2 = x$ for all $x \in S$, so d^2 is an additive identity. Since it is also multiplicatively absorbing, it follows that d^2 is actually a zero. $\square$

Combining the above lemmas, we arrive at the following proposition.

Proposition 3.21 *Let S be a finite, additively commutative, c-simple semiring with* $\mathrm{Center}(S) \neq \emptyset$*. Then S has a zero, a one or an infinity.*

Finally, the relevance to additively idempotent semirings is given by the following lemma.

Lemma 3.22 *Let S be a finite, additively idempotent, additively commutative, c-simple semiring. Set*

$$\sigma = \sum_{x \in S} x.$$

Then either $\sigma \in \mathrm{Center}(S)$ or $\sigma^2 = \sigma$.

Proof: Suppose that $\sigma \notin \mathrm{Center}(S)$. Then there exists $c \in S$ such that $c\sigma \neq \sigma c$. Notice that $c\sigma + \sigma^2 = (c + \sigma)\sigma = \sigma^2$ and $\sigma c + \sigma^2 = \sigma(c + \sigma) = \sigma^2$. Thus

$$c\sigma + \sigma^2 = \sigma c + \sigma^2. \tag{3.2}$$

Consider the relation $\approx$ defined by

$$x \approx y \quad \text{if} \quad x + \sigma^2 = y + \sigma^2.$$

Certainly $\approx$ is an additive equivalence relation. To see that it is multiplicative, suppose $x \approx y$ and $\gamma \in S$. Then

$$\begin{aligned} x \approx y \;&\Rightarrow\; x + \sigma^2 = y + \sigma^2 \\ &\Rightarrow\; \gamma x + \gamma\sigma^2 = \gamma y + \gamma\sigma^2 \\ &\Rightarrow\; \gamma x + \gamma\sigma^2 + \sigma^2 = \gamma y + \gamma\sigma^2 + \sigma^2. \end{aligned}$$

Furthermore,

$$\begin{aligned}\gamma x+\sigma^2 &= \gamma x+(\gamma\sigma+\sigma)\sigma\\ &= \gamma x+\gamma\sigma^2+\sigma^2\end{aligned}$$

and

$$\begin{aligned}\gamma y+\sigma^2 &= \gamma y+(\gamma\sigma+\sigma)\sigma\\ &= \gamma y+\gamma\sigma^2+\sigma^2.\end{aligned}$$

Thus $\gamma x+\sigma^2=\gamma y+\sigma^2$, so $\gamma x\approx\gamma y$. Similarly, $x\gamma\approx y\gamma$, so $\approx$ is a congruence relation. But $c\sigma\neq\sigma c$ and $c\sigma\approx\sigma c$ by Equation 3.2, so $\approx = S\times S$. It follows that $\sigma\approx\sigma^2$, so

$$\sigma=\sigma+\sigma^2=\sigma^2+\sigma^2=\sigma(\sigma+\sigma)=\sigma(\sigma)=\sigma^2.$$

□

3.6 Main theorem

Combining Theorems 3.12, 3.13, 3.16 and Proposition 3.21 we have the following theorem.

Theorem 3.23 *Let S be a finite, additively commutative, congruence-simple semiring. Then one of the following holds:*

1. $|S|=2$.
2. $S\cong \mathrm{Mat}_n(\mathbb{F}_q)$ *for some finite field* $\mathbb{F}_q$ *and some* $n\geq 1$.
3. *S is a zero multiplication ring of prime order.*
4. *S is additively idempotent. Furthermore, if* $\mathrm{Center}(S)\neq\emptyset$ *and* $|S|>2$*, then S has a zero, one or an infinity.*

Observe the similarity between this theorem and Theorem 3.9. Recall that for a finite group G, $V(G)$ is a finite, additively commutative, c-simple semiring and is additively idempotent. So the semirings $V(G)$ do fall into the fourth case of Theorem 3.23. While we do believe that most of the additively idempotent semirings in the fourth case of this theorem are isomorphic to some $V(G)$, we should note that there are exceptions. In particular, the semiring given in Table 3.2 is one such exception.

This semiring is additively idempotent with a one. However, observe that the element a is 'almost' a zero and b is 'almost' an infinity. We conjecture that, except

Table 3.2. A C-SIMPLE SEMIRING OF ORDER 3

+	a	1	b
a	a	1	b
1	1	1	b
b	b	b	b

$\cdot$	a	1	b
a	a	a	b
1	a	1	b
b	a	b	b

for some such semirings of order 3, the additively idempotent c-simple semirings are those of the form $V(G)$ for finite groups G. This conjecture is based on an admittedly small amount of empirical evidence; specifically, the random generation of several thousand additively commutative semirings with small order (< 15).

3.7 Conclusion

We examine now the implications of Theorem 3.23 for the DLP in finite, additively commutative, c-simple semirings. The DLP is certainly trivial in a semiring of order two. In the second case of Theorem 3.23, S is actually a ring so nothing new is gained. Every DLP in a zero-multiplication ring has solution either zero or one, so this is trivial as well. The conclusion is the following: If S is a finite, additively commutative, c-simple semiring where the DLP is hard, then either $S \cong \mathrm{Mat}_n(\mathbb{F}_q)$ for some n and q, or S is additively idempotent.

Furthermore, if our conjecture on the structure of the semirings in the fourth case of Theorem 3.23 is correct, then the DLP can be maximally hard only if $(S \setminus \{\infty\}, \cdot) \cong G$ for some finite group G. If this is the case, the Pohlig-Hellman algorithm directly applies to any cyclic subgroup $\langle\alpha\rangle \subseteq G$. The DLP in this situation is at most as hard as in $G' \subseteq \langle\alpha\rangle$, where $|G'| = p$ and p is the largest prime dividing $|\langle\alpha\rangle|$. In particular, the conjecture would imply the strong result that if the DLP in a finite, additively commutative, c-simple semiring S is hard, then it can be assumed that $S \cong \mathrm{Mat}_n(\mathbb{F}_q)$ or $(S \setminus \{\infty\}, \cdot) \cong \mathbb{Z}_p$.

CHAPTER 4

SEMIGROUP ACTIONS

Previously, we examined a generalization of DLP settings to semirings. In this chapter, we consider a generalization of the discrete logarithm problem itself as the building block for an asymmetric cipher. We will look carefully at the minimum requirements for a DLP-type problem to extend to a DHP-type problem that is usable for a key-exchange protocol. In addition, we will see what requirements are necessary for such a generalized DHP-type problem to extend to an ElGamal type protocol.

Some of the research in this chapter was done in collaboration with Joachim Rosenthal, Gerard Maze, and Josep Climent [42, 30].

4.1 Extended Diffie-Hellman and ElGamal

In this section we show how the DLP can be considered as a special case of an action by a semigroup. The idea of using group actions to construct cryptographic protocols is not a new one; Yamamura [51] has been considering a group action of $Sl_2(\mathbb{Z})$ and Blackburn and Galbraith [7] have been investigating the system of [51]. Our goal, however, is different than that of the related papers in the literature; we would like to first generalize the existing framework to find some minimum requirements for a DH-type key exchange. We will then explore a particular example, showing that one must choose parameters judicially.

Problem 4.1 Let H be a finite semigroup [group] acting on a finite set X. Given $x, y \in X$ with $y = gx$ for some $g \in H$, the *Semigroup Action Problem* (SAP) [*Group Action Problem* (GAP)] asks for $\gamma \in H$ such that $y = \gamma x$.

Comparing this with Problem 1.3, we see that the discrete logarithm problem is a special case of the group action problem with $H = (\mathbb{Z}, \cdot)$ and $X = G$, a group, and the action given by

$$(n, g) \longmapsto g^n.$$

In the same way the DLP gives rise to the Diffie-Hellman key exchange, the SAP gives rise to a generalized Diffie-Hellman key exchange.

Protocol 4.2 (Extended Diffie-Hellman Key Exchange) Let X be a finite set and H an abelian semigroup acting on X. The *Extended Diffie-Hellman key exchange* is the following protocol:

1. Alice and Bob agree on an element $x \in X$.

2. Alice chooses $a \in H$ and computes ax. Alice's private key is a, her public key is ax.

3. Bob chooses $b \in H$ and computes bx. Bob's private key is b, his public key is bx.

4. Their shared secret key is then

$$a(bx) = (a \cdot b)x = (b \cdot a)x = b(ax).$$

If, in addition, X has a group structure, this extends to an ElGamal-type cryptosystem.

Protocol 4.3 (Extended ElGamal cryptosystem) Let $(X, \circ)$ be a finite group and H an abelian semigroup acting on X. The *Extended ElGamal cryptosystem* is the following protocol:

1. Alice chooses $a \in H$, $x \in X$, and computes $\alpha = ax$. She publishes her public key (x, α).

2. Bob wishes to send Alice the message $m \in X$. He first obtains her public key (x, α).

3. Bob chooses a random element $b \in H$ and computes $\beta = bx$, $\gamma = (b\alpha) \circ m$ and sends the pair (β, γ) to Alice.

4. Alice recovers m by computing

$$(a\beta)^{-1} \circ \gamma = (abx)^{-1} \circ (bax \circ m) = (abx)^{-1} \circ (abx \circ m) = m.$$

4.2 Pollard-rho for group actions

We present here a Pollard-rho type birthday attack for the special case when H is actually a group. It is well known that if H satisfies left-cancellation it is embeddable in a finite group [48]. Thus this attack may also apply if H satisfies left-cancellation.

To facilitate an understanding of the algorithm, we first give the motivation. Observe that if G is a group acting on a set X, $x, y \in X$ and $y = gx$ for some $g \in G$, then $g^{-1}y = x$. Thus x and y have the same orbit $\mathcal{O}_x$. Suppose now that $\{a_i\}, \{b_j\}$ are two random sequences in G and consider the sequence

$$\{a_1 y, b_1 x, a_2 y, b_2 x, \ldots\} \subseteq \mathcal{O}_x.$$

We can expect to find a collision in this sequence after approximately $\sqrt{|\mathcal{O}_x|}$ elements. With probability 1/2, the collision will take the form $a_i y = b_j x$, in which case we have $y = a_i^{-1} b_j x$, solving the group action problem.

The question remains if we may find this collision without requiring the storage of $\sqrt{|\mathcal{O}_x|}$ elements. For this, suppose $f : X \longrightarrow G$ is a randomly chosen function. Choose $a_1, b_1 \in G$ randomly and define

$$\begin{aligned} a_{n+1} &= f((\prod_{i=1}^{i=n} a_i) y), \\ b_{n+1} &= f((\prod_{i=1}^{i=n} b_i) x). \end{aligned}$$

So if $a_i y = b_j x$ for some i, j, we have $a_{i+k} = b_{j+k}$ for all $k \geq 1$.

Algorithm 4.4 (Pollard-rho for group actions)
Input: A finite group G acting on a set X, and elements $x, y \in X$ with $y = \alpha x$ for some $\alpha \in G$.
Output: An element $\gamma \in G$ such that $y = \gamma x$.

1. Choose a random function $f : X \longrightarrow G$, and a random element $a_1 \in G$. Compute $a_2 \leftarrow f(a_1 y)$. Set $a \leftarrow a_1, \hat{a} \leftarrow a_2 a_1, i \leftarrow 1$.

2. [Find a loop] If $a_i = a_{2i}$, goto 4.

3. Compute $a_{i+1} \leftarrow f(ay), a_{2i+1} \leftarrow f(\hat{a}y), a_{2i+2} \leftarrow f(a_{2i+1}\hat{a}y)$. Set $a \leftarrow a_{i+1}a$, $\hat{a} \leftarrow a_{2i+2}a_{2i+1}\hat{a}$ and $i \leftarrow i + 1$. Goto 2.

4. Choose a random element $b_1 \in G$. Compute $b_2 \leftarrow f(b_1 x)$. Set $b \leftarrow b_1$, $\hat{b} \leftarrow b_2 b_1, j \leftarrow 1$.

5. [Find b loop] If $b_j = b_{2j}$, goto 7.

6. Compute $b_{j+1} \leftarrow f(bx), b_{2j+1} \leftarrow f(\hat{b}x), b_{2j+2} \leftarrow f(b_{2j+1}\hat{b}x)$. Set $b \leftarrow b_{j+1}b$, $\hat{b} \leftarrow b_{2j+2}b_{2j+1}\hat{b}$ and $j \leftarrow j+1$. Goto 5.

7. [Find a and b collision] Compute $b_{j+1} \leftarrow f(bx)$, $b \leftarrow b_{j+1}b$. Set $j \leftarrow j+1$.

8. If $bx = ay$, goto 10. Otherwise, if $b = \hat{b}$, goto 1 and choose different starting parameters.

9. Compute $b_{j+1} \leftarrow f(bx)$, $b \leftarrow b_{j+1}b$. Set $j \leftarrow j+1$. Goto 8.

10. [Calculate solution] Set $\gamma \leftarrow a^{-1}b$. Output γ and stop.

Observe that this algorithm has a small, fixed memory requirement. In particular, one needs only ten variables to store the current values of

$$a, \hat{a}, a_i, a_{2i}, i, b, \hat{b}, b_j, b_{2j}, j.$$

We wish now to determine the expected number of operations performed by Algorithm 4.4. Notice that the algorithm may be restarted in step 8. Suppose that the probability of the algorithm being restarted in step 8 is $p < 1$. If we let $N = |\mathcal{O}_x|$, the expected number of times that step 2 will be executed is certainly $O(\frac{\sqrt{N}}{p})$. The same is also true for steps 5 and 8. We wish now to find the probability p. At any point during the algorithm when a or b is defined, we have

$$a = \prod_{k=1}^{i} a_k \quad \text{and} \quad b = \prod_{k=1}^{j} b_k,$$

where $\prod$ denotes the left product. So if $(\prod_{k=1}^{n} a_k)y = (\prod_{k=1}^{m} b_k)x$ for some n, m, then $a_{n+1} = b_{m+1}$, whence

$$(\prod_{k=1}^{n+1} a_k)y = (\prod_{k=1}^{m+1} b_k)x.$$

By induction,

$$(\prod_{k=1}^{n+l} a_k)y = (\prod_{k=1}^{m+l} b_k)x \ \text{ for all } l \geq 1$$

This shows that step 10 will be reached if $(\prod_{k=1}^{n} a_k)y = (\prod_{k=1}^{m} b_k)x$ for some n, m. If we let $y_i = (\prod_{k=1}^{i} a_k)y$ and $x_j = (\prod_{k=1}^{j} b_k)x$, then the sequences $\{y_i\}, \{x_j\} \subseteq \mathcal{O}_x$

are 'nearly random', by assumption. That is, they are random upto the points where they begin to repeat, since f is a randomly chosen function. Thus, we expect that the sequence

$$\{y_1, x_1, y_2, x_2, \ldots\}$$

contains two identical elements somewhere in the first $\sqrt{|O_x|}$ elements. Since the sequences are assumed to be random, with probability 1/2 the collision will take the form $y_i = x_j$ for some i, j, whence $p \geq 1/2$. Thus the expected runtime of the algorithm is $O(\sqrt{|\mathcal{O}_x|})$.

For a particular group action, it is possible that some of the improvements to the Pollard-rho algorithm given in [45, 50] may give corresponding improvements to the constant in the runtime estimate here.

Remark 4.5 It may be possible to extend Algorithm 4.4 so that it solves the semigroup action problem as well. In particular, Algorithm 1.8 can be used to find the cycle length of an element in a semigroup. From this, one can easily find the cycle start of such an element. It may then be possible to modify Algorithm 4.4 in such a way that element inversion is not necessary – only the ability to perform subtraction on exponents. While we do not have such a modification at present, it would be prudent to assume that such an extension is possible.

4.3 Matrix action on abelian groups

We have already seen one example of a cryptosystem based on the semigroup action problem–the Diffie-Hellman system. In general, much care must be taken in choosing suitable parameters for which the semigroup action problem is hard. We present here an example for which the SAP may be hard, though we have no concrete example with efficient key sizes at present.

Take as a semigroup $\mathrm{Mat}_n(\mathbb{Z})$ with multiplication. Fix a finite abelian semigroup $(G, \cdot)$ and consider

$$X = G^n = \underbrace{G \times G \times \cdots \times G}_{n \text{ copies}}.$$

Since X is a $\mathbb{Z}$-module, the semigroup $\mathrm{Mat}_n(\mathbb{Z})$ acts on X via the formal multiplication

$$\left(\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}, \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \right) \longmapsto \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \text{where } y_i = \prod_{i=j}^{n} x_j^{a_{ij}}. \tag{4.1}$$

Remark 4.6 If $n = 1$, the action reduces to $(a, x) \mapsto x^a$, and so the SAP is actually a DLP in the semigroup G.

Since examples of finite abelian groups are well-known and plentiful, it is natural to wonder what the implications are if G is a group. In the sequel, we assume that G is a group and derive some implications on possible choices of parameters in that case.

Let $l = |G|$ so that we may consider the action as $\mathrm{Mat}_n(\mathbb{Z}_l)$ acting on G^n. It was required in Protocol 4.2 that we have an abelian semigroup acting on a set. To meet this requirement we fix some $A \in \mathrm{Mat}_n(\mathbb{Z}_l)$ and consider the restricted action of the abelian sub-semigroup $\mathbb{Z}_l[A] = \{f(A) \mid f(t) \in \mathbb{Z}_l[t]\}$ on G^n.

The EDH cryptosystem based on this semigroup action then works as follows:

- Alice and Bob agree on a finite abelian group G with $|G| = l$, and a positive integer n. They also agree on a matrix $A \in \mathrm{Mat}_n(\mathbb{Z}_l)$ and a column vector $x = [x_1, \ldots, x_n] \in G^n$.
- Alice chooses $M_a \in \mathbb{Z}_l[A]$ and computes $\alpha = M_a x$, sending α to Bob.
- Bob chooses $M_b \in \mathbb{Z}_l[A]$ and computes $\beta = M_b x$, sending β to Alice.
- Alice and Bob can each compute the shared secret

$$y = M_a M_b x = M_a \beta = M_b \alpha.$$

An eavesdropper Eve can certainly break the system if she can find $M \in \mathbb{Z}_l[A]$ with $Mx = \alpha$, since

$$M\beta = M(M_b x) = MM_b x = M_b(Mx) = M_b \alpha = y.$$

In general, $\mathbb{Z}_l[A]$ is not a multiplicative group, and so Algorithm 4.4 does not directly apply. However, the algorithm can be adapted to fit this situation.

Let $\chi_A(t) \in \mathbb{Z}_l[t]$ be the characteristic polynomial of A. Since $\mathbb{Z}_l[t]$ is not necessarily a unique factorization domain, the minimal polynomial of A may not be well-defined. However, since $\chi_t(A) = 0$, there does exist a monic polynomial $m_A(t)$ of minimal degree $d < n$, such that $m_A(A) = 0$. If $m_A(t) = c_0 + c_1 t + \cdots + c_{d-1} t^{d-1} + t^d$, we have

$$A^d = -c_0 I - c_1 A - \cdots - c_{d-1} A^{d-1}.$$

In particular, this implies that

$$\mathbb{Z}_l[A] = \{a_0 I + a_1 A + \cdots + a_{d-1} A^{d-1} \mid a_i \in \mathbb{Z}_l\}.$$

Algorithm 4.4 can certainly find u_i, v_i such that

$$Ux = (u_0 I + u_1 A + \cdots + u_{d-1}A^{d-1})x = (v_0 I + v_1 A + \cdots + v_{d-1}A^{d-1})\alpha = V\alpha. \quad (4.2)$$

If the matrix on the RHS happens to be invertible, this certainly solves Eve's problem (since the inverse will also be in $\mathbb{Z}_l[A]$). We wish, however, to solve the problem in the general case.

It follows from Equation 4.1 that for any $M_1, M_2 \in \mathrm{Mat}_n(\mathbb{Z}_l)$

$$(M_1 + M_2)\alpha = (M_1\alpha)(M_2\alpha). \quad (4.3)$$

Suppose now that Algorithm 4.4 has been used $k+1$ times to find matrices $U_i, V_i \in \mathbb{Z}_l[A]$ such that $U_0 x = V_0\alpha, \ldots U_k x = V_k\alpha$. Equation 4.3 implies that that for all $c_0, \ldots, c_k \in \mathbb{Z}_l$

$$(c_0 U_0 + \cdots + c_k U_k)x = (c_0 V_0 + \cdots + c_k V_k)\alpha.$$

Our goal is to find such matrices and constants such that $c_0 V_0 + \cdots + c_k V_k = I$, which will clearly solve the problem. Observe that the output of Algorithm 4.4 actually gives each V_i as a sum $V_i = v_{i,0}I + v_{i,1}A + \cdots + v_{i,d-1}A^{d-1}$. We then have

$$\begin{aligned} U_0 x &= (v_{0,0}I + \cdots + v_{0,d-1}A^{d-1})\alpha, \\ &\vdots \\ U_k x &= (v_{k,0}I + \cdots + v_{k,d-1}A^{d-1})\alpha. \end{aligned}$$

Since the $v_{i,j}$ are chosen randomly with a uniform distribution, $v_{i,0}$ will be invertible with probability $\phi(l)/l \geq 1/3$. We may thus assume that some $v_{i,0}$ is invertible. Consequently, it is expected that elementary row operations will the desired result so long as $k \geq \max\{3, d\}$. This gives a solution to this particular matrix action problem in time $O(d\sqrt{|\mathcal{O}_x|})$, where $\mathcal{O}_x = \{Mx \mid M \in \mathbb{Z}_l[A]\}$.

We give now a Pohlig-Hellman type reduction. Suppose $l = rs$ with $(r,s) = 1$. Then $\mathbb{Z}_l \cong \mathbb{Z}_r \times \mathbb{Z}_s$ and so $\mathrm{Mat}_n(\mathbb{Z}_l) \cong \mathrm{Mat}_n(\mathbb{Z}_r) \times \mathrm{Mat}_n(\mathbb{Z}_s)$. Recall that $l = |G|$ and $x, \alpha \in G^n$. Thus, x^r, α^r lie in a subgroup of G^n with order dividing s. Similarly, x^s, α^s lie in a subgroup of G^n with order dividing r. Applying the previous algorithm twice will give $M_r \in \mathbb{Z}_r[\pi_r(A)], M_s \in \mathbb{Z}_s[\pi_s(A)]$ with

$$\begin{aligned} M_r x^s &= \alpha^s, \\ M_s x^r &= \alpha^r, \end{aligned}$$

where π_r and π_s are the canonical projections from $\mathbb{Z}_l$ onto $\mathbb{Z}_r$ and $\mathbb{Z}_s$ respectively. One may use the Euclidean algorithm to find integers c, d such that $cr + ds = 1$. Setting $M = dsM_r + crM_s$ we have

$$\begin{aligned} Mx &= (dsM_r + crM_s)x \\ &= (dsM_r x)(crM_s x) \\ &= (dM_r x^s)(cM_s x^r) \\ &= (dI\alpha^s)(cI\alpha^r) \\ &= \alpha^{ds}\alpha^{cr} \\ &= \alpha^{ds+cr} \\ &= \alpha. \end{aligned}$$

If $l = p_1^{e_1} \cdots p_k^{e_k}$ with $p_1^{e_1} < \cdots < p_k^{e_k}$, this process may be inductively carried out so that the expected number of operations is $O(d \sum_{i=1}^{k} \sqrt{|\mathcal{O}_{x,i}|})$ where

$$\mathcal{O}_{x,i} = \{Mx^{l/p_i^{e_i}} \mid M \in \mathbb{Z}_{p_i^{e_i}}[\pi_{p_i^{e_i}}(A)]\}.$$

In particular, since

$$|\mathcal{O}_{x,i}| \leq |\{M \in \mathbb{Z}_{p_i^{e_i}}[\pi_{p_i^{e_i}}(A)]\}| \leq p_i^{de_i},$$

the expected number of operations is bounded above by $O\left(kd\sqrt{p_k^{de_k}}\right)$.

Neglecting the size of the matrix A, the key size in this example is the size of an element of G^n. Since $|G| = l$, the key size is presumably

$$N = \log_2 l = e_1 \log_2 p_1 + \cdots + e_k \log_2 p_k.$$

Recall that other systems exist where the best known attack with key size N requires $O(2^{N/2})$ operations. To compete with the efficiency of these systems, we need to choose parameters so that

- $l = |G| = p^k$ for some prime p and some $k > 1$ (we will see shortly that $k = 1$ is a poor choice).

- $d = n$ (i.e., the rank of A is n).

If parameters are chosen that meet these conditions, we will have a key size of $N = nk \log_2 p$. The expected number of operations needed for this attack to succeed will be $O(n\sqrt{p^{nk}}) = O(n2^{N/2})$. However, choosing A with full rank makes more than half of the elements of $\mathbb{Z}_{p^k}[A]$ invertible. Using this observation, the algorithm can be tweaked to need only $O(2^{N/2+1})$ operations, which does still compare favorably with the best known systems, if there exist parameters satisfying the assumptions.

The following propositions impose further conditions on the choice of parameters.

Proposition 4.7 *Let $G, A, x = [x_1, \ldots, x_n]$, and $\alpha = [\alpha_1, \ldots, \alpha_n]$ be as above. Suppose further that the x_i all lie in a common cyclic subgroup $\langle g \rangle$ of G. Then the problem of finding $M \in \mathbb{Z}_l[A]$ such that $Mx = \alpha$ polynomial-time reduces to $2n$ discrete log problems in $\langle g \rangle$.*

Proof: Since $x_i, \alpha_i \in \langle g \rangle$, there exist unique integers, $e_i, a_i \in \{0, 1, \ldots, |g| - 1\}$ such that $x_i = g^{e_i}$ and $\alpha_i = g^{a_i}$. Then

$$M \begin{bmatrix} g^{e_1} \\ \vdots \\ g^{e_n} \end{bmatrix} = \begin{bmatrix} g^{a_1} \\ \vdots \\ g^{a_n} \end{bmatrix} \Longleftrightarrow M \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

Where the product on the right is the usual matrix product on a vector. One can certainly in polynomial time find such an M. $\square$

The computation of $2n$ discrete logs in $\langle g \rangle$ can be performed with $O(2np^{k/2})$ operations. Since this is far less than the $O\left(2^{(\text{keysize})/2}\right)$ we hope for, it is a situation to be avoided.

By previous observations, we may now assume that $l = p^k$.

Proposition 4.8 *Let G, A, x, and α be as above. Further suppose that A is diagonalizable over $\mathbb{Z}_{p^k}$. Then the problem of finding $M \in \mathrm{Mat}_n(\mathbb{Z}_{p^k})$ such that $MA = AM$ and $Mx = \alpha$ polynomial-time reduces to n discrete log problems in G.*

Proof: One may compute the characteristic polynomial $\chi_A(t)$ of A in polynomial time. Since A is assumed to be diagonalizable, $\chi_A(t)$ factors (not necessarily uniquely) as a product of linear factors over $\mathbb{Z}_{p^k}$. Using algorithms derived from Hensel's lemma [4], one may factor $\chi_A(t)$ into a such product of linear factors in probabilistic polynomial time. In the usual way, one may find then find a set of n linearly independent eigenvectors in $(\mathbb{Z}_{p^k})^n$ and hence an invertible matrix U such that $UAU^{-1} = Z$ for some diagonal matrix $Z \in \mathrm{Mat}(\mathbb{Z}_{p^k})$.

Observe now that UMU^{-1} is a diagonal matrix for all $M \in \mathbb{Z}_{p^k}[A]$. In particular, if $M_a \in \mathbb{Z}_{p^k}[A]$ is a matrix such that $M_a x = \alpha$, then $UM_aU^{-1} = D_a$ for some diagonal matrix D_a and

$$D_a(U^{-1}x) = U^{-1}\alpha.$$

This implies that there necessarily exists a solution D to the general equation

$$D(U^{-1}x) = U^{-1}\alpha.$$

Such a solution may obviously be found by computing n discrete logarithms in G. If D_0 is such a solution, then

$$(UD_0U^{-1})x = \alpha.$$

Furthermore, if $M = UD_0U^{-1}$, then

$$MA = (UD_0U^{-1})(UZU^{-1}) = UD_0ZU^{-1} = UZD_0U^{-1} = AM.$$

□

Also observe that if G is an abelian group with $|G| = p^k$, then there exists an isomorphism

$$\phi : H \longrightarrow \mathbb{Z}_{p^{e_1}} \oplus \cdots \oplus \mathbb{Z}_{p^{e_r}}$$

for some r and $e_1, \ldots, e_r$. By Proposition 4.7, G should be chosen so that $r > 1$. Furthermore, if the isomorphism ϕ can be efficiently computed, the induced action of $\mathrm{Mat}_n(\mathbb{Z}_{p^k})$ on $(\mathbb{Z}_{p^{e_1}})^n \oplus \cdots \oplus (\mathbb{Z}_{p^{e_r}})^n$ can be easily computed. One may then solve the induced semigroup action problem coordinate-wise on cyclic groups. Thus, if G is chosen to be a group, it should be chosen so that the isomorphism ϕ is hard to compute on arbitrary elements.

Remark 4.9 Recall that these requirements were all derived with the assumption that G is a group. If G is a semigroup, the same considerations no longer directly apply. Indeed it may be the case that this semigroup action problem is significantly harder in the case where G is an abelian semigroup that is not a group.

Example 4.10 Here we will give an illustrative with small parameters. Consider the elliptic curve equation

$$E: \quad y^2 = x^3 + x + 47$$

over $\mathbb{F}_{71}$. The group of rational points of this curve is

$$E(\mathbb{F}_{71}) \cong \mathbb{Z}_5 \oplus \mathbb{Z}_{15}$$

Let $P_1 = (1, 7)$, $P_2 = (51, 11)$ and $P_3 = (49, 58)$. In this small example, we may verify by brute force that these points do not lie in a common cyclic subgroup.

$$\begin{aligned}
\langle P_1 \rangle &= \{(1,7), (43,52), (43,19), (1,64), (\mathrm{id})\} \\
\langle P_2 \rangle &= \{(51,11), (70,20), (70,51), (51,60), (\mathrm{id})\} \\
\langle P_3 \rangle &= \{(49,58), (60,57), (60,14), (49,13), (\mathrm{id})\}
\end{aligned}$$

Thus, the subgroup $G = \langle P_1, P_2, P_3 \rangle \subseteq E(\mathbb{F}_{71})$ is isomorphic to $\mathbb{Z}_5 \oplus \mathbb{Z}_5$.

Suppose now that Alice and Bob have agreed on

$$A = \begin{pmatrix} 3 & 1 & 1 \\ 2 & 2 & 4 \\ 1 & 2 & 3 \end{pmatrix} \in \mathrm{Mat}_3(\mathbb{Z}_5)$$

and will use the action of $\mathbb{Z}_5[A]$ on G^3 to agree upon a secret key. Observe that the characteristic polynomial of A is $\chi_M(t) = t^3 + 2t^2 + 1$, which is irreducible over $\mathbb{Z}_5$. This guarantees that A is not diagonalizable over $\mathbb{Z}_5$ and that A has maximal order: $|A| = 124$. Alice and Bob then agree on the vector

$$x = \begin{bmatrix} (1,7) \\ (51,11) \\ (49,58) \end{bmatrix} \in G^3$$

Alice chooses

$$M_a = \begin{pmatrix} 3 & 4 & 1 \\ 3 & 0 & 2 \\ 1 & 1 & 2 \end{pmatrix} \in \mathbb{Z}_5[A].$$

and computes

$$\alpha = \begin{pmatrix} 3 & 4 & 1 \\ 3 & 0 & 2 \\ 1 & 1 & 2 \end{pmatrix} \begin{bmatrix} (1,7) \\ (51,11) \\ (49,58) \end{bmatrix} = \begin{bmatrix} (1,7) \\ (58,31) \\ (44,69) \end{bmatrix}.$$

She sends α to Bob. Bob chooses

$$M_b = \begin{pmatrix} 3 & 2 & 4 \\ 4 & 2 & 4 \\ 4 & 2 & 2 \end{pmatrix} \in \mathbb{Z}_5[A].$$

and computes

$$\beta = \begin{pmatrix} 3 & 2 & 4 \\ 4 & 2 & 4 \\ 4 & 2 & 2 \end{pmatrix} \begin{bmatrix} (1,7) \\ (51,11) \\ (49,58) \end{bmatrix} = \begin{bmatrix} (51,11) \\ (39,7) \\ (51,60) \end{bmatrix},$$

sending β to Alice. Alice and Bob then share the secret key

$$k = M_a\beta = M_b\alpha = \begin{bmatrix} (69,45) \\ (51,11) \\ (1,7) \end{bmatrix}.$$

Note that this example is not optimally efficient because the orbit of x, $\mathcal{O}_x$, is too small:

$$|\mathcal{O}_x| = 25 < |\mathbb{Z}_5[A]| = 125.$$

Choosing $n = 2$ (i.e., 2×2 matrices) would have corrected this flaw, but such an example would have been less illustrative.

4.4 Conclusions

In this dissertation we have examined some generalizations of asymmetric ciphers based on the difficulty of the discrete log problem.

It was first illustrated how using the DLP in haphazardly chosen rings may not be optimal. In particular, the DLP in the ring $\mathbb{F}_q[\underline{x}]/I$ with I zero-dimensional is a poor choice since it polynomial-time reduces to some DLPs in finite fields. The difficulty of the reduced problem is maximized if I is simple, in which case $\mathbb{F}_q[\underline{x}]/I$ is itself a finite field. However, the examination of the DLP in this ring did give rise to an algorithm for computing the primary decomposition of zero-dimensional ideals over $\mathbb{Q}$.

In Chapter 3 we discussed the possibility of using the DLP in finite, additively commutative semirings. Since the use of simple structures is the most reliable way to avoid Pohlig-Hellman type attacks, we restricted our attention to congruence-simple semirings. We classified such semirings except for the additively idempotent ones and showed that if the DLP is hard in such a semiring S, then $S \cong \mathrm{Mat}_n(\mathbb{F}_q)$, or S is additively idempotent. We also conjectured that if $|S| > 3$, this latter case implies $(S \setminus \{\infty\}, \cdot) \cong G$ for some finite group G.

Finally, in this last chapter, it was shown how every action of a finite semigroup H on a finite set gives rise to an extended Diffie-Hellman key exchange. We gave a Pollard-rho type attack for the case where H is a group. Section 4.3 presented a matrix action on finite abelian semigroups that may give rise to a hard problem on which to build an asymmetric cipher.

BIBLIOGRAPHY

[1] W. W. Adams and P. Loustaunau. *An Introduction to Gröbner Bases*, volume 3 of *Graduate Studies in Mathematics*. American Mathematical Society, Rhode Island, 1994.

[2] I. Anshel, M. Anshel, and D. Goldfeld. An algebraic method for public-key cryptography. *Mathematical Research Letters*, 6:287–291, 1999.

[3] M. Anshel and D. Goldfeld. Zeta functions, one-way functions, and pseudorandom number generators. *Duke Math. J.*, 88:371–390, 1997.

[4] E. Bach and J. Shallit. *Algorithmic Number Theory, Vol. 1.* The MIT Press, Cambridge, 1996.

[5] R. El Bashir, J. Hurt, A. Jančařék, and T. Kepka. Simple commutative semirings. *Journal of Algebra*, 236:277–306, 2001.

[6] T. Becker and V. Weispfenning. *Gröbner Bases*. Graduate Texts in Mathematics. Springer-Verlag, New York, 1993. A computational approach to commutative algebra. In cooperation with Heinz Kredel.

[7] S.R. Blackburn and S.D. Galbraith. Cryptanalysis of two cryptosystems based on group actions. In *Advances in Cryptology – ASIACRYPT '99*, volume 1716 of *Lecture Notes in Computer Science*, pages 52–61. Springer Verlag, Berlin, 1999.

[8] I. Blake, G. Seroussi, and N. Smart. *Elliptic Curves in Cryptography.* Lecture Note Series 265. London Mathematical Society, Cambridge, 1999.

[9] H. Cohen. *A Course in Computational Algebraic Number Theory.* Springer-Verlag, New York, third edition, 1993. Corrected Printing, 1996.

[10] D. Coppersmith. Fast evaluation of logarithms in fields of characteristic two. *IEEE Trans. Inform. Theory*, 30:587–594, 1984.

[11] D. Cox, J. Little, and D. O'Shea. *Using Algebraic Geometry.* Springer-Verlag, New York, 1998.

[12] W. Diffie and M. E. Hellman. New directions in cryptography. *IEEE Trans. Inform. Theory*, IT-22(6):644–654, 1976.

[13] W. Diffie and M.E. Hellman. Multiuser cryptographic techniques. In *Proc. of AFIPS National Computer Conference*, pages 109–112. AFIPS Press, Montvale, NJ, 1976.

[14] D. Eisenbud, C. Huneke, and W. Vasconcelos. Direct methods for primary decomposition. *Inventiones Mathematicae*, 110:207–235, 1992.

[15] T. ElGamal. A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Trans. Inform. Theory*, 31(4):469–472, 1985.

[16] J. C. Faugère, P. Gianni, D. Lazard, and T. Mora. Efficient computation of zero-dimensional Gröbner bases by a change of ordering. *Journal of Symbolic Computation*, 16:329–344, 1993.

[17] P. Gianni, B. Trager, and G. Zacharias. Gröbner bases and primary decomposition of polynomial ideals. *J. Symbolic Computation*, 6:149–167, 1988.

[18] D. M. Gordon. Discrete logarithms in GF(p) using the number field sieve. *SIAM Journal on Discrete Mathematics*, 6:124–138, 1993.

[19] Udo Hebisch and Hanns Joachim Weinert. *Semirings and Semifields*, pages 425–462. Handbook of Algebra, Vol. 1. Elsevier Science B.V., Amsterdam, 1996.

[20] T. W. Hungerford. *Algebra*. Graduate Texts in Mathematics. Springer, New York, 1980.

[21] D.E. Knuth. *The Art of Computer Programming: Vol. 2 / Seminumerical Algorithms*. Addison-Wesley, Reading, MA, 3rd edition, 1988.

[22] N. Koblitz. Elliptic curve cryptosystems. *Math. Comp.*, 48(177):203–209, 1987.

[23] N. Koblitz. *Algebraic Aspects of Cryptography*. Springer-Verlag, Berlin, 1998. With an appendix by A. J. Menezes, Y.-H. Wu and R. J. Zuccherato.

[24] W. Kuich and A. Salomaa. *Semirings, Automata, Languages*, volume 5 of *EATCS Monographs on Theoretical Computer Science*. Springer-Verlag, Berlin, 1986.

[25] D. Lazard. Solving zero-dimensional algebraic systems. *Journal of Symbolic Computation*, 13:117–131, 1992.

[26] A.K. Lenstra. *Polynomial Time Algorithms for the Factorization of Polynomials*. PhD thesis, University of Amsterdam, 1984.

[27] H. Lenstra. Finding isomorphisms between finite fields. *Mathematics of Computation*, 56:329–347, 1991.

[28] H.W. Lenstra, A.K. Lenstra, and L. Lovász. Factoring polynomials with rational coefficients. *Math. Ann.*, 261:515–534, 1982.

[29] U. M. Maurer. Towards the equivalence of breaking the Diffie-Hellman protocol and computing discrete logarithms. In *Advances in cryptology—CRYPTO '94 (Santa Barbara, CA, 1994)*, pages 271–281. Springer, Berlin, 1994.

[30] G. Maze, C. Monico, J. Climent, and J. Rosenthal. Public key cryptography based on simple modules over simple rings. to appear in Proceedings of MTNS 2002, 2002.

[31] R. J. McEliece. A public-key cryptosystem based on algebraic coding theory. Technical report, DSN Progress report # 42-44, Jet Propulsion Laboratory, Pasadena, Californila, 1978.

[32] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone. *Handbook of Applied Cryptography*. CRC Press Series on Discrete Mathematics and its Applications. CRC Press, Boca Raton, FL, 1997. With a foreword by Ronald L. Rivest.

[33] A. J. Menezes and Y.-H. Wu. The discrete logarithm problem in $\mathrm{Gl}(n, q)$. *Ars Combin.*, 47:23–32, 1997.

[34] A.J. Menezes and S.A. Vanstone. A note on cyclic groups, finite fields, and the discrete logarithm problem. *AAECC*, 3:67–74, 1992.

[35] S.S. Mitchell and P.B. Fenoglio. Congruence-free commutative semirings. *Semigroup Forum*, 37:79–91, 1988.

[36] C. Monico. Computing the primary decomposition of zero-dimensional ideals. Preprint, 2000.

[37] C. Monico, J. Rosenthal, and A. Shokrollahi. Using low density parity check codes in the McEliece cryptosystem. In *Proceedings of the 2000 IEEE International Symposium on Information Theory*, page 215, Sorrento, Italy, 2000.

[38] J. Patarin. *Hidden field equations (HFE) and isomorphisms of polynomials (IP): two new families of asymetric algorithms*, pages 33–48. Eurocrypt 96. Springer Verlag, 1996.

[39] G. Pfister, W. Decker, and G. Greuel. In B. Matzat, G. Greuel, and G. Hiss, editors, *Algorithmic Algebra and Number Theory*, pages 187–220. Springer, Berlin, 1999.

[40] S.C. Pohlig and M.E. Hellman. An improved algorithm for computing logarithms over GF(p) and its cryptographic significance. *IEEE Trans. Inform. Theory*, 24:106–110, 1978.

[41] J. M. Pollard. Monte Carlo methods for index computation (mod p). *Mathematics of Computation*, 32:918–924, 1978.

[42] J. Rosenthal, G. Maze, and C. Monico. A public-key cryptosystem based on actions by semigroups. to appear in Proceedings of ISIT 2002, 2002.

[43] C. E. Shannon. Communication theory of secrecy systems. *Bell System Tech. J.*, 28:656–715, 1949.

[44] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, 1997.

[45] E. Teske. On random walks for Pollard's rho method. *Mathematics of Computation*, 70:809–825, 2000.

[46] University of Kaiserslautern. Singular : A computational algebra package. URL:http://www.mathematik.uni-kl.de/zca/Singular, 2002.

[47] H.S. Vandiver. Note on a simple type of algebra in which the cancellation law of addition does not hold. *Bulletin of the American Mathematical Society*, 40:916–920, 1934.

[48] S. Warner. *Modern Algebra*. Dover Publications, Inc., New York, 1965, 1990.

[49] H.J. Weinert. On 0-simple semirings, semigroup semirings, and two kinds of division semirings. *Semigroup Forum*, 28:313–333, 1984.

[50] M. Wiener and R. Zuccherato. *Faster attacks on elliptic curve cryptosystems*, volume 1556 of *Lecture Notes in Computer Science*, pages 190–200. Springer-Verlag, 1999.

[51] A. Yamamura. Public-key cryptosystems using the modular group. In *Public Key Cryptography*, volume 1431 of *Lecture Notes in Computer Science*, pages 203–216. Springer, Berlin, 1998.