Universidade de São Paulo
Instituto de Matemática e Estatística
Bacharelado em Matemática Aplicada e Computacional

Anderson Reis Rosa

Uma Aplicação da Teoria dos Números e Curvas Elípticas à Criptografia

Supervisor: Prof. Dr. Kostiantyn Iusenko

São Paulo
Setembro de 2019

Resumo

A criptografia é uma ferramenta indispensável para proteger informações em sistemas computacionais, ela também é uma disciplina fascinante pois cruza diversas áreas de conhecimentos (ciência da computação, matemática, estatística, engenharia elétrica, física e teoria da comunicação). Ela tem uma função fundamental que é encapsular mensagens, por isso, ela tem um papel imprescindível em uma comunicação minimamente segura.

Apesar de existirem outros fatores que tornam uma comunicação segura, vamos nos concentrar na encriptação de uma mensagem, embora esta não impeça a interceptação dos dados que trafegam em um canal de comunicação. Ela torna o conteúdo da mensagem inelegível para um humano compreender, portanto, encriptar uma mensagem nos garante a privacidade do conteúdo. Porém, para garantir um sistema de encriptação confiável, ela precisa ser implementada de forma correta como a que discutiremos brevemente.

Este trabalho aborda a teoria que muitos sistemas criptográficos modernos utilizam para implementar seus algoritmos. Entraremos em temas estudados em teoria dos números e curvas elípticas.

Palavras-chaves: teoria dos números, curvas elípticas, criptografia, segurança de dados.

Abstract

Encryption is an indispensable tool for protecting information in computational systems, it is also a fascinating discipline because it crosses several areas of knowledge (science of computing, mathematics, statistics, electrical engineering, physics and communication theory). She has one fundamental function that is to encapsulate messages, so it has a indispensable in a minimally secure communication.

Although there are other factors that make communication secure, let us concentrate on encrypting a message, although this does not prevent the interception of the data that travels on a communication channel. It makes the content of the message ineligible for a human to understand, therefore, to encrypt a message we guarantee the privacy of the content. However, to ensure a system that encryption, it must be implemented in the correct way we will discuss briefly.

This work addresses the theory that many modern cryptographic systems use to implement their algorithms. We will studied in number theory and elliptic curves.

Keywords: number theory, elliptical curves, encryption, data security.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

1.1 Motivação

O matemático alemão Johann Carl Friedrich Gauss (1777−1855) disse: "A matemática é a rainha das ciências, e a teoria dos números é a rainha da matemática" [14], este (teoria dos números) é um ramo da matemática pura dedicado ao estudo das propriedades dos números. Vamos abordar sua aplicação na computação, por exemplo, na criptografia.

Cabe aqui uma curiosidade da origem da expressão *teoria dos números*. O termo mais antigo para se referir a ele é aritmética. No início do século XX, ele foi substituído por "teoria dos números" (a palavra "aritmética", usada pelo público em geral para significar "cálculos elementares", também adquiriu outros significados na lógica matemática, como na aritmética de Peano, e na ciência da computação, como na aritmética de ponto flutuante).

Uma das primeiras obras a compilar resultados aritméticos na história da matemática foi o livro Arithmetica, de Diofanto de Alexandria (nascido entre 201 e 214 — falecido entre 284 e 298). Neste livro, ele trabalha com uma série de problemas cuja resolução busca as soluções racionais para sistemas de

equações polinomiais, geralmente com o padrão $f(x, y) = z^2$ ou $f(x, y, z) = w^2$. Por este motivo, equações polinomiais onde se propõe encontrar soluções racionais ou inteiras são hoje conhecidas como Equações Diofantinas.

Pode-se dizer que Diofanto estudava pontos racionais (ou seja, pontos cujas coordenadas são racionais em curvas e variedades algébricas); entretanto, ao contrário dos gregos do período clássico, que faziam o que agora conhecemos como álgebra básica em termos geométricos, Diofanto fazia o que chamamos de geometria algébrica básica em termos puramente algébricos. Em uma linguagem moderna, o trabalho de Diofanto encontrava parametrizações racionais de variedades, isto é, dada uma equação, por exemplo, seguindo o padrão $f(x_1, x_2, x_3) = 0$, seu objetivo (em essência) era encontrar três funções racionais g_1, g_2 e g_3 tais que, para todos os valores de r e s, de $x_1 = g_1(r, s)$, $x_2 = g_2(r, s)$ e $x_3 = g_3(r, s)$ sejam uma solução de $f(x_1, x_2, x_3) = 0$.

Diofanto também estudou as equações de algumas curvas não-racionais (para as quais nenhuma parametrização racional é possível) e, neste caso, ele conseguiu encontrar alguns pontos racionais sobre essas curvas (o que parece ser a primeira ocorrência das curvas elípticas) por meio de uma construção (não-geométrica) equivalente a traçar uma tangente à curva E em um ponto racional P conhecido e encontrar o outro ponto de intersecção entre a tangente e a curva.

O problema central da geometria diofantina era determinar se uma equação diofantina tinha soluções e quantas eram. A abordagem adotada foi pensar nas soluções de uma equação como um objeto geométrico.

Por exemplo, uma equação em duas variáveis define uma curva no plano. Mais geralmente, uma equação, ou sistema de equações, em duas ou mais variáveis define uma curva, uma superfície ou algum outro objeto no espaço n-dimensional. Na geometria diofantina, pergunta-se se existem pontos ra-

cionais (pontos cujas coordenadas são racionais) ou pontos inteiros (pontos cujas coordenadas são inteiras) na curva ou na superfície. Se houver algum desses pontos, o próximo passo é perguntar quantos existem e como são distribuídos. Duas questões básicas nessa direção é: *há uma quantidade finita ou infinita de pontos racionais em uma dada curva (ou superfície)? E quanto aos pontos inteiros?*

Um exemplo aqui pode ser útil. Considere a equação de Pitágoras $x^2 + y^2 = 1$; gostaríamos de estudar suas soluções racionais, isto é, suas soluções (x, y) tais que x e y são ambos racionais. Isto é o mesmo que pedir por todas as soluções inteiras para $a^2+b^2 = c^2$, pois qualquer solução da última equação nos dá como resultado $x = a/c$, $y = b/c$ para a primeira. É também o mesmo que pedir todos os pontos com coordenadas racionais na curva descrita por $x^2 + y^2 = 1$. (Esta curva é um círculo de raio 1 em torno da origem.)

Este exemplo ilustra a importância do trabalho de Diofanto de Alexandria em seu livro, Arithmetica. Para determinar se uma equação diofantina tinha soluções e quantas eram, a estratégia adotada por ele era pensar nas soluções de uma equação como um objeto geométrico, por exemplo, curvas elípticas.

Cabe aqui sinalizar que as curvas elípticas são especialmente relevantes na teoria dos números e constituem uma área importante da pesquisa acadêmica atual; por exemplo, elas também foram usadas na prova, por Andrew Wiles [20], do último Teorema de Fermat, elas também podem ser usadas em criptografia, a Criptografia de Curvas Elípticas (CCE), que é uma aproximação para criptografia de chaves públicas com base na estrutura algébrica de curvas elípticas sobre corpos finitos. Neste trabalho vamos nos concentrar nesta aplicação das curvas elípticas e também nos fundamentos de teoria dos números e das curvas elípticas, por este motivo, apesar da abordagem conceitual dos últimos parágrafos, para prosseguir no trabalho, será necessário

uma definição forma, que faremos a seguir.

Uma *curva elíptica* é uma curva algébrica plana definida por uma equação da forma $y^2 = x^3 + ax + b$, e não singular; isto é, que não possui singularidade ou auto-interseções, por exemplo, seja $a = -1$, $b = 1$, então temos a seguinte curva elíptica:

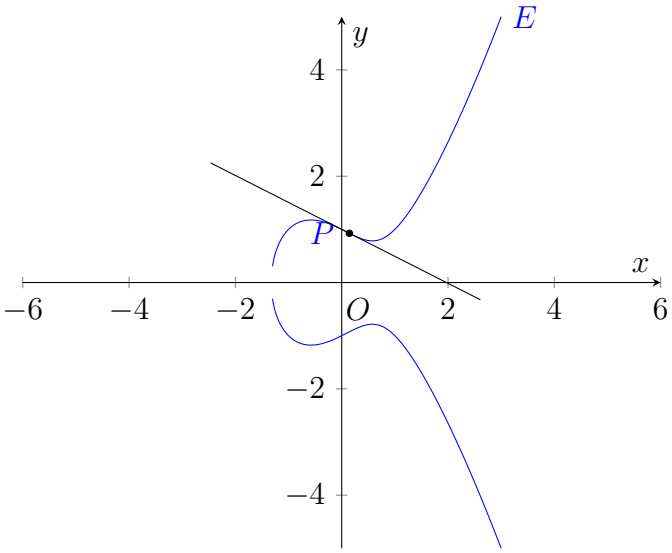


Figura 1.1: Exemplo de curva elíptica $y^2 = x^3 - x + 1$

Se $y^2 = f(x)$, onde f é qualquer polinômio de grau três em x sem raízes repetidas, o conjunto solução é uma curva plana não-singular, uma curva elíptica. Se f tem grau quatro e é livre de quadratura, esta equação descreve novamente uma curva plana; no entanto, não tem escolha natural de elemento identidade. Mais genericamente, qualquer curva algébrica, por exemplo, a curva obtida a partir da intersecção de duas superfícies quadráticas embutidas no espaço projetivo tridimensional, é chamada de curva elíptica, desde que tenha pelo menos um ponto racional para atuar como a identidade.

1.2 Objetivos

Buscaremos abordar conceitos relacionados à Criptografia de Curvas Elípticas e estudar o algoritmo utilizado para gerar chaves públicas baseadas na estrutura algébrica de curvas elípticas sobre corpos finitos. Esta criptografia requer chaves menores em comparação com a criptografia não-CCE (com base em corpos de Galois simples) para fornecer uma segurança equivalente.

As possibilidades de aplicação das curvas elípticas a criptografia são várias, como por exemplo, para assinaturas digitais e geradores pseudo-aleatórios. Indiretamente, elas também podem ser usadas para encriptação, combinando a chave pública com um algoritmo de chave simétrica.

Provavelmente, o termo criptografia é conhecido pelo leitor devido a sua utilização nos mais variados contextos. Geralmente, o termo criptografia se refere aos processos matemáticos de tornar uma mensagem impossível de ser lida, a não ser pela pessoa que tem a chave para "desencriptar", fazendo com que o texto se torne legível novamente.

A criptografia se expandiu para além de simples mensagens secretas; hoje, você pode usar a criptografia para propósitos mais elaborados, como, por exemplo, para identificar o autor das mensagens.

A criptografia é a melhor tecnologia que nós temos para proteger informações, por exemplo, bancos online usam esta técnica para "esconder" os dados de seus usuários de terceiros, e provedores de serviços de internet fazem o mesmo, aliás, estes serviços não conseguiriam garantir o mínimo de segurança para os dados dos seus usuários sem a criptografia. Ela se desenvolveu de tal maneira que é praticamente impossível quebrar o código − quando utilizada de forma correta.

1.3 Organização do texto

O restante desta monografia está organizado da seguinte forma. O capítulo 2 apresenta um pouco sobre a trilha histórica da criptografia e também os conceitos fundamentais sobre criptografia simétrica e assimétrica. O capítulo 3 e capítulo 4 descrevem os fundamentos de teoria dos números, que servem de base para compreender os conceitos de criptografia moderna. O capítulo 5 descreve o conceito de curvas elípticas. Por fim, falamos no capítulo 6 sobre a aplicação dos conceitos apresentados nos capítulos anteriores à CCE.

Capítulo 2

Conceitos de Criptografia

Neste capítulo, apresentamos conceitos fundamentais da criptografia e suas aplicações. Tendo isto em mente, esta parte discute brevemente a história da criptografia, abordando desde seus primórdios até algumas das avançadas técnicas existentes atualmente. Sendo assim, deve ficar claro para o leitor a diferença entre **criptografia simétrica** e **assimétrica**, também como a forma pela qual a utilização *destas técnicas* é capaz de garantir o estabelecimento de uma comunicação segura, este capítulo foi retirado quase que, literalmente, da dissertação de mestrado de Marcos Simplicio [8] embora tenham sido feitas melhorias e ajustes necessários conforme o trabalho de David Kahn (1996) [11].

2.1 Breve história da criptografia

A história da criptografia está intimamente ligada à da escrita. Os primeiros traços de sua utilização datam de 2000 A.C., no Egito. Durante séculos, a criptografia se baseou em dois grandes princípios de cifras:

- **Substituição:** Em um alfabeto, pode-se trocar um conjunto de letras entre si. O mais famoso exemplo desta técnica de codificação é, provavelmente, o alfabeto de César, que consiste na substituição de cada letra de um alfabeto, por exemplo o alfabeto latino, pela i-ésima letra em um alfabeto cíclico (a letra "x" sendo seguida pela letra "y", esta sendo seguida pela letra "z', então retornando para "a"), ou seja, substituímos a letra "d" por "a" do alfabeto cíclico, "e" deste mesmo alfabeto por "b" e assim por diante, chegando até a substituição de "a" por "x", "b" por "y" e "c" por "z", havendo assim um número de possibilidades de substituições igual ao número de letras do alfabeto cíclico, por esta razão ele é chamado de alfabeto de César e chamá-lo cifra de César é um erro, porque para ser considerada uma cifra, a troca entre as letras precisa ser feita de maneira aleatória e não dependendo da sequência alfabética. Um aprimoramento do alfabeto de César é a cifra de Vigenère (do século XVI, Roma), que consiste na separação do alfabeto em n conjuntos, cada um utilizando uma chave diferente. Desta maneira, cada conjunto respeita uma chave de substituição. Por exemplo, definimos duas chaves K_1 e K_2, as letras localizadas em uma posição ímpar do alfabeto serão enviadas para K_1 posições adiante, enquanto que a chave K_2 será utilizada para as demais.

  Saindo do período renascentista, entrando no século XIV, temos a criação das máquinas de rotores que utilizam a técnica de substituição, como por exemplo, a máquina de Herbern (utilizando apenas um rotor) e a máquina Enigma, criada pela força militar alemã.

- **Permutação:** modifica-se a ordem das letras do texto claro, ou seja, que possa ser compreendido por um humano. O mais antigo exemplo de utilização desta técnica é atribuído aos Espartanos, na antiga Grécia, que utilizavam-se de um bastão conhecido pelo nome de *scytale* [1] ou *skytale*. O processo de codificação consistia em enrolar uma faixa de pergaminho no *skytale*, escrevendo então a mensagem clara sobre este pergaminho. Para decifrar o texto, basta enrolar a faixa em um outro *skytale* de mesmo tamanho, alinhando o texto de forma a torná-lo legível.

Hoje, no contexto dos documentos digitais, as "palavras" de uma mensagem foram substituídas pelos seus bits ou bytes. Entretanto, as técnicas de substituição e permutação continuam válidas perfeitamente.

Além desta breve explicação sobre estes dois conceitos (substituição e permutação), cabe aqui destacar como o trabalho *Communication Theory of Secrecy Systems* de Claude Shannon (1949) [17] foi importante na área da Teoria da Informação.

Ele enuncia neste trabalho o que hoje é chamado de **Segredo Perfeito:** a probabilidade de se obter um texto claro a partir de quaisquer conjuntos de textos cifrados deve ser a mesma, ou seja, a dificuldade de se obter textos claros deve ser independente do número de textos cifrados de que dispõe um atacante. Ele também mostra que a combinação da mensagem original com uma chave completamente aleatória de mesmo tamanho satisfaz este requisito, resultado em uma cifra de segurança máxima, denominada **One-Time Pad**. No entanto, como as chaves somente podem ser utilizadas uma única vez e são potencialmente grandes, a utilização de tal cifra na maioria

[1] Em criptografia, *scytale* é uma ferramenta usada para executar uma cifra de transposição, consistindo de um cilindro com uma tira de pergaminho enrolada em torno dele sobre a qual está escrita uma mensagem.

das aplicações práticas é inviável.

Com trabalhos científicos como este de Shannon e também com o desenvolvimento dos computadores e de redes de telecomunicação ao longo das últimas décadas, a criptografia deixou de ser restrita aos meios diplomático e militar, permitindo a criação de diversas cifras e também maneiras de ataques às mesmas. É desta forma que, em 1975, foi promovido um concurso semi-público para a escolha de um padrão criptográfico destinado a aplicações civis. O vencedor do concurso foi um algoritmo proposto pela IBM e alterado pela Agência de Segurança Nacional americana (*Natinal Security Agency* − NSA) segundo critério não divulgados na época, dando origem ao *Data Encryption Standard* − DES (NIST, 1977) [7]. O projeto da cifra baseia-se na chamada Estrutura de Feistel, sendo efetivamente adotado como padrão de codificação pelo governo americano em 1977. Porém, o DES é atualmente considerado obsoleto, principalmente, pelo seu reduzido tamanho de chave (56 bits), pouco segura se considerada a capacidade computacional disponível atualmente, e pelo seu pequeno tamanho de bloco (64 bits), que pode facilitar ataques futuros.

Apesar de o DES ter sido, definitivamente, aposentado pelo Instituto Nacional de Padrões e Tecnologia Americano (*Nacional Institute of Standard and Technology* − NIST) em 2004, foi criada uma solução paliativa para aproveitar a ampla base existente de implementações desta cifra em hardware e software, conhecida como 3-DES (ou DES Triplo). Ele consiste em efetuar por 3 vezes a codificação usando DES simples, com 3 chaves distintas de 56 bits, K_1, K_2 e K_3 resultando em uma segurança, aproximadamente, equivalente à de uma cifra de 112 bits. Existe ainda uma variante bastante popular que consiste em utilizar a chave K_2 com o algoritmo de decodificação, de forma que o 3-DES executa a sequência codificação-decodificação-codificação (CDC).

Apesar de esta construção não ser diferente em termos de segurança, ela é interessante pela possibilidade de simular o DES simples usando o 3-DES com 3 chaves.

O padrão de codificação atual foi definido em outubro do ano 2000, após um novo concurso (desta vez verdadeiramente público) patrocinado pelo NIST, que teve início oficial em junho de 1997. Dentre os 5 finalistas, o algoritmo Rijindael (Daemen; Rijmen, 2002) [3], desenvolvido pelos criptógrafos belgas Joan Daemen e Vincent Rijmen, foi escolhido como o *Advanced Encryption Standard* − AES (NIST, 2001) [19]. O Rijindael foi desenvolvido a partir de um outro algoritmo chamado de SQUARE (Daemen; Knudsen; Rijmen, 1997).

2.2 Segurança na Comunicação

O termo Criptografia (**do grego**: *kryptós*,"esconder", e *gráphein*, "escrever") é, historicamente, associado à arte de "esconder informações" ou "encapsular informação", termo que pode ser interpretado como a capacidade de fornecer confidencialidade à informação.

Resumidamente, pode-se então definir um algoritmo criptográfico (ou **cifra**) como uma *função reversível* que transforma **textos claros** (também denominados **mensagens claras**) P em **textos cifrados** (ou **mensagens cifradas**) C e vice-versa, utilizando, no processo, uma ou mais *chaves criptográficas* K.

Portanto, supondo que duas pessoas estão trocando textos, se Alice quer enviar uma mensagem a Bob (seguindo a tradição de nomes, frequentemente, adotados no universo da Criptografia) de forma segura, mesmo que o canal de comunicação usado seja, caracteristicamente, inseguro como é o caso da

Internet, ela irá codificar sua mensagem usando uma *função de codificação* E e uma *chave de codificação* (K_e).

Para ser capaz de decifrar e, portanto, de ler a mensagem de Alice, Bob irá utilizar uma *função de decodificação* D e também uma *chave de decodificação* (K_d). Um terceiro indivíduo que não tenha acesso a estas chaves deve então ser incapaz de espreitar o diálogo.

Apesar da confidencialidade ser uma das principais áreas de aplicação da criptografia, atualmente, esta apresenta um campo mais amplo, respondendo às seguintes necessidades:

- **Confidencialidade:** garantir que as mensagens trocadas poderão ser compreendidas somente pelos usuários desejados, de tal forma que apenas eles sejam capazes de extrair a informação nela contida.
- **Integridade:** possibilidade de verificar a consistência da informação contida em uma mensagem. Tal serviço não garante que as mensagens não sejam alteradas durante a transmissão, mas sim que a ocorrência da alteração possa ser detectada.
- **Autenticação:** possibilidade de comprovar a identidade de um indivíduo que participa da comunicação.
- **Irretratabilidade:** garantia de que nem o remetente nem o destinatário de uma determinada mensagem possam negar sua transmissão, recepção ou posse.

2.3 Dois métodos para trocar informações

Formalmente, os processos de codificação e decodificação podem ser definidos da seguinte forma: Seja $\mathcal{P}$ um conjunto de textos claros, $\mathcal{C}$ um conjunto de

textos cifrados e $\mathcal{K}$ um conjunto de chaves. Uma função de codificação E associa a cada par $(P, K_e) \in \mathcal{P} \times \mathcal{K}$ um elemento $C = E(P, K_e) \in \mathcal{C}$, para obtermos a mensagens claras novamente usamos a função de decodificação D, de forma que $D(E(P, K_e), K_d) = P$.

2.3.1 Criptografia Simétrica

Até o ano de 1976, a única forma conhecida de criptografia era a **criptografia de chave secreta ou criptografia simétrica**: para que dois indivíduos pudessem se comunicar de forma segura, ambos precisavam compartilhar uma mesma chave secreta para codificação e decodificação, conhecida apenas por eles. Isto gerava problemas, principalmente, com relação à distribuição destas chaves, que precisavam ser trocadas através de um meio seguro antes que fosse possível utilizar qualquer forma de criptografia. Portanto, se Alice e Bob decidem utilizar um esquema de codificação simétrica para se comunicar, eles devem partilhar uma mesma chave K, conhecida apenas por eles, pelo menos na teoria porque estamos descartando a possibilidade de um terceiro indivíduo ter acesso a estas chaves. Esta chave será usada tanto na operação de codificação quanto decodificação, ou seja, $K_e = K_d = K$. A figura 2.1 ilustra o processo. [2]

[2] P_{Alice} é a mensagem clara de Alice e P_{Bob} é a mensagem clara para Bob. Além disso, C é mensagem cifrada, E_{Alice} é função de codificação do texto (ou mensagem) de Alice e D_{Bob} é a função de decodificação do texto de Alice por Bob.

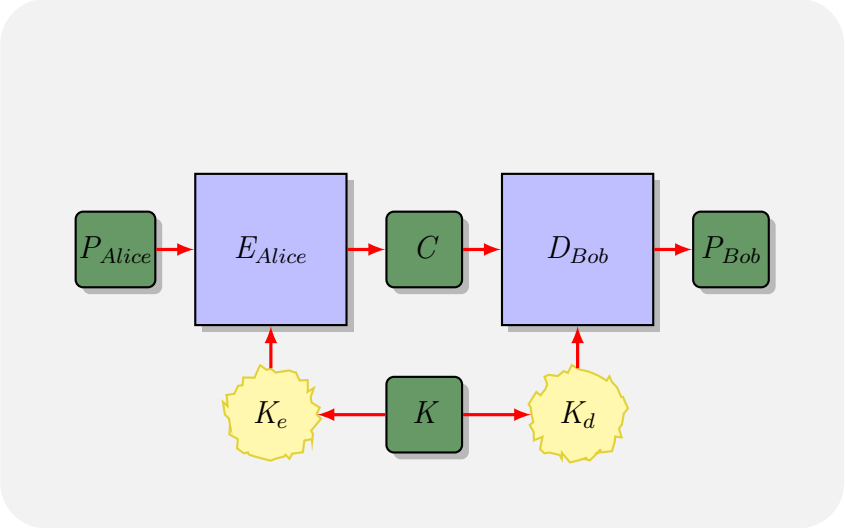


Figura 2.1: Criptografia Simétrica

O AES e o DES são dois exemplos de algoritmos que adotam o esquema de criptografia simétrica.

2.3.2 Criptografia Assimétrica

Apesar de algumas fontes (Williamson, 1976) [9] alegarem que este tipo de criptografia já era conhecido no meio militar, foi apenas em 1976 que os criptógrafos ingleses Diffie e Hellman (Diffie; Hellman, 1976) [4] apresentaram ao meio civil o esquema conhecido como **criptografia assimétrica (ou criptografia de chave pública).**

Sua utilização permite que indivíduos estabeleçam uma comunicação segura sem a necessidade de um compartilhamento prévio de chave criptográfica simétrica.

Portanto, são usadas duas chaves diferentes para codificação e decodificação: uma *chave pública* K_u e sua correspondente *chave privada* K_r. Quando

Alice deseja enviar uma mensagem confidencial P a Bob, ela deve codificar a mensagem usando a chave pública de Bob, que pode ser encontrada abertamente na Internet ou junto a uma entidade com esta atribuição (Entidade Certificadora), por exemplo. Desta forma, é gerada uma mensagem cifrada C que apenas Bob é capaz de decifrar, posto que ele é o único que conhece sua chave privada. A figura 2.2 ilustra este processo. [3]

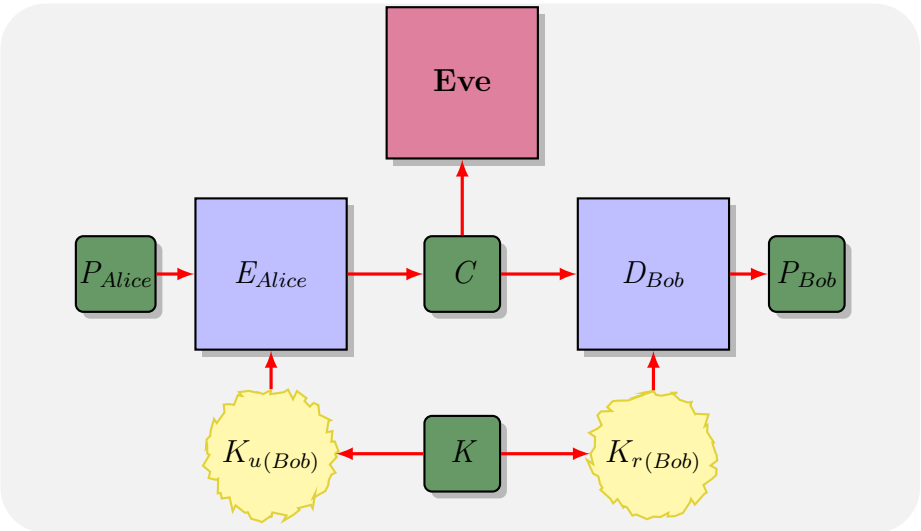


Figura 2.2: Criptografia Assimétrica

Um dos exemplos mais conhecidos deste tipo de cifra é o algoritmo RSA (Rivest, Shamir, Adelmen; 1977) [19], cuja segurança baseia-se na dificuldade computacional de se fatorar números grandes.

[3] P_{Alice} é a mensagem clara de Alice e P_{Bob} é a mensagem clara para Bob. Além disso, C é mensagem cifrada, E_{Alice} é função de codificação do texto (ou mensagem) de Alice e D_{Bob} é a função de decodificação do texto de Alice por Bob. Quando Alice assina uma mensagem usando sua chave privada, Bob pode ter certeza que é com ela (e não com Eve, por exemplo), que ele está iniciando uma conversa.

2.3.3 Simétrica Vs Assimétrica

Existe um grande interesse na utilização de criptografia de chave pública, pela sua capacidade de prover diversas funcionalidades essenciais no domínio da segurança da informação. Dentre as suas aplicações, destacam-se: assinatura digital, autenticação de usuários e distribuição de chaves.

Uma assinatura digital é uma forma de comprovar a autoria de uma mensagem ou documento. Sucintamente, o procedimento para sua geração consiste no cálculo de um resumo criptográfico de tamanho fixo (também conhecido como *hash*) da mensagem que se deseja assinar. Isto pode ser feito utilizando, por exemplo, uma função da família SHA (NIST, 2002) [1].

O *hash* é então codificado com a chave privada do autor da mensagem. O resultado é a assinatura digital da mesma: como apenas este indivíduo tem acesso à sua chave privada, apenas ele seria capaz de criar tal assinatura, que pode se verificar por meio da sua chave pública. Assim, caso Bob receba uma mensagem supostamente assinada por Alice, ele pode decodificá-la com a chave pública de Alice e comparar o resultado com o *hash* da mensagem recebida.

Caso os mesmos sejam idênticos, pode-se concluir que aquele documento foi realmente assinado por Alice; caso contrário, é possível que a assinatura não pertença a Alice (por exemplo, Eve pode tê-lo assinado) ou que o documento foi alterado na transmissão (acidental ou intencionalmente). O fato do *hash* da mensagem ser assinado ao invés da mensagem completa está relacionado a questões de desempenho: não importando o tamanho da mensagem, a codificação se dá sobre dados de tamanho fixo e arbitrariamente pequeno.

O processo de autenticação é muito semelhante à assinatura digital, porém, possui uma sutil (e importante) diferença: enquanto a assinatura digital prova que um documento pertence a uma determinada pessoa, mesmo muito tempo

após a geração do mesmo, o processo de autenticação visa à identificação dos interlocutores antes do início da comunicação. Desta forma, quando Alice assina uma mensagem usando sua chave privada, Bob pode ter certeza que é com ela (e não com Eve, por exemplo), que ele está iniciando uma conversa.

Por fim, o compartilhamento das chaves simétricas pode ser feito por Alice e Bob sem a necessidade de uma canal de comunicação seguro: basta Alice enviar para Bob uma mensagem contendo a chave simétrica desejada, codificada com um algoritmo assimétrico e a chave pública de Bob. Desta forma, Bob será o único capaz de recuperar a chave simétrica enviada por Alice.

A razão pela qual o próprio algoritmo de chave pública não é utilizado durante toda a comunicação é simples: desempenho. É uma fato conhecido que algoritmos assimétricos apresentam um desempenho bem inferior àquele dos algoritmos simétricos.

O RSA, por exemplo, é de 100 a 1000 vezes mais lento do que o DES e usa chaves bem maiores, sendo 1024 bits um tamanho bastante comum. Portanto, após as fases iniciais de comunicação usando algoritmos de chave pública, os pares comunicantes, geralmente, passam a utilizar um algoritmo simétrico com a chave compartilhada.

2.4 Resumo

Este capítulo abordou alguns dos conceitos mais básicos e essenciais relativos à área de estudo da criptografia, além de um breve histórico foram enumerados os serviços básicos aos quais a mesma se presta: **confidencialidade**, **integridade**, **autenticação** e **irretrabilidade**. Foi ainda explicada a diferença entre **criptografia simétrica** e **assimétrica**, e como as mesmas são utilizadas.

Capítulo 3

Fundamentos em Teoria dos Números

Neste capítulo e no próximo apresentamos alguns conceitos fundamentais de teoria dos números, mais especificamente, falaremos sobre divisibilidade e algoritmo de divisão, algoritmo de euclides, aritmética modular, e no próximo capítulo falaremos sobre grupos, anéis e corpos, temas usados para compreender **cifras simétricas**, além disso, falaremos sobre teorema fundamental da aritmética, Fermat, Euler e Chinês do Resto, temas fundamentais para entender **cifras assimétricas**.

O leitor deve conhecer estes assuntos para se aprofundar no trabalho, mais detalhes podem ser encontrados nos livros "Understanding Cryptography" [5], "Cryptography and Network Security" [18], "A Course in Number Theory and Cryptography" [13] e "Números: Uma Introdução à Matemática" [15].

3.1 Divisibilidade e Algoritmo de Divisão

Vamos abordar algumas propriedades, definições e teoremas para explicar os conceitos de divisibilidade e algoritmo de divisão.

3.1.1 Princípio da Boa Ordem

Todo subconjunto não vazio $S \in \mathbb{Z}$ de elementos não-negativos possui elemento *minimal*, isto é, o elemento $x \in S$ tal que para todo $y \in S$ temos que $x \leq y$.

3.1.2 Divisibilidade

Uma equação do tipo $a = xb$ pode ou não ter solução em $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$; isto dependerá dos coeficientes a e b da equação. Quando tal solução existe, diz-se que a é **divisível** por b. Mais precisamente:

Definição 1(Divisibilidade). Sejam a e $b \in \mathbb{Z}$, diz-se que b, diferente de zero, **divide** a, ou que b é **divisor** de a ou, ainda, que a é um **múltiplo** de b, se existe um inteiro m tal que $a = mb$.

Usaremos a notação $b \mid a$ para indicar que b divide a. A negação dessa afirmação será indicada por $b \nmid a$. Por outro lado, no que $0 \mid a$ se, e somente se, $a = 0$. Neste caso, o quociente não é único pois $0m = 0$, para todo inteiro m.

Posteriormente, vamos precisar de algumas propriedades simples de divisibilidade para números inteiros, que são as seguintes, lembrando que quaisquer que sejam os números inteiros a, b, c, d, e também assumimos que os divisores são diferentes de zero, valem:

- **(i)** Se $a \mid 1$, então $a = \pm 1$.
- **(ii)** Se $a \mid b$ e $b \mid a$, então $a = \pm b$.
- **(iii)** Todo $b \neq 0$ divide 0.
- **(iv)** Se $a \mid b$ e $b \mid c$, então $a \mid c$.
- **(v)** Se $a \mid b$ e $a \mid c$, então $a \mid (mb + nc)$ $\forall m, n \in \mathbb{Z}$.

3.1.3 Algoritmo da Divisão

Dado qualquer número inteiro positivo b e qualquer inteiro não negativo a, se dividirmos b por a, nós obtemos um quociente inteiro q e um resto r que obedecem o seguinte teorema:

Teorema 1(Algoritmo da Divisão). Sejam a e b inteiros, com $b > 0$. Então, existem q, $r \in \mathbb{Z}$, únicos, tais que $a = bq + r$ e $0 \leq r < b$.

Demonstração. Considere o conjunto $S = \{a - bt \mid t \in \mathbb{Z}, a - bt \geq 0\}$. Suponha $a \geq 0$, neste caso, $a - b0 = a \in S$. Por outro lado, se $a < 0$, $a - ba = a(1 - b) \geq 0$, portanto $a(1 - b) \in S$, logo, $S \neq \emptyset$. Como todo conjunto não vazio de inteiros não negativos contém um mínimo pelo Princípio da Boa Ordem, seja r o menor elemento de S. Por definição, r é da forma $r = a - bq$ para um $q \in \mathbb{Z}$ qualquer, e $r \geq 0$. Além disso, temos que $r < b$, pois, caso contrário, $r - b$ seria um elemento de S menor que r, contradizendo a definição de r; de fato, se $r \geq b$, então nós teríamos $0 \leq r - b = a - b(q + 1)$. Isso prova a existência de r e q. Para provar a unicidade, suponha que $a = bq + r$ e $a = bq' + r'$, onde $0 \leq r < b$ e $0 \leq r' < b$. Então subtraindo estas duas

equações e rearranjar os termos, obtemos:

$$r' - r = b(q - q')$$

Então, $r' - r$ é um múltiplo de b; porém, $0 \leq r < b$ e $0 \leq r' < b$ implica que $|r' - r| < b$, assim sendo, a única possibilidade é $r' - r = 0$. Além disso, $0 = b(q - q_1)$ e $b \neq 0$ isto implica que $q - q' = 0$. □

Agora, vamos definir a *função piso*, *teto* e também o *operador* mod, por fim, definiremos o conceito de *ideal* e refinaremos ainda mais o Teorema 1, estes assuntos servirão de base para nosso trabalho.

Função Piso e **Teto**

Seja $\lfloor \cdot \rfloor$ e $\lceil \cdot \rceil$, respectivamente, a função piso e teto, que são funções de $\mathbb{R}$ para $\mathbb{Z}$. Para $x \in \mathbb{R}$, a função piso de x, $\lfloor x \rfloor$, é o maior número inteiro $m \leq x$; é equivalentemente dizer que $\lfloor x \rfloor$ é o único número inteiro m tal que $m \leq x < m + 1$, ou de outra forma, tal que $x = m + \epsilon$ para algum $\epsilon \in [0, 1[$. Além disso, a função teto de x, $\lceil x \rceil$, é o menor número inteiro $m \geq x$; é equivalentemente dizer que $\lceil x \rceil$ é o único número inteiro m tal que $m - 1 < x \leq m$, ou de outra forma, tal que $x = m - \epsilon$ para alguns $\epsilon \in [0, 1[$.

Significado do Mod

O operador mod é usado neste trabalho e na literatura de duas maneiras diferentes: primeiro, veremos como um *operador*, que, a partir de dois argumentos inteiros, retorna o resto entre eles, e mais tarde o veremos como uma relação de *congruência*. Vou explicar as distinções e definições deles nesta subseção e nas próximas:

Definição 2(Operador Mod). Sejam a, b inteiros, com $b > 0$. Pelo Teorema 1, existem $q, r \in \mathbb{Z}$ únicos que satisfazem $a = bq + r$ e $0 \leq r < b$. Portanto, podemos definir:

$$a \bmod b \coloneqq r;$$

isto é, por $a \bmod b$ denotamos o resto da divisão de a por b. Como já vimos que $b \mid a$ se, somente se, $a \bmod b = 0$. Dividindo ambos os lados da equação $a = bq + r$ por b, obtemos $a/b = q + r/b$. Como $q \in \mathbb{Z}$ e $r/b \in [0, 1[$, temos que $q = \lfloor a/b \rfloor$. Então,

$$a \bmod b = a - b \lfloor a/b \rfloor .$$

Pode-se usar esta equação para estender a definição de $a \bmod b$ para todos os inteiros a e b com $b \neq 0$; para $b < 0$, simplesmente definimos $a \bmod b$ como $a - b \lfloor a/b \rfloor$. Logo, podemos generalizar o Teorema 1 quando dividimos o inteiro a por um inteiro positivo b, o resto desta divisão está em um intervalo diferente de $[0, b[$. Então, seja x qualquer número real, e considere o intervalo $[x, x+b[$. Este intervalo contém precisamente b inteiros, ou seja, $\lceil x \rceil, \ldots, \lceil x \rceil + b - 1$. Portanto, aplicando o Teorema 1 com $a - \lceil x \rceil$ no lugar de a, temos o teorema:

Teorema 2. Sejam $a, b \in \mathbb{Z}$ com $b > 0$, e seja $x \in \mathbb{R}$. Então existem $q, r \in \mathbb{Z}$ únicos, tais que $a = bq + r$ e $r \in [x, x + b[$.

Significado de Ideal

Vamos iniciar agora o conceito de ideal, gostaríamos de alertar para o fato de que a definição clássica de ideal é feita para um anel em geral, porém, como a utilização neste trabalho se restringe a ideias sobre inteiros, vamos restringir, também, as definições para nossas necessidades.

Definição 3(Ideal). Um conjunto não-vazio $I \subseteq \mathbb{Z}$ é chamado de *ideal* se $\forall a, b \in I$ e para todo $z \in \mathbb{Z}$, temos:

$$a + b \in I \quad \text{e} \quad az \in I.$$

É fácil ver que todo ideal I contém 0: como $a \in I$ para algum inteiro a, temos que $0 = a \cdot 0 \in I$, aliás, note que se um ideal I contém um inteiro a, ele também contém $-a$, pois $-a = a \cdot (-1) \in I$. Assim, se um ideal contém a e b, ele também contém $a - b$. É claro que $\{0\}$ e $\mathbb{Z}$ são ideais. Além disso, um ideal I é igual a $\mathbb{Z}$ se somente se $1 \in \mathbb{Z}$; para ver isto, note que $1 \in \mathbb{Z}$ implica que para cada $z \in \mathbb{Z}$, temos $z = 1 \cdot z \in I$, e portanto, $I = \mathbb{Z}$; caso contrário, se $I = \mathbb{Z}$, então em particular, $1 \in \mathbb{Z}$.

Para $a \in \mathbb{Z}$, definimos $a\mathbb{Z} \coloneqq \{az : z \in \mathbb{Z}\}$; isto é, $a\mathbb{Z}$ é conjunto de todos os múltiplo de a. Se $a = 0$, então claramente $a\mathbb{Z} = \{0\}$; de outra forma, $a\mathbb{Z}$ consiste os inteiros distintos:

$$\ldots, -3a, -2a, -a, 0, a, 2a, 3a, \ldots$$

É fácil de ver que $a\mathbb{Z}$ é um ideal: para todo $az, az' \in a\mathbb{Z}$ e $z'' \in \mathbb{Z}$, temos que $az + az' = a(z + z') \in a\mathbb{Z}$ e $(az)z'' \in a\mathbb{Z}$. O ideal $a\mathbb{Z}$ é chamado de *ideal gerado* por a, e um ideal da forma $a\mathbb{Z}$ para algum $a \in \mathbb{Z}$ é chamado de *ideal principal*.

Observe que para todo $a, b \in \mathbb{Z}$, temos que $b \in a\mathbb{Z}$ se, e somente se, $a \mid b$, aliás, note que para cada ideal I, temos que $b \in I$ se, e somente se, $b\mathbb{Z} \subseteq I$. Ambas observações são consequências simples das definições, como o leitor pode verificar. Se combinarmos estas duas observações, nós percebemos que $b\mathbb{Z} \subseteq a\mathbb{Z}$ se, e somente se, $a \mid b$. Vamos supor que I_1 e I_2 são ideais. Então não é difícil perceber que o conjunto

$$I_1 + I_2 \coloneqq \{a_1 + a_2 : a_1 \in I_1, a_2 \in I_2\}$$

é também um ideal. De fato, suponha que $a_1 + a_2 \in I_1 + I_2$ e $b_1 + b_2 \in I_1 + I_2$. Então nós temos que $(a_1 + a_2) + (b_1 + b_2) = (a_1 + b_1) + (a_2 + b_2) \in I_1 + I_2$, e para cada $z \in \mathbb{Z}$, temos que $(a_1 + a_2)z = a_1z + a_2z \in I_1 + I_2$.

Exemplo 1. Considere o ideal principal $3\mathbb{Z}$. Este consiste em todos os múltiplos de 3; isto é, $3\mathbb{Z} = \{\ldots, -9, -6, -3, 0, 3, 6, 9, \ldots\}$.

Exemplo 2. Considere o ideal $3\mathbb{Z}+5\mathbb{Z}$. Este ideal contém $3\cdot 2+5\cdot(-1) = 1$; já que contém 1, ele também contém todos os inteiros; isto é, $3\mathbb{Z} + 5\mathbb{Z} = \mathbb{Z}$.

Neste dois exemplos, mostramos a formação de ideais, observando com mais detalhes, quando definimos um *ideal principal* não foi por acaso: o seguinte Teorema 3 diz que todos os ideais de $\mathbb{Z}$ são principais:

Teorema 3. Seja I um ideal de $\mathbb{Z}$. Então existe um inteiro não-negativo d tal que $I = d\mathbb{Z}$.

Demonstração. Primeiro vamos provar a existência do Teorema 3. Se $I = \{0\}$, então $d = 0$.

Agora, vamos supor que $I \neq \{0\}$. Existe pelo menos um inteiro $a \neq 0$ tal que $a \in I$, então pela Definição 3 (ideal) temos que $-a \in I$. Como a e $-a$ pertencem a I, podemos afirmar que I contém inteiros positivos. Assim, o conjunto $I_+ = \{a \in I : a > 0\}$ é não-vazio. Pelo Princípio de Boa Ordem existe um $d = \min I_+$.

De fato, como $d \in I$, a Definição 3 mostra que, para todo $z \in \mathbb{Z}$, tem-se que $dz \in I$, logo, $d\mathbb{Z} \subset I$. Para provar a inclusão contrária, consideraremos um elemento qualquer $a \in I$ e provaremos que é um múltiplo de d. É suficiente mostrar que $d \mid a$ e podemos determinar q e r tais que $a = dq + r$, tal que $0 \leq r < d$. Se $r \neq 0$, como $r = a - dq$ e tanto a quanto dq pertencem a I, teríamos que $r \in I_+$. Mas, $r < d = \min I_+$, uma contradição. Assim, $r = 0$, logo, $a = dq$ é um múltiplo de d. □

3.2 Algoritmo de Euclides

Uma das técnicas básicas da teoria dos números é o Algoritmo Euclidiano, que é um procedimento simples para determinar o *maior divisor comum* de dois inteiros positivos.

Maior Divisor Comum

Sejam $a, b \in \mathbb{Z}$, denotamos $d \in \mathbb{Z}$ um **divisor comum** de a e b se $d \mid a$ e $d \mid b$; além disso, nós chamamos d de **maior divisor comum** de a e b se d é não-negativo e todos os outros divisores comuns de a e b dividem d.

Teorema 4. Para todo $a, b \in \mathbb{Z}$, existe um maior divisor comum único d de a e b, e além disso, $a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z}$.

Demonstração. Aplicamos o Teorema 3 ao ideal $I := a\mathbb{Z} + b\mathbb{Z}$. Seja $d \in \mathbb{Z}$ com $I = d\mathbb{Z}$. Queremos mostrar que d é o maior divisor comum de a e b. Note que $a, b, d \in \mathbb{Z}$ e d é positivo.

Como $a \in I = d\mathbb{Z}$, nós temos que $d \mid a$; similarmente, $d \mid b$. Então temos que d é um divisor comum de a e b.

Como $d \in \mathbb{Z} = a\mathbb{Z} + b\mathbb{Z}$, existe $s, t \in \mathbb{Z}$ tal que $as + bt = d$. Suponha agora que $a = a'd'$ e $b = b'd'$ para algum $a', b', d' \in \mathbb{Z}$. Então a equação $as + bt = d$ implica que $d'(a's + b't) = d$, diz que $d' \mid d$. Assim, qualquer divisor comum d' de a e b divide d. Isso prova que d é o maior divisor comum de a e b. Por exclusividade, observe que se e é o maior divisor comum de a e b, então $d \mid e$ e $e \mid d$, e portanto, $d = \pm e$; já que ambos d e e são ambos não negativos por definição, logo nós temos que $d = e$. □

Para $a, b \in \mathbb{Z}$, vamos denotar $mdc(a, b)$ como o maior divisor comum de a e b.

Nós dizemos que $a, b \in \mathbb{Z}$ são **primos entre si** se $mdc(a, b) = 1$, que é o mesmo dizer que os únicos divisores comuns de a e b são +1 e -1.

O que se segue é essencialmente apenas uma reafirmação do Teorema 4, vamos enfatizar neste teorema:

Teorema 5. Sejam $a, b, r \in \mathbb{Z}$ e seja $d := mdc(a, b)$. Então existe $s, t \in \mathbb{Z}$ tal que $as + bt = r$ se, e somente se, $d \mid r$. Em particular, a e b são primos entre si se, e somente se, existem inteiros r e t tal que $as + bt = 1$.

Demonstração.

$$\begin{aligned} & as + bt = r \quad \text{para algum} \quad s, t \in \mathbb{Z} \\ \iff & r \in a\mathbb{Z} + b\mathbb{Z} \\ \iff & r \in d\mathbb{Z} \quad \text{(pelo Teorema 4)} \\ \iff & d \mid r. \end{aligned}$$

Isso prova a primeira afirmação. A segunda declaração segue a primeira, definindo $r := 1$. □

Teorema 6. Sejam $a, b, c \in \mathbb{Z}$ tal que $c \mid ab$ e $mdc(a, c) = 1$. Então $c \mid b$.

Proposição 1. Sejam a, b inteiros, $d = mdc(a, b)$ e c um inteiro não nulo. Então:

- **(i)** $mdc(ac, bc) = d|c|$
- **(ii)** Se $c \mid a$ e $c \mid b$, então $mdc(a/c, b/c) = d/|c|$.

Teorema 7(Teorema de Euclides). Sejam a, b, $c \in \mathbb{Z}$ tais quer $a \mid bc$. Se $mdc(a, b) = 1$, então $a \mid c$

Demonstração. Supondo que $a \mid bc$ e $mdc(a, b) = 1$. Então pelo Teorema 5 nós temos que $as + bt = 1$ para algum $s, t \in \mathbb{Z}$. Multiplicando por c ambos os lados desta equação, nós temos que:

$$acs + bct = c. \tag{3.1}$$

Como a divide acs por hipótese, e também a divide bct, então dividindo a por ambos os lodos da equação 3.1, temos que $a \mid |c|$, logo $a \mid c$. □

Encontrando o Maior Divisor Comum

Descrevemos agora um algoritmo creditado a Euclides para encontrar facilmente o maior divisor comum de dois inteiros. Este algoritmo tem significado subsequente neste capítulo. Suponha que temos inteiros a, b tais que $d = mdc(a, b)$. Porque $mdc(|a|, |b|) = mdc(a, b)$, podemos assumir que $a \geq b > 0$. Agora dividindo a por b e aplicando o algoritmo de divisão, nós podemos concluir:

$$a = q_1 b + r_1 \quad \text{como} \quad 0 \leq r_1 < b \tag{3.2}$$

Se acontecer que $r_1 = 0$, então $b \mid a$ e $d = mdc(a, b) = b$. Mas se $r_1 \neq 0$, podemos afirmar que $d \mid r_1$. Isto é devido às propriedades básicas da divisibilidade: as relações $d \mid a$ e $d \mid b$ juntas implicam que $d \mid (a - q_1 b)$, que é o mesmo que $d \mid r_1$. Antes de prosseguir com o Algoritmo Euclidiano, precisamos responder a pergunta: O que é o $mdc(b, r_1)$? Nós sabemos que $d \mid b$ e $d \mid r_1$. Agora pegue qualquer inteiro c arbitrário que divida tanto b como r_1. Portanto, $c \mid (q_1 b + r_1) = a$. Porque c divide tanto a quando b, devemos ter $c \leq d$, que é o maior divisor comum de a e b. Portanto $d = mdc(b, r_1)$.

Vamos agora retornar à equação (3.2) e assumir que r_1 é diferente de 0. Porque $b > r_1$, podemos dividir b por r_1 e aplicar o algoritmo de divisão para obtemos:

$$b = q_2 r_1 + r_2 \quad \text{como} \quad 0 \leq r_2 < r_1$$

Como foi dito antes, se $r_2 = 0$, então $d = r_1$ e se $r_2 \neq 0$, então $d =$

$mdc(r_1, r_2)$. O processo de divisão continua até que o resto seja igual a zero, vamos ver isto no esquema (3.3), digamos, no $(n+1)$ passo no qual r_{n-1} é dividido por r_n.

O resultado é o seguinte:

$$\left.\begin{array}{rcrl} a &=& q_1 b + r_1, & 0 \leq r_1 < b; \\ b &=& q_2 r_1 + r_2, & 0 \leq r_2 < r_1; \\ r_1 &=& q_3 r_2 + r_3, & 0 \leq r_3 < r_2; \\ \vdots & & & \\ r_{n-2} &=& q_n r_{n-1} + r_n, & 0 \leq r_n < r_{n-1}; \\ r_{n-1} &=& q_{n+1} r_n + 0, & \\ d &=& mdc(a,b) = r_n. & \end{array}\right\} \quad (3.3)$$

Em cada iteração, temos $d = mdc(r_i, r_{i+1})$ até que finalmente $d = mdc(r_n, 0) = r_n$. Assim, podemos encontrar o maior divisor comum de dois inteiros por repetitiva aplicação do algoritmo de divisão.

Argumentamos essencialmente de cima para baixo que o resultado final é o $mdc(a, b)$. Nós também podemos argumentar de baixo para cima. O primeiro passo é mostrar que r_n divide a e b. Segue-se da última divisão na equação (3.3) que r_n divide r_{n-1}. A penúltima divisão mostra que r_n divide r_{n-2} porque divide os dois termos à direita. Sucessivamente, vê-se que r_n divide todos os r_i termos e finalmente a e b. Resta mostrar que r_n é o maior divisor que divide a e b. Se pegarmos qualquer inteiro arbitrário que divida a e b, ele também deve dividir r_1, como explicado anteriormente. Podemos seguir a sequência de equações na equação (3.3) abaixo e mostrar que c deve dividir todos os r_i. Portanto, c deve dividir r_n, de modo que $r_n = mdc(a, b)$.

Vamos agora olhar como seria o pseudocódigo deste algoritmo:

Algoritmo 1 Algoritmo de Euclides

```
1: procedure EUCLIDES(a, b)                    ▷ O mdc de a e b
2:     r ← a mod b
3:     while r ≠ 0 do                  ▷ Nós temos a resposta se r é 0
4:         a ← b
5:         b ← r
6:         r ← a mod b
7:     end while
8:     return b                                 ▷ O mdc é b
9: end procedure
```

3.3 Aritmética Modular

Quase todos os algoritmos de criptografia, tanto de **cifras simétricas** quanto **assimétricas** são baseados em aritmética com um número finito de elementos. Por isso, nós vamos introduzir o conceito de aritmética modular, que é uma maneira simples de executar aritmética em um conjunto finito de inteiros.

Congruência

Definição 4. Seja $m \neq 0$ um inteiro fixo. Dois inteiros a e b dizem-se *congruentes* módulos m, se m divide a diferença $a - b$.

Portanto, pela definição acima dois inteiros a e b dizem-se congruentes módulos m, se $a \bmod m = b \bmod m$, escrevemos $a \equiv b \pmod{m}$ [1]. Para indicar que a e b não são congruentes módulo m, escrevemos $a \not\equiv b \pmod{m}$.

[1] Usamos o operador mod de duas maneiras diferentes: primeiro como um **operador** que, a partir de dois argumentos inteiros, retornar o resto entre eles, como na expressão $a \bmod b$ (veja Definição 2); segundo como uma **relação de congruência** que mostra a equivalência de dois inteiros, como $a \equiv b \pmod{m}$.

O número inteiro m é chamado de *módulo* e Gauss escreveu no seu livro, Disquisitiones Arithmeticae, que o símbolo $\equiv$ foi induzido pelo símbolo de igualdade, já que são conceitos semelhantes.

Com a nossa definição, $a \equiv b \pmod{m}$ se, e somente se, $m \mid (a-b)$, ou, equivalentemente, se houver um inteiro q tal que $a = b + mq$.

Como $m \mid (a-b)$ se, e somente se, $|m| \mid (a-b)$, nós nos limitaremos a considerar o caso em que $m > 0$.

Proposição 2. Seja m um inteiro fixo. Dois inteiros a e b são *congruentes módulo* m se, e somente se, eles têm como resto o mesmo inteiro quando dividimos por m.

Propriedades de Congruências

Sejam $m > 0$ um inteiro fixo, e $a, b, c, d \in \mathbb{Z}$. Então, valem as seguintes propriedades:

- **(i)** $a \equiv a \pmod{m}$.
- **(ii)** Se $a \equiv b \pmod{m}$, então $b \equiv a \pmod{m}$;
- **(iii)** Se $a \equiv b \pmod{m}$ e $b \equiv c \pmod{m}$, então $a \equiv c \pmod{m}$.
- **(iv)** Se $a \equiv b \pmod{m}$ e $c \equiv d \pmod{m}$, então $a+c \equiv b+d \pmod{m}$.
- **(v)** Se $a \equiv b \pmod{m}$, então $a + c \equiv b + c \pmod{m}$.
- **(vi)** Se $a \equiv b$ e $c \equiv d \pmod{m}$, então $ac \equiv bd \pmod{m}$.
- **(vii)** Se $a \equiv b \pmod{m}$, então $a^n \equiv b^n \pmod{m}$, para todo inteiro positivo n.

- **(viii)** Se $m \mid (a-b)$, então $a \equiv b \pmod{m}$

Demonstração. Para provar **(i)**, observamos que m divide $0 = a - a$, logo $a \equiv a \pmod{m}$. Para provar **(ii)**, notamos que m divide $a-b$, então também divide $-(a-b) = b-a$, logo $b \equiv a \pmod{m}$. Para **(iii)**, note também que se m divide $a-b$ e $b-c$, então também divide $(a-b)+(b-c) = a-c$, logo $a \equiv c \pmod{m}$. A demonstração de **(iv)** é análoga à anterior, e **(v)** segue de **(iv)**, observando por **(i)** que $c \equiv c \pmod{m}$. Para provar **(vi)**, mudaremos levemente a estratégia. Se $a \equiv b \pmod{m}$ e $c \equiv d \pmod{m}$, existem inteiros q_1 e q_2 tais que $a = b+q_1m$ e $c = d+q_2m$, logo $ac = bd+(bq_2+dq_1+q_1q_2m)m$, isto é, $m \mid (ac-bd)$, donde $ac \equiv bd \pmod{m}$. Novamente, **(vii)** segue de **(vi)**, tomando-se $c = a$, $d = b$ e usando a indução em n. Finalmente, para demonstrar **(viii)**, observamos que, se $m \mid (a-b)$, temos diretamente que, $m \mid ((a)-(b))$, então, $a \equiv b \pmod{m}$. □

Teorema 8. Sejam $a, m \in \mathbb{Z}$ com $m > 0$. Então existe um único inteiro a tal que $z \equiv a \pmod{m}$ e $0 \leq z < m$, isto é, $z := a \bmod m$. Mais geralmente, para cada $x \in \mathbb{R}$, existe um único inteiro $z \in [x, x+m[$ tal que $z \equiv a \pmod{m}$.

Vejamos alguns exemplos:

1. Vamos encontrar o conjunto de soluções a para a congruência:

$$3a + 4 \equiv 6 \pmod{7} \tag{3.4}$$

Suponto que a é uma solução da Equação (3.4). Subtraindo 4 de ambos os lados desta equação, nós obtemos:

$$3a \equiv 2 \pmod 7 \tag{3.5}$$

Em seguida, gostaríamos de dividir ambos os lados dessa congruência por 3, para obter a. Nós não podemos fazer isso diretamente, porém, desde que $3 \cdot 5 \equiv 1 \pmod 7$, podemos alcançar o mesmo efeito multiplicando ambos os lados em (3.5) por 5. Se fizermos isso obtemos:

$$a \equiv 3 \pmod 7$$

Assim, se a é uma solução para (3.4), então devemos ter $a \equiv 3 \pmod 7$; Por outro lado, pode-se verificar que, se $a \equiv 3 \pmod 7$, então (3.4) é válido. Concluímos que os inteiros a que são soluções para (3.4) são precisamente aqueles inteiros que são congruentes a 3 (mod 7), que podem ser listados:

$$\ldots, -18, -11, -4, 3, 10, 17, 24, \ldots$$

2. Considere as horas de um relógio analógico. Se você anotar o horário a cada uma hora, você obtém:

$$1h, 2h, 3h, \ldots, 11h, 12h, 1h, 2h, 3h, \ldots, 11h, 12h, 1h, 2h, 3h, \ldots$$

Mesmo se continuássemos adicionando, hora a hora, esta nesta lista, nós nunca saíremos dela. Com este exemplo, podemos agora abordar de uma formar mais geral como funciona aritmética modular (também

chamada aritmética do relógio).

3. Considere o conjunto dos doze números:

$$\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$$

Nós podemos fazer aritmética regular, desde que os resultados sejam menores do que 12. Por exemplo:

$$2 \cdot 3 = 6$$
$$4 + 4 = 8$$

Porém se somarmos dois elementos deste conjunto, por exemplo, $10 + 4 = 14$ e se dividirmos esta soma por 12. E consideramos **apenas o resto** desta divisão temos que:

$$14 \equiv 2 \pmod{12}.$$

Há algumas implicações a partir desta definição que vão além da regra casual "dividir pelo módulo e considerar o resto", nós abordaremos estas implicações brevemente.

Operações com Aritmética Modular

Note que, pelo Teorema 1, o operador (mod m) mapeia todos os inteiros, para o conjunto $\{0, 1, \ldots, (m-1)\} \in \mathbb{Z}$. Isso sugere a seguinte pergunta: Podemos realizar operações aritméticas dentro dos limites deste conjunto? A resposta é sim, e as operações possuem as propriedades esperadas como veremos abaixo. Esta técnica é conhecida como **aritmética modular**:

- **1.** $(a \bmod m + b \bmod m) \bmod m = (a + b) \bmod m$
- **2.** $(a \bmod m - b \bmod m) \bmod m = (a - b) \bmod m$
- **3.** $(a \bmod m \cdot b \bmod m) \bmod m = (a \cdot b) \bmod m$

Demonstração. Nós vamos demonstrar a primeira, as outras duas são análogas. Seja $a \bmod m = r_a$ e $b \bmod m = r_b$. Então nós podemos escrever $a = r_a + jm$ para um inteiro j e $b = r_b + km$ para um inteiro k. Como

$$\begin{aligned}(a + b) \bmod m &= (r_a + jm + r_b + km) \bmod m \\ &= (r_a + r_b + (k + j)m) \bmod m \\ &= (r_a + r_b) \bmod m \\ &= (a \bmod m + b \bmod m) \bmod m\end{aligned}$$

□

Vamos observar um exemplo destas operações na Tabela 3.1 logo abaixo:

+	0	1	2	3	4	5	6	7
0	**0**	1	2	3	4	5	6	7
1	1	2	3	4	5	6	7	**0**
2	2	3	4	5	6	7	**0**	1
3	3	4	5	6	7	**0**	1	2
4	4	5	6	7	**0**	1	2	3
5	5	6	7	**0**	1	2	3	4
6	6	7	**0**	1	2	3	4	5
7	7	**0**	1	2	3	4	5	6

(a) Adição

·	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	**1**	2	3	4	5	6	7
2	0	2	4	6	0	2	4	6
3	0	3	6	**1**	4	7	2	5
4	0	4	0	4	0	4	0	4
5	0	5	2	7	4	**1**	6	3
6	0	6	4	2	0	6	4	2
7	0	7	6	5	4	3	2	**1**

(b) Multiplicação

w	$-w$	w^{-1}
0	**0**	–
1	**7**	**1**
2	**6**	–
3	**5**	**3**
4	**4**	–
5	**3**	**5**
6	**2**	–
7	**1**	**7**

(c) Inversa

Tabela 3.1: Aritmética Módulo 8.

Olhando para a adição, os resultados são diretos e há um padrão regular na matriz. Ambas as matrizes, (a) e (b), são simétricas em relação à diagonal principal, em conformidade com a propriedade comutativa de adição e

multiplicação. Aliás, existe um inverso aditivo, ou negativo, para cada inteiro na aritmética modular. Para achá-lo varre-se a linha correspondente da matriz (a) para encontrar o valor 0, o inteiro no topo dessa coluna é o inverso (pois o inverso aditivo é, por definição, aquele cuja soma resulta em 0); assim $(2 + 6) \bmod 8 = 0$. Isso significa que existe um inverso. Da mesma forma, na matriz (b). Mas o inverso para multiplicação é dado quando procura-se na matriz as entradas que contenham 1; o inteiro no topo dessa coluna é o inverso multiplicativo (analogamente, o inverso multiplicativo é o elemento cuja soma resulta em 1). Assim, $3 \cdot 3 \bmod 8 = 1$, ou seja, 9 dividido por 8 tem resto 1. Note que nem todos os inteiros $\bmod 8$ têm inverso, por exemplo, 2, 4 e 6.

Propriedades da Aritmética Modular

Seja a um inteiro. Chama-se $[r]$ as *classes de congruências de a módulo m*, ou seja, o conjunto formado, normalmente, de todos os inteiros que são congruentes a *a módulo m*.

$$\begin{aligned}[r] = r + m\mathbb{Z} &\coloneqq \{r + mb : b \in \mathbb{Z}\} \\ &= \{a \in \mathbb{Z} \quad | \quad a \equiv r \pmod{m}\}\end{aligned}$$

e, portanto,

$$a \in [r] \iff a \equiv r \pmod{m} \iff a = r + mb, \quad b \in \mathbb{Z},$$

Historicamente, classes de congruências são chamadas de **classe de resíduos** ou **resíduo** m, nós vamos adotar esta terminologia aqui também.

Denotaremos pelo símbolo $\mathbb{Z}_m$ o conjunto de classe de resíduos módulo m, o que se segue é simplesmente uma reafirmação do Teorema 7:

Definição 5. Seja m um inteiro positivo. Então $\mathbb{Z}_m$ consiste nas m classe resíduos distintas $[0], [1], [2], \ldots, [m-1]$, além disso, para todo $x \in \mathbb{Z}$, cada classe de resíduos módulo m contém uma representação exclusiva no intervalo $[x, x+m[$.

Podemos "equipar" $\mathbb{Z}_m$ com operações que definem adição e multiplicação. É natural pensarmos para $a, b \in \mathbb{Z}$ a seguintes definições:

$$[a] + [b] \coloneqq [a+b],$$
$$[a] \cdot [b] \coloneqq [a \cdot b].$$

É preciso verificar se estas definições são inequívocas, ou seja, mais precisamente, deve-se verificar que se $[a] = [a']$ e $[b] = [b']$, então $[a+b] = [a'+b']$ e $[a \cdot b] = [a' \cdot b']$. Porém, esta propriedade segue imediatamente das propriedades de congruências **(iv)** e **(vi)**.

Observe que para todo $a, b, c \in \mathbb{Z}$, nós temos:

$$[a] + [b] = [c] \iff a + b \equiv c \pmod{m},$$
$$[a] \cdot [b] = [c] \iff a \cdot b \equiv c \pmod{m}.$$

Por exemplo, a classe de resíduos módulo 6 são:

$$[0] = \{\ldots, -12, -6, 0, 6, 12, \ldots\}$$
$$[1] = \{\ldots, -11, -5, 1, 7, 13, \ldots\}$$
$$[2] = \{\ldots, -10, -4, 2, 8, 14, \ldots\}$$
$$[3] = \{\ldots, -9, -3, 3, 9, 15, \ldots\}$$
$$[4] = \{\ldots, -8, -2, 4, 10, 16, \ldots\}$$
$$[5] = \{\ldots, -7, -1, 5, 11, 17, \ldots\}.$$

A Tabela 3.2 mostra a adição e multiplicação das classe de resíduos módulo 6:

+	[0]	[1]	[2]	[3]	[4]	[5]
[0]	[0]	[1]	[2]	[3]	[4]	[5]
[1]	[1]	[2]	[3]	[4]	[5]	[0]
[2]	[2]	[3]	[4]	[5]	[0]	[1]
[3]	[3]	[4]	[5]	[0]	[1]	[2]
[4]	[4]	[5]	[0]	[1]	[2]	[3]
[5]	[5]	[0]	[1]	[2]	[3]	[4]

(a) Adição

·	[0]	[1]	[2]	[3]	[4]	[5]
[0]	[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]	[5]
[2]	[0]	[3]	[4]	[0]	[2]	[4]
[3]	[0]	[3]	[0]	[3]	[0]	[3]
[4]	[0]	[4]	[2]	[0]	[4]	[2]
[5]	[0]	[5]	[4]	[3]	[2]	[1]

(b) Multiplicação

Tabela 3.2: Aritmética com as classe de resíduos módulo 6.

Em vez de usarmos o intervalo $[0, 6[$, poderíamos também usar um outro intervalo, como, por exemplo, $[3, -3[$. Então, em vez de nomear as classe de resíduos [0], [1], [2], [3], [4], [5], poderíamos usar [-3], [-2], [-1], [0], [1], [2]. Observe que [-3] = [3], [-2] = [4] e [-1] = [5].

De todos os inteiros em uma classe de resíduos, o menor inteiro não negativo é aquele usado para representar a classe de resíduos. Encontrar o menor inteiro não negativo ao qual k é congruente módulo m é chamado de *redução k módulo m.*

Há uma peculiaridade da aritmética modular que a diferencia da aritmética comum. Primeiro, observe que (na aritmética modular) nós podemos escrever:

$$\textbf{Se} \quad a + b \equiv a + c \pmod{m}, \quad \textbf{então} \quad b \equiv c \pmod{m} \tag{3.6}$$

$$5 + 23 \equiv 5 + 7 \pmod{8}; \quad 23 \equiv 7 \pmod{8}$$

A equação (3.6) é consistente com a existência de um inverso aditivo. Adicionando inverso aditivo de a para ambos os lados equação (3.6), nós temos:

$$(-a) + a + b \equiv (-a) + a + c \pmod{m}$$
$$b \equiv c \pmod{m}$$

Porém, a propriedade abaixo (análoga à anterior, mas para multiplicação) somente é verdadeira quanto $mdc(a, m) = 1$. Ou seja, somente quando a e m são primos entre si, nós podemos afirmar que:

$$\textbf{Se} \quad a \cdot b \equiv a \cdot c \pmod{m}, \quad \textbf{então} \quad b \equiv c \pmod{m}. \tag{3.7}$$

Para compreender melhor isso, vamos pegar os inteiros, por exemplo, $a = 6$, $b = 3$, $c = 7$ e $m = 8$. Portanto, aplicando na equação 3.7 temos que:

$$6 \cdot 3 \equiv 6 \cdot 7 \pmod{8}$$

Portanto, $3 \not\equiv 7 \pmod 8$ pois o $mdc(6, 8) = 2$.

Semelhante ao caso da equação (3.6), podemos dizer que a equação (3.7) é consistente com a existência de um inverso multiplicativo. Aplicando o inverso multiplicativo a em ambos os lado da equação (3.7), temos:

$$(a^{-1})ab \equiv (a^{-1})ac \pmod m$$
$$b \equiv c \pmod m$$

Sendo assim, podemos realizar aritmética modular dentro de $\mathbb{Z}_m$ e esta segue as propriedades que descreveremos a seguir. Sejam $x, y, z \in \mathbb{Z}$ temos a seguinte Tabela 3.3:

Propriedades	***Expressões***
Comutatividade	$(x + y) \bmod m = (y + x) \bmod m$ $(x \cdot y) \bmod m = (y \cdot x) \bmod m$
Associatividade	$((z + x) + y) \bmod m = (z + (x + y)) \bmod m$ $((z \cdot x) \cdot y) \bmod m = (z \cdot (x \cdot y)) \bmod m$
Distributiva	$(z \cdot (x \cdot y)) \bmod m = ((z \cdot x) + (z \cdot y)) \bmod m$
Identidade	$(0 + z) \bmod m = z \bmod m$ $1 \cdot z \bmod m = z \bmod m$
Inverso Aditivo	Para todo inteiro $w \in \mathbb{Z}_m$, existem um z tal que $(w + z) \equiv 0 \pmod m$

Tabela 3.3: Propriedades da aritmética modular para inteiros em $\mathbb{Z}_m$.

Mostraremos na próxima seção que isso implica que $\mathbb{Z}_m$ é um anel comutativo com um elemento de identidade.

Algoritmo Euclidiano Revisado

Para qualquer inteiro a não-negativo e qualquer inteiro b positivo temos,

$$mdc(a, b) = mdc(b, a \bmod b), \tag{3.8}$$

por exemplo, $mdc(55, 22) = mdc(22, 55 \bmod 22) = mdc(22, 11) = 11$.

Seja $d = mdc(a, b)$, então, pela definição de maior divisor comum, temos que $d \mid a$ e $d \mid b$. Para qualquer b inteiro positivo, podemos expressar a desta forma

$$a = qb + r \equiv r \pmod{b}$$
$$a \bmod b = r$$

com $q, r \in \mathbb{Z}$. Assim sendo, $a \bmod b = a - qb$ para algum inteiro q. Além disso, sabemos que se $d \mid b$, então ele divide qb. Nós também temos que $d \mid a$, logo, $d \mid (a \bmod b)$. Isso mostra que d é um divisor comum de b e também $a \bmod b$. Por outro lado, se d é um divisor comum de b e também $a \bmod b$, então $d \mid qb$ e também é verdade que $d \mid qb + a \bmod b$, que é equivalentemente a dizer que $d \mid a$. Então, o conjunto de divisores comuns de a e também de b é igual ao conjunto de divisores comuns de b e $a \bmod b$. Portanto, isso comprova que o $mdc(a, b) = mdc(b, a \bmod b)$.

Note que a equação 3.8 pode ser usada repetidamente para determinar o maior divisor comum.

$$mdc(18, 12) = mdc(12, 6) = mdc(6, 0) = 6$$

Sendo assim, o esquema 3.3 pode ser reescrito da seguinte maneira:

$$\left.\begin{array}{rcll} a & = & q_1 b + r_1; & r_1 = a \bmod b \\ b & = & q_2 r_1 + r_2; & r_2 = b \bmod r_1 \\ r_1 & = & q_3 r_2 + r_3; & r_3 = r_1 \bmod r_2 \\ \vdots & & & \\ r_{n-2} & = & q_n r_{n-1} + r_n; & r_n = r_{n-2} \bmod r_{n-1} \\ r_{n-1} & = & q_{n+1} r_n + 0, & r_{n+1} = r_{n-1} \bmod r_n = 0 \\ d & = & mdc(a,b) = r_n. & \end{array}\right\} \quad (3.9)$$

Podemos implementar o Algoritmo Euclidiano de forma recursiva:

Algoritmo 2 Algoritmo de Euclides

```
1: procedure EUCLIDES(a, b)
2:     if b = 0 then
3:         return a
4:     else
5:         return EUCLIDES(b, a mod b)
6:     end if
7: end procedure
```

Algoritmo Euclidiano Estendido

Passamos agora a olhar para uma melhoria do Algoritmo Euclidiano que será importante para cálculos posteriores na área de corpos finitos e também em algoritmos de criptografia, como o RSA. Dados dois inteiros a e b, o Algoritmo Euclidiano Estendido não apenas calcula o maior divisor comum d, mas também dois inteiros adicionais x e y a seguinte equação:

$$ax + by = d = mdc(a, b) \quad (3.10)$$

Antes de examinarmos o algoritmo, vamos olhar para alguns dos valores de x e y quando $a = 42$ e $b = 30$. Note que $mdc(42, 30) = 6$.

Aqui está uma representação parcial em forma de matriz dos possíveis valores gerados d pela equação (3.10) acima:

$y \setminus x$	-3	-2	-1	0	1	2	3
-3	-216	-174	-132	-90	-48	-6	36
-2	-186	-144	-102	-60	-18	24	66
-1	-156	-114	-72	-30	12	54	96
0	-126	-84	-42	0	42	84	126
1	-96	-54	-12	30	72	144	186
2	-66	-24	18	60	102	144	188
3	-36	6	48	90	132	174	216

Observe que todas as entradas desta matriz são divisíveis por 6. Isto não é uma surpresa porque ambos os valores 42 e 30 são divisíveis por 6, então todo número da forma $42x + 30y = 6 \cdot (7x + 5y)$ é um múltiplo de 6. Observe também que o $mdc(42, 30) = 6$ aparece na matriz acima. Em geral, pode-se mostrar que, dados dois inteiros a e b, o menor valor positivo de $ax+by$ é igual ao $mdc(a, b)$. Agora vamos demonstrar o Algoritmo de Euclidiano Estendido para determinar (x, y, d) dado a, b. Passamos novamente pela equação (3.10), e para cada passo i podemos encontrar inteiros x_i e y_i que satisfazem $r_i =$

$ax_i + by_i$. Nós terminamos com a seguinte sequência:

$$\left.\begin{array}{rcll} a & = & q_1 b + r_1, & r_1 = ax_1 + by_1; \\ b & = & q_2 r_1 + r_2, & r_2 = ax_2 + by_2; \\ r_1 & = & q_3 r_2 + r_3, & r_3 = ax_3 + by_3; \\ \vdots & & & \\ r_{n-2} & = & q_n r_{n-1} + r_n, & r_n = ax_n + by_n; \\ r_{n-1} & = & q_{n+1} r_n + 0. & \end{array}\right\} \tag{3.11}$$

Observe, que agora podemos reorganizar os termos para escrever

$$r_i = r_{i-2} - r_{i-1} q_i \tag{3.12}$$

Também na linha $i-1$ e $i-2$, nós encontramos os valores

$$r_{i-2} = ax_{i-2} + by_{i-2}, \quad \text{e} \quad r_{i-1} = ax_{i-1} + by_{i-1}$$

Substituindo na equação (3.12), temos:

$$r_i = (ax_{i-2} + by_{i-2}) - (ax_{i-1} + by_{i-1})q_i = a(x_{i-2} - q_i x_{i-1}) + b(y_{i-2} - q_i y_{i-1})$$

Porém, nós já assumimos que $r_i = ax_i + by_i$. Portanto,

$$x_i = x_{i-2} - q_i x_{i-1}, \quad \text{e} \quad y_i = y_{i-2} + q_i y_{i-1}$$

Precisamos fazer vários comentários adicionais aqui. Em cada linha, calculamos um novo resto r_i com base nos restos das duas linhas anteriores,

sabendo r_{i-1} e r_{i-2}. Para iniciar o algoritmo, precisamos dos valores r_0 e r_{-1}, que são apenas a e b. Portanto, é simples determinar os valores necessários para x_{-1}, y_{-1}, x_0 e y_0. Sabemos do Algoritmo Euclidiano que o processo termina com um resto igual a zero, e que o maior divisor comum de a e b é $d = mdc(a, b) = r_n$. Mas também determinamos que $d = r_n = ax_n + by_n$. Portanto, na equação (3.10), $x = x_n$ e $y = y_n$. Por exemplo, vamos usar $a = 1759$ e $b = 550$, e resolver $1759x + 550y = mdc(1759, 550)$. Os resultados são mostrados na Tabela 3.4. Assim, nós temos: $1759 \cdot (-111) + 550 \cdot 355 = -195249 + 195250 = 1$.

i	r_i	q_i	x_i	y_i
-1	1759		1	0
0	550		0	1
1	109	3	1	-3
2	5	5	-5	16
3	4	21	106	-339
4	1	1	-111	355
5	0	4		

Resultado: $d = 1$; $x = -111$; $y = 355$.

Tabela 3.4: Exemplo do Algoritmo Eucledeano Estendido.

3.4 Teorema Fundamental da Aritmética

Uma preocupação central da Teoria dos Números é o estudo dos números primos. Nesta seção forneceremos uma visão geral sobre este assunto e mostraremos que todo inteiro não igual a 0, 1, -1 pode-se expressar como produto

de números primos, de forma única, a menos da ordem dos fatores. Esse resultado, conhecido como o Teorema Fundamental da Aritmética, já aparece no livro do IX de *Os Elementos* de Euclides e destaca a importância dos primos na Teoria dos Números: eles desempenham um papel análogo ao dos átomos na estrutura da matéria. Todos os outros números podem ser obtidos através de produtos entre os números primos.

Definição 6(Números Primos). Um inteiro p diz-se *primo* se tem exatamente dois divisores positivos, 1 e $|p|$.

Note que a definição exclui propositadamente o 0, que tem infinitos divisores positivos, e os inteiros 1 e -1 que têm *um divisor positivo.*

Um número diferente de 0, 1 e -1 que não é primo diz-se *composto.* Note que, da definição, vem imediatamente que, se um inteiro não-nulo a é composto, ele admite um divisor b tal que $|b|$ seja diferente de 1 e de $|a|$, isto é, um divisor b tal que $1 < |b| < |a|$. Um divisor nessas condições diz-se um *divisor próprio* de a.

Suponha que p é um primo, $a \in \mathbb{Z}$ e tomando a Definição 6 temos que

$$p \mid a \implies mdc(a,p) = p,$$
$$p \nmid a \implies mdc(a,p) = 1.$$

Sendo assim combinando esta observação com o Teorema 7, temos:

Teorema 9. Seja p um número primo e sejam $a, b \in \mathbb{Z}$. Se $p \mid ab$, então $p \mid a$ ou $p \mid b$.

Demonstração. Assumindo que $p \mid ab$. Se $p \mid a$, a tese está verificada. Caso

contrário, como observado logo acima o $mdc(a, p) = 1$, assim pelo Teorema 7, temos que $p \mid b$. □

Cololário 1. Se um *inteiro primo p divide um produto* $a_1 a_2 \dots a_n$, então $p \mid a_k$ para algum $k = 1, \dots, n$, ou seja, $1 \leq k \leq n$.

Demonstração. Vamos provar por indução sobre n. Para $n = 1$ temos que $p \mid a_1$ de fato.

Agora seja $n > 1$, e assumindo que temos $n - 1$ termos. Então, pelo Teorema 9, $p \mid a_1$ ou $p \mid a_2 \dots a_n$; se $p \mid a_1$, que é verdade pelo caso base; caso contrário, por indução, p divide um dos $a_2 \dots a_n$. □

Teorema 10. Seja p um inteiro diferente de 0, 1 e -1. Então, p é primo se, e somente se, toda vez que p divide um produto de dois números, p divide pelo menos um dos fatores.

Demonstração. Suponha que p tenha a propriedade do enunciado mas não seja primo. Então, $|p|$ pode ser escrito da forma $|p| = a \cdot b$, onde a e b são divisores próprios positivos, isto é, verificam

$$1 < a < |p|,$$
$$1 < b < |p|$$

Consequentemente, $p \mid ab$, mas $p \nmid a$ e $p \nmid b$; logo, uma contradição. □

Lema 1. Todo inteiro $a > 1$ pode ser escrito *como produto de números primos.*

Teorema 11. Seja $n > 1$. Então, *existem primos positivos* $p_1 \leq p_2 \leq \ldots \leq p_t$ tais que $a = p_1 p_2 \ldots p_t$, e essa decomposição é única.

Teorema 12(Teorema Fundamental da Aritmética). Seja a um inteiro diferente de 0. Então, *existem primos positivos* $p_1 < p_2 < \ldots < p_r$ e *inteiros positivos* $n_1, n_2, \ldots, n_r$ tais que $a = E p_1^{n_1} \ldots p_r^{n_r}$, em que $E = \pm 1$, conforme a seja positivo ou negativo, além disso, essa decomposição é única.

Demonstração. Temos que $a = E\,|a|$, onde $E = 1$ ou $E = -1$, conforme a seja positivo ou negativo. Como $|a|$ é positivo, do Teorema 11, temos que existe primos $p_1 \leq p_2 \leq \ldots \leq p_t$ tais que

$$a = E p_1 p_2 \ldots p_t.$$

Agrupando os primos eventualmente repetidos, podemos escrever

$$a = E p_1^{n_1} \ldots p_r^{n_r}$$

A unicidade segue diretamente do Teorema 11. □

Cololário 2. Sejam a e d *inteiros diferentes* de 0. Então, *existem primos positivos* $p_1 < p_2 < \cdots < p_t$ e inteiros não-negativos $n_1, \ldots, n_t$, $m_1, \ldots, m_t$ (mas eventualmente iguais a zero, se necessário) tais que

$$a = E_1 p_1^{n_1} \ldots p_t^{n_t},$$
$$d = E_2 p_1^{m1} \ldots p_t^{m_t},$$

em que $E_i = \pm 1 para i = 1, 2$.

Usando essas decomposições, podemos dar um critério de divisibilidade que formulamos apenas para inteiros positivos (observe que isso não é uma perda de generalidade, já que $d \mid a$ se, e somente se, $|d|$ divide $|a|$).

Lema 2. Sejam $a = p_1^{n_1} \dots p_t^{n_t}$ e $d = p_1^{m_1} \dots p_t^{m_t}$ *inteiros positivos*, onde $p_1, \dots, p_t$ são *primos positivos* e n_i, m_i são inteiros não-negativos para todo $1 \leq i \leq t$. Então, $d \mid a$ se, e somente se, $m_i \leq n_i$, $1 \leq i \leq t$.

Teorema 13. Sejam $a = p_1^{n_1} \dots p_t^{n_t}$ e $b = p_1^{m_1} \dots p_t^{m_t}$ inteiros nas condições do lema 2. Então,

$$d = mdc(a,b) = p_1^{\alpha_1} \dots p_t^{\alpha_t}, \quad \text{em que} \quad \alpha_i = \min(n_i, m_i), 1 \leq i \leq t,$$
$$m = mmc(a,b) = p_1^{\beta_1} \dots p_t^{\beta_t}, \quad \text{em que} \quad \beta_i = \max(n_i, m_i), 1 \leq i \leq t.$$

3.5 Teorema Chinês do Resto

Teorema 14(Teorema Chinês do Resto). Sejam $\{m_i\}_{i=1}^{k}$ inteiros, primos entre si dois a dois (isto é, $i \neq j$ para todo $1 \leq i, j \leq k$), e sejam $b_1, \ldots, b_k$ inteiros arbitrários. Então, existe uma solução $a \in \mathbb{Z}$ para o sistema de congruências lineares

$$\begin{aligned} a &\equiv b_1 \pmod{m_1} \\ a &\equiv b_2 \pmod{m_2} \\ &\vdots \\ a &\equiv b_k \pmod{m_k}. \end{aligned}$$

Além disso, qualquer $b \in \mathbb{Z}$ é uma solução para este sistema de congruências lineares se, e somente se, $a \equiv b \pmod{m}$, onde $m \coloneqq \prod_{i=1}^{k} m_i$.

Demonstração. Para provar a existência de uma solução a para o sistema de congruências lineares, primeiro vamos mostrar como construir inteiros do tipo $e_1, \ldots, e_k$ tal que para $i, j = 1, \ldots, k$, temos que

$$e_j \equiv \begin{cases} 1 \pmod{m_i} & \text{se} \quad i = j \\ 0 \pmod{m_i} & \text{se} \quad i \neq j. \end{cases} \tag{3.13}$$

Se fizermos isso, em seguida, definindo

$$a \coloneqq \sum_{i=1}^{k} a_i e_i,$$

para $j = 1, \ldots, k$, temos que

$$a = \sum_{i=1}^{k} a_i e_i \equiv a_j \pmod{m_j},$$

uma vez que todos os termos nesta soma são zero módulo m_j, exceto para o termo $i = j$, que é congruente com a_j módulo m_j.

Construiremos agora os inteiros do tipo $e_1, \ldots, e_k$ que sejam satisfatórios para equação 3.13. Seja $m \coloneqq \prod_{i=1}^{k} m_i$ com $i = 1, \ldots,$ k, e também seja $m_i^* \coloneqq m/m_i$; isto é, m_i^* é o produto de todos os números m_j com $i \neq j$. Como $\{m_i\}_{i=1}^{k}$ são inteiros, primos entre si dois a dois, segue que para todo i entre 1 e k, temos que $mcd(m_i, m_i^*) = 1$, e podemos definir também $t_i \coloneqq (m_i^*)^{-1} \bmod m_i$ e $e_i \coloneqq m_i^* t_i$. Uma vez que $e_i \equiv 1 \pmod{m_i}$, $m_i \mid m_j^*$ para todo $i \neq j$ e $e_j \equiv 0 \pmod{m_i}$, temos que a equação 3.13 é satisfeita. Isso prova a existência de uma solução a para o sistema de congruências. Se $a \equiv b \pmod{m}$, então temos que $m_i \mid m$ para $i = i, \ldots, k$, vemos que $a \equiv b_i \equiv b \pmod{m_i}$ para $i = 1, \ldots, k$, então b também resolve o sistema de congruências lineares.

Finalmente, se b é uma solução para o sistema de congruências lineares, então $a \equiv b_i \equiv b \pmod{m_i}$ para $i = 1, \ldots, k$. Portanto, $m_i \mid (a - b)$ para $i = 1, \ldots,$ k. Sendo assim $\{m_i\}_{i=1}^{k}$ são inteiros primos entre si, isto implica que $m \mid (a - b)$, ou equivalentemente dizer que, $a \equiv b \pmod{m}$. □

3.6 Teorema de Fermat e Euler

Dois teoremas que desempenham papéis importantes na criptografia de chave pública são o Teorema de Fermat e o Teorema de Euler.

Função Totient de Euler

Antes de apresentarmos o Teorema de Euler, precisamos definir uma

função importante em teoria dos números. Estamos nos referindo à **Função Totient**, ela é definida como:

Definição 7. Seja $n \in \mathbb{Z}$ e $n \geq 1$, indicaremos por $\varphi(n)$ a função:

$$\varphi(n) \coloneqq |\mathbb{Z}_n^*| \,.$$

Equivalentemente, $\varphi(n)$ é a quantidade de números compreendidos entre 1 e $n-1$ que são relativamente primos a n. Denotamos o conjunto destes números como $A = \left\{x_1, x_2, \ldots, x_{\varphi(n)}\right\}$, onde cada x_i é tal que $1 \leq x_i \leq n$ e $mdc(x_i, n) = 1$. A função assim definida chama-se Função Totient φ de Euler.

Por exemplo, se $n = 4$, os inteiros positivos menores que ou iguais a 4 que são primos em relação a ele são os inteiros 1 e 3, ou seja, os elementos de $A = \{1, 3\}$ que são primos com relação a 4, assim $\varphi(4) = 2$. Analogamente, temos $\varphi(5) = 4$, $\varphi(6) = 2$.

Usando o Teorema 3.5 é fácil obter uma boa fórmula para φ em termos da fatoração em primos de n, conforme estabelecemos com o teorema a seguir.

Teorema 15. Sejam $\{n_i\}_{i=1}^{n}$ inteiros positivos e primos entre si dois a dois, e $n \coloneqq \prod_{i=1}^{k} n_i$.

$$\varphi(n) = \prod_{i=1}^{k} \varphi(n_i).$$

A Tabela 3.5 lista os primeiros 30 valores de $\varphi(n)$. O valor de $\varphi(1)$ é sem significado, porém, é definido como 1. Para um número primo p deve ficar

claro que:

$$\varphi(p) = p - 1.$$

Supondo que nós temos dois números primos p e q com $p \neq q$. Então, podemos mostrar que, para $n = pq$,

$$\varphi(n) = \varphi(pq) = \varphi(p) \cdot \varphi(q) = (p-1) \cdot (q-1)$$

Demonstração. Considerando que $\varphi(n) = \varphi(p) \cdot \varphi(q)$ e o conjunto de números inteiros positivos menores do que n, isto é, $\{1, \ldots, (pq-1)\}$, e o conjunto de inteiros positivos que não são primos com relação a n, ou seja, $\{p, 2p, \ldots, (q-1)p\}$, e $\{q, 2q, \ldots, (p-1)q\}$. Então, temos que:

$$\begin{aligned}
\varphi(n) &= (pq-1) - [(q-1) + (p-1)] \\
&= pq - (p+q) + 1 \\
&= (p-1) \cdot (q-1) \\
&= \varphi(p)\varphi(q)
\end{aligned}$$

□

Por exemplo, se $\varphi(21) = \varphi(3) \cdot \varphi(7) = (3-1) \cdot (7-1) = 2 \cdot 6 = 12$. Os elementos de $A = \{1, 2, 4, 5, 8, 10, 11, 13, 16, 17, 19, 20\}$ são todos primos com relação a 21.

n	φ
1	1
2	1
3	2
4	2
5	4
6	2
7	6
8	4
9	6
10	4

n	φ
11	10
12	4
13	12
14	6
15	8
16	8
17	16
18	6
19	18
20	8

n	φ
21	12
22	10
23	22
24	8
25	20
26	12
27	18
28	12
29	28
30	8

Tabela 3.5: Alguns valores da Função Totient de Euler $\varphi(n)$.

3.6.1 Teorema de Euler

Teorema 16(Teorema de Euler). Sejam a e n inteiros com $n \geq 1$, tais que $mdc(a, n) = 1$. Então, $a^{\varphi(n)} \equiv 1 \pmod{n}$.

Não vamos fazer a demonstração deste teorema pois ela é bem extensa. Fica a cargo do leitor. Sendo assim, vamos ver alguns exemplos:

- **1.** $a = 3; n = 10; \varphi(10) = 4; a^{\varphi(n)} = 3^4 = 81 \equiv 1 \pmod{10}$;
- **2.** $a = 2; n = 11; \varphi(11) = 10 a^{\varphi(n)} = 2^{10} = 1024 \equiv 1 \pmod{11}$.

3.6.2 Teorema de Fermat

Teorema 17(Teorema de Fermat). Sejam p um primo e a um inteiro tal que $p \nmid a$, então

$$a^{p-1} \equiv 1 \pmod{p}$$

Demonstração. Considere o conjunto de inteiros **(i)** $= \{a, 2a, 3a, \ldots, (p-1)a\}$. Dados dois elementos quaisquer desse conjunto, eles não são congruentes entre si módulo p, pois, se $xa \equiv ya \pmod{p}$ com $1 \leq x,\ y \leq p-1$, como $mdc(a,p) = 1$, cancelando teríamos $x \equiv y \pmod{p}$, o que não acontece, já que os elementos do conjunto **(ii)** $= \{1, 2, 3, \ldots, p-1\}$ não são congruentes entre si módulo p. Além disso, nenhum dos elementos de **(i)** é congruente a 0 módulo p, já que, se $p \mid xa$, com $1 \leq x \leq p-1$, então $p \mid x$ ou $p \mid a$, o que não acontece. Segue-se então que os elementos de **(i)** são congruentes aos elementos de **(ii)**. Temos, então $p-1$ congruências da forma

$$\begin{aligned} a &\equiv x_1 \pmod{p} \\ 2a &\equiv x_2 \pmod{p} \\ &\vdots \\ (p-1)a &\equiv x_{p-1} \pmod{p} \end{aligned}$$

onde $x_1, x_2, \ldots, x_{p-1}$ são os inteiros $1, 2, \ldots, p-1$, eventualmente em uma outra ordem. Multiplicando essas congruências, temos

$$a \cdot 2a \cdot \ldots (p-1)a \equiv 1 \cdot 2 \cdot (p-1) \pmod{p}$$

ou seja,

$$(p-1)!a^{p-1} \equiv (p-1)! \pmod{p}$$

Como $mdc((p-1)!, p) = 1$, podemos cancelar e obtemos

$$a^{p-1} \equiv 1 \pmod{p}$$

□

Note que, se p é primo, $\varphi(p) = p - 1$; então o Teorema de Fermat segue como um caso particular do Teorema de Euler.

$$\boxed{a^{\varphi(n)+1} \equiv a \pmod{n}} \tag{3.14}$$

Cololário 3. Sejam p um primo e a um inteiro arbitrário. Então, $a^p \equiv a \pmod{p}$.

Demonstração. Se $p \nmid a$, do Teorema 17 temos que $a^{p-1} \equiv 1 \pmod{p}$; multiplicando os membros dessa congruência por a segue que $a^p \equiv a \pmod{p}$.

Se $p \mid a$, então $p \mid a^p$, e consequentemente $p \mid (a^p - a)$; logo $a^p \equiv a \pmod{p}$. □

O Teorema de Fermat pode ser usado para provar diversos resultados sobre divisibilidade. Ilustraremos esta afirmação com alguns exemplos:

- **1.** Seja a um inteiro arbitrário. Provaremos que o algarismos das unidades de a e de a^5 é o mesmo (quando escritos em base 10).

  Se r e s indicam esses algarismos, como $a \equiv r \pmod{10}$ e $a^5 \equiv s \pmod{10}$, para concluir a igualdade bastará mostrar que $a^5 \equiv a \pmod{10}$. Do Teorema de Fermat, temos que $a^5 \equiv a \pmod{5}$, logo, $5 \mid (a^5 - a)$. Por outro lado, $2 \mid (a^5 - a)$, pois a e a^5 são ambos pares ou ambos ímpares. Como $mdc(2, 5) = 1$, temos que $10 \mid (a^5 - a)$, como queríamos demostrar.

- **2.** Dados a e b inteiros arbitrários e p um primo, tem-se que

$$(a + b)^p \equiv a^p + b^p \pmod{p}.$$

De fato, tomando módulo p e usando o Teorema de Fermat, temos que

$$(a+b)^p \equiv a+b \equiv a^p + b^p \pmod{p}.$$

3.7 Resumo

Neste capítulo demos uma breve explicação das principais definições e teoremas que servem como base para o leitor compreender os próximos capítulos e também os conceitos usados para aplicar **cifras simétricas**, por exemplo, AES, e **cifras assimétricas** RSA e CCE. Estes exemplos são largamente usados hoje para codificar uma mensagem como foi citado no capítulo 3.

Capítulo 4

Grupos, Anéis e Corpos

No capítulo 4 abordaremos estes três temas: grupos, anéis e corpos, eles são elementos fundamentais de um ramo da Matemática conhecido como álgebra abstrata ou álgebra moderna.

Estamos preocupados com conjuntos cujos elementos podemos operar algebricamente; isto é, podemos combinar dois elementos do conjunto, talvez de várias maneiras, para obter um terceiro elemento do conjunto.

Essas operações estão sujeitas a regras específicas, que definem a natureza do conjunto. Por convenção, a notação para as duas classes principais de operações nos elementos do conjunto é geralmente a mesma que a notação para adição e multiplicação em números ordinários. No entanto, é importante notar que, na álgebra abstrata, não estamos limitados a operações aritméticas comuns. Tudo isso deve ficar claro quando prosseguirmos neste capítulo.

4.1 Grupos

Um **grupo** é um conjunto, G, juntamente com uma operação $\bullet$ (chamada de lei do grupo G) que combina quaisquer dois elementos a e b para formar outro elemento, denotado de $a \bullet b$. Para se qualificar como um grupo, o conjunto e a operação, $\{G, \bullet\}$, devem satisfazer quatro requisitos conhecidos como axiomas do grupo:

- **(A1) Operação fechada em G:** Para todo a e $b \in G$, o resultado da operação, $a \bullet b$, também está em G.
- **(A2) Associatividade:** Para todo a, b e $c \in G$, $a \bullet (b \bullet c) = (a \bullet b) \bullet c$.
- **(A3) Identidade:** Há um elemento $e \in G$ tal que $a \bullet e = e \bullet a = a$ para todo $a \in G$.
- **(A4) Inverso:** Para cada $a \in G$, existe um elemento $a' \in G$, comumente denotado a^{-1} (ou $-a$, se a operação for denotada $+$) tal que $a \bullet a' = a' \bullet a = e$, no qual e é o elemento de identidade.

Se um grupo tem um número finito de elementos, é denominado de **grupo finito**, e a **ordem** deste grupo é igual ao número de elementos no grupo. Caso contrário, o grupo é chamado de **grupo infinito**.

Um grupo se diz ser **abeliano** se ele satisfaz o seguinte axioma, além dos citados acima:

- **(A5) Comutatividade:** $a \bullet b = b \bullet a$, $\forall a, b \in G$.

4.2 Anéis

Um **anel** é um conjunto, R, composto por duas operações, $+$ e $\cdot$, chamadas de adição e multiplicação respectivamente, tal que $\forall a, b, c \in R$, os seguintes axiomas são obedecidos:

- **(A1 a A5)** R é um grupo abeliano em relação à adição; isto é, R satisfaz os axiomas $A1$ a $A5$ em relação à operação $+$. Assim, o inverso de um elemento a com respeito à operação adição no anel é $-a$ e o elemento neutro desta operação é o 0.
- **(M1) Operação fechada em R:** Se a e $b \in R$, então $a \cdot b$ está também em R.
- **(M2) Associatividade:** $a \cdot (b \cdot c) = (a \cdot b) \cdot c, \forall a, b, c \in R$.
- **(M3) Distributividade:** $a \cdot (b + c) = a \cdot b + a \cdot c, \forall a, b, c \in R$.

Um anel é dito **comutativo** se satisfaz a seguinte condição adicional:

- **(M4) Comutatividade:** $a \cdot b = b \cdot a, \forall a, b \in R$.

Definimos um **domínio de integridade**, que é um anel comutativo que obedece aos seguintes axiomas:

- **(M5) Identidade:** Existe um elemento $1 \in R$ tal que $a \cdot 1 = 1 \cdot a = a, \forall a, b \in R$
- **(M6) Sem divisores de zero:** Se $a, b \in R$ e $ab = 0$, então temos que $a = 0$ ou $b = 0$.

4.3 Corpos

Primeiramente, vamos definir o que é um **corpo**. Um corpo F, é um conjunto de elementos com duas operações, $+$ e $\cdot$, chamadas de adição e multiplicação respectivamente, todos os elementos $a, b, c \in F$ seguem os seguintes axiomas:

- **(A1 a A5 e M1 a M6)** F é um domínio de integridade, ou seja, F satisfaz axiomas $A1$ a $A5$ e $M1$ a $M6$.
- **(M7) Inverso Multiplicativo** Para todo $a \in F$, exceto 0, existe um elemento $a^{-1} \in F$ tal que $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

Em essência, um corpo é um conjunto no qual podemos fazer adição e multiplicação comutativa sem sair do conjunto e que contém o inverso multiplicativo de cada elemento.

4.3.1 Corpos Finitos da forma $GF(p)$

Nós definimos um corpo como um conjunto que obedece a todos os axiomas da seção anterior. Estes não são de interesse particular no contexto de criptografia. Porém, corpos finitos desempenham um papel crucial em muitos algoritmos criptográficos.

Pode ser mostrado que a ordem de corpos finitos (número de elementos) deve ser uma potência de um primo, p^n, onde n é um inteiro positivo. Nós discutimos números primos na seção 3.4.

O corpo finito de ordem p^n é geralmente escrito da forma $GF(p^n)$; GF significa "Galois Field", em homenagem ao matemático que primeiro estudou

corpos finitos. Dois casos especiais são de interesse para nossos propósitos neste trabalho.

Quando $n = 1$, nós temos o corpo finito $GF(p)$; este corpo finito tem uma estrutura diferente dos corpos finitos quando $n > 1$, por esta razão nós também falaremos sobre os corpos finitos da forma $GF(2^n)$.

Corpos Finitos de Ordem p

Sejam p um primo e $GF(p)$ o **corpo finito de ordem** p, representado como um conjunto $\mathbb{Z}_p$ de inteiros $\{0, 1, \ldots, p-1\}$ juntamente com as operações aritméticas módulo p.

Lembrem-se de que mostramos na seção 3.3 que o conjunto $\mathbb{Z}_n$ de inteiros $\{0, 1, \ldots, n-1\}$, juntamente com as operações aritméticas módulo n, é um anel comutativo (veja a Tabela 3.3). Observamos que qualquer inteiro em $\mathbb{Z}_n$ tem um inverso se, e somente se, esse inteiro é relativamente primo em relação a n [1]. Se n é primo, então todos os inteiros não negativos em $\mathbb{Z}_n$ são relativamente primos em relação a n, e portanto, existe um inverso para todos os inteiros diferentes de zero em $\mathbb{Z}_n$, portanto, para $\mathbb{Z}_p$, nós podemos adicionar mais uma propriedade:

- **Inverso Multiplicativo** (w^{-1}) Para todo $w \in \mathbb{Z}_p, w \neq 0$, existe um $z \in \mathbb{Z}_p$ tal que $w \cdot z \equiv 1 \pmod{p}$

Porque w é relativamente primo em relação a p, se nós multiplicarmos todos os elementos de $\mathbb{Z}_p$ por w, os restos resultantes são todos os elementos de $\mathbb{Z}_p$ permutados. Assim, exatamente um dos restos tem o valor 1. Portanto, existe um inteiro em $\mathbb{Z}_p$ (denominado w^{-1}) que, quando multiplicado por w, resulta em 1; portanto, $\mathbb{Z}_p$ é de fato um corpo finito, aliás, podemos conferir

[1] Como afirmamos na discussão da equação (3.7), dois inteiros são **relativamente primos** se o máximo divisor comum de ambos é 1.

a equação (3.7) através do inverso de a:

$$\text{Se} \quad (a \cdot b) \equiv (a \cdot c) \pmod{m}, \quad \text{então} \quad b \equiv c \pmod{p} \tag{4.1}$$

Basta multiplicar ambos os lados da equação (4.1) pelo inverso de a.

$$((a^{-1}) \cdot a \cdot b) \equiv ((a^{-1}) \cdot a \cdot c) \pmod{p}$$
$$b \equiv c \pmod{p}$$

Encontrando o Inverso Multiplicativo em $GF(p)$

É relativamente fácil encontrar o inverso de um elemento em $GF(p)$ para pequenos valores de p. Simplesmente montamos uma Tabela 4.1, e o resultado desejado pode ser lido diretamente. No entanto, para grandes valores de p, esta abordagem não é prática.

+	0	1	2	3	4	5	6
0	**0**	1	2	3	4	5	6
1	1	2	3	4	5	6	**0**
2	2	3	4	5	6	**0**	1
3	3	4	5	6	**0**	1	2
4	4	5	6	**0**	1	2	3
5	5	6	**0**	1	2	3	4
6	6	**0**	1	2	3	4	5

(a) Adição

·	0	1	2	3	4	5	6
0	0	**1**	2	3	4	5	6
1	0	0	0	0	0	0	0
2	0	2	4	6	**1**	3	5
3	0	3	6	2	5	**1**	4
4	0	4	**1**	5	2	6	3
5	0	5	3	**1**	6	4	2
6	0	6	5	4	3	2	**1**

(b) Multiplicação

w	$-w$	w^{-1}
0	**0**	–
1	**6**	**1**
2	**5**	**4**
3	**4**	**5**
4	**3**	**2**
5	**2**	**3**
6	**1**	**6**

(c) Inversa

Tabela 4.1: Aritmética em $GF(7)$.

Se a e b são relativamente primos, então existe um inverso de b módulo a, ou seja, se $mdc(a, b) = 1$ existe um $b^{-1} \leq a$ tal que $b \cdot b^{-1} = 1 \pmod{a}$.

Se a é um número primo e $b \leq a$, então a e b são relativamente primos e o $mdc(a, b) = 1$. Com isso podemos encontrar b^{-1} usando o Algoritmo de Euclides Estendido.

Sendo assim, retomando a equação (3.10), podemos resolvê-la pelo Algo-

ritmo de Euclides Estendido:

$$ax + by = d = mdc(a, b)$$

Agora, se $mdc(a, b) = 1$, então nós temos que $ax + by = 1$. Usando a igualdade da aritmética modular, definida na seção 3.3, nós podemos dizer que:

$$(ax \bmod a + by \bmod a) \bmod a = 1 \bmod a$$
$$\implies 0 + (by \bmod a) = 1$$

Porém, se $by \bmod a = 1$, então $y = b^{-1}$ e aplicando o Algoritmo Euclidiano Estendido para equação (3.10) encontramos o valor do inverso multiplicativo b se $mdc(a, b) = 1$. Considere o exemplo que apresentamos na Tabela 3.4, temos que a = 1759 e $b = 550$. A solução da equação $1759x + 550y = 1$, y é igual 355.

Portanto, $y = b^{-1} = 355$. Para verificar que y é de fato o inverso de b, calculamos $550 \cdot 355 \bmod 1759 = 195250 \bmod 1759 = 1$.

4.3.2 Corpos Finitos da Forma $GF(2^n)$

Na subseção anterior, nós mencionamos que a ordem de um corpo finito deve ser do tipo p^n, no qual p é um número primo e n é um inteiro positivo. Na subseção 4.3.1 nós analisamos um caso especial de corpos finitos de ordem p. Descobrimos que, usando aritmética modular em $\mathbb{Z}_p$, todos os axiomas de um corpo, citados na seção 4.1 estão satisfeitos. Para um polinômio sobre p^n, com $n > 1$, operações módulo p^n não formam um corpo. Nesta seção, nós apresentamos que estruturas que satisfazem os axiomas para um corpo em um conjunto com p^n elementos e também $GF(2^n)$.

Praticamente, todos os algoritmos, sejam eles de chaves simétricas ou de chaves públicas (assimétricas), envolvem operações com aritmética nos inteiros. Se uma das operações usadas no algoritmo é divisão, então precisamos trabalhar com aritmética sobre um corpo. Por conveniência e por eficiência de implementação, nós precisamos operar com inteiros que se encaixam exatamente em um determinado número de bits, sem padrões de bits desperdiçados. Isto é, desejamos trabalhar com números inteiros no intervalo de 0 a $2^n - 1$ que se encaixem em uma palavra com n-bits.

Suponha que nós queremos definir um algoritmo de encriptação convencional que opera dados de 8 bits de cada vez, e queremos realizar a divisão. Com 8 bits, nós podemos representar inteiros no intervalo de 0 a 255. Porém, 256 não é um número primo, de modo que com a aritmética módulo $\mathbb{Z}_{256}$, o conjunto $\mathbb{Z}_{256}$ não é corpo (pois para um polinômio sobre p^n, com $n > 1$, operações módulo p^n não formam um corpo). O número primo mais próximo de 256 é 251. Então, o conjunto $\mathbb{Z}_{251}$, usando o módulo 251, é um corpo. Contudo, neste caso, o padrões de 8 bits representando os números inteiros de 251 a 255 não seriam usados, resultando num uso ineficiente de armazenamento.

Como o exemplo anterior indica, se todas as operações aritméticas forem usadas e desejarmos representar uma gama completa de inteiros em n-bits, então as operações módulo 2^n não funcionarão. Equivalentemente, o conjunto de inteiros módulo 2^n para $n > 1$, não é um corpo finito de ordem p^n. Além disso, mesmo que o algoritmo de criptografia use apenas adição e multiplicação, mas não divisão, o uso do conjunto $\mathbb{Z}_{2^n}$ é questionável.

Suponha que nós queremos usar um bloco de 3-bits em nosso algoritmo de encriptação, usando apenas operações de adição e multiplicação. Então, as operações 8 são bem definidas, como mostra a Tabela 3.1. Porém, note que

na matriz (b) de multiplicação, os inteiros diferentes de zero não aparecem na mesma proporção de vezes. Por exemplo, existem apenas quatro ocorrências de 3, mas doze ocorrências de 4. Por outro lado, como foi dito existem corpos finitos da forma $GF(2^n)$, então existe em particular um corpo finito de ordem $2^3 = 8$. Isso é mostrado na Tabela 4.2, neste caso o número de ocorrências de inteiros diferentes de zero é uniforme, para resumir temos:

Inteiro	1	2	3	4	5	6	7
Ocorrências em $\mathbb{Z}_8$	4	8	4	12	4	8	4
Ocorrências em $GF(2^3)$	7	7	7	7	7	7	7

Por enquanto, vamos deixar de lado a questão de como as matrizes da Tabela 4.2 foram construídas e, em vez disso, vamos fazer algumas observações.

- **1.** As tabelas de adição e multiplicação são simétricas em relação a diagonal principal, em conformidade com a propriedade da comutatividade de adição e multiplicação. Esta propriedade é também exibida na Tabela 3.1, que usa mod 8.

- **2.** Todos os elementos não nulos possuem um inverso multiplicativo, como pode ser visto na matriz (b) da Tabela 4.2, ao contrário do caso da Tabela 3.1.

- **3.** O esquema definido pela Tabela 4.2 satisfaz todos requisitos para um corpo finito. Portanto, podemos nos referir a este esquema para $GF(2^3)$.

- **4.** Por convenção, nós vamos usar uma notação de 3 bits para cada um dos elementos de $GF(2^3)$.

Intuitivamente, um algoritmo que mapeie os inteiros de maneira desigual pode ser criptograficamente mais fraco do que aquele que fornece um ma-

peamento uniforme. Então, os corpos finitos da forma $GF(2^n)$ são atrativos para algoritmos criptográficos.

Podemos concluir que o ideal é procura um conjunto composto de 2^n elementos, juntamente com a definição de adição e multiplicação sobre o conjunto que define um corpo. Podemos atribuir um único inteiro no intervalo de 0 a $2^n - 1$ para cada elemento do conjunto.

	000	001	010	011	100	101	110	111
+	0	1	2	3	4	5	6	7
000	0	**0**	1	2	3	4	5	6	7
001	1	1	**0**	3	2	5	4	7	6
010	2	2	3	**0**	1	6	7	4	5
011	3	3	2	1	**0**	7	6	5	4
100	4	4	5	6	7	**0**	1	2	3
101	5	5	4	7	6	1	**0**	3	2
110	6	6	7	4	5	2	3	**0**	1
111	7	7	5	5	4	3	2	1	**0**

(a) Adição

	000	001	010	011	100	101	110	111
·	0	1	2	3	4	5	6	7
000	0	0	0	0	0	0	0	0	0
001	1	0	**1**	2	3	4	5	6	7
010	2	0	2	4	6	3	**1**	7	5
011	3	0	3	6	5	7	4	**1**	2
100	4	0	4	3	7	6	2	5	1
101	5	0	5	**1**	4	2	7	3	6
110	6	0	6	7	**1**	5	3	2	4
111	7	0	7	5	2	**1**	6	4	3

(b) Multiplicação

w	$-w$	w^{-1}
0	**0**	–
1	**1**	**1**
2	**2**	**5**
3	**3**	**6**
4	**4**	**7**
5	**5**	**2**
6	**6**	**3**
7	**7**	**4**

(c) Inversa

Tabela 4.2: Aritmética em $GF(2^3)$.

Tenha em mente que não usaremos aritmética modular porque como vimos isso não resulta em um corpo. Em vez disso, mostraremos como a aritmética polinomial fornece um meio para construir o corpo desejado.

Aritmética Polinomial Modular

Considere o conjunto S de todos os polinômios de grau $n - 1$ ou menor sobre o corpo $\mathbb{Z}_p$. Então, cada polinômio tem a forma

$$f(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1x + a_0 = \sum_{i=0}^{n-1} a_i x^i$$

na qual cada a_i assume um valor no conjunto $\{0, 1, \ldots, p-1\}$. Existe um total de p^n polinômios diferentes em S.

Exemplo 1. Os polinômios neste conjunto são:

Com a definição apropriada das operações aritméticas, cada conjunto S é um corpo finito. A definição consiste nos seguintes elementos:

0	x	$2x$
1	$x+1$	$2x+1$
2	$x+2$	$2x+2$

para $p = 3$ e $n = 2$ e $3^2 = 9$

0	x	x^2	x^2+x
1	$x+1$	x^2+1	x^2+x+1

para $p = 2$ e $n = 3$ e $2^3 = 8$

- **1.** A aritmética segue as regras da aritmética polinomial usando as regras básicas da álgebra.

- **2.** Aritmética sobre os coeficientes é realizada módulo p, ou seja, usamos as regras de aritmética para o corpo finito $\mathbb{Z}_p$.

- **3.** Se a multiplicação resultar em um polinômio de grau maior que $n-1$, então o polinômio é reduzido em módulo por algum polinômio irredutível fixado $m(x)$ de grau n, ou seja, dividimos por $m(x)$ e mantemos o resto. Para um polinômio $f(x)$, o resto é expresso como $r(x) = f(x) \bmod m(x)$.

Nós vamos chamar o conjunto $\{f(x) \mod m(x) \mid f(x) \in \mathbb{Z}_p[x]\}$ de polinômios modulares de m.

O AES para 8 bits usa aritmética no corpo finito $GF(2^8)$, com o polinômio irredutível $m(x) = x^8 + x^4 + x^3 + x + 1$. Considere os dois polinômios $f(x) = x^6 + x^4 + x^2 + x + 1$ e $g(x) = x^7 + x + 1$. Então

$$f(x) + g(x) = x^7 + x^6 + x^4 + x^2$$
$$f(x) \cdot g(x) = x^{13} + x^{11} + x^9 + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + 1$$

Logo, $f(x) \cdot g(x) \bmod m(x) = x^7 + x^6 + 1$.

Tal como acontece com a aritmética modular, temos a noção de classe resíduos, na aritmética polinomial modular. Se o polinômio $m(x)$ tem grau n, então a quantidade da classe resíduos módulo $m(x)$ consiste de p^n elementos.

A classe resíduos $[x+1]$, (mod $m(x)$), consiste em todos os polinômios $a(x)$ tais que $a(x) \equiv x + 1 \pmod{m(x)}$. Equivalentemente, a classe resíduos $[x + 1]$ consiste em todos os polinômios $a(x)$ que satisfazem a igualdade $a(x) \bmod m(x) = x + 1$.

Pode-se mostrar que o conjunto de todos os polinômios modulares de um polinômio de grau n irredutível $m(x)$ satisfaz os axiomas da seção 4.1 e, portanto, forma um corpo finito. Além disso, todos os corpos finitos de determinada ordem são isomorfos; isto é, quaisquer duas estruturas de corpo finito de uma determinada ordem são a mesma estrutura, mas a representação ou rótulos dos elementos podem ser diferentes.

Para construir o corpo finito de ordem $GF(2^3)$, nós precisamos escolher um polinômio irredutível de grau 3. Existem dois possíveis: $(x^3 + x^2 + 1)$ e $(x^3 + x + 1)$. Da Tabela 4.3 temos $GF(2^3)$ para o polinômio $m(x) = x^3 + x + 1$, mas poríamos escolher $m(x) = (x^3 + x^2 + 1)$. Note que este conjunto de tabelas tem a mesma estrutura que a Tabela 4.2. Assim, conseguimos encontrar um maneira de definir um corpo de ordem 2^3.
Agora podemos ler adições e multiplicações da tabela facilmente. Por exemplo, considere o binário $100 + 010 = 110$. Isto é equivalente a $x^2 + x$. Também considere $100 \cdot 010 = 011$, isto é equivalente a $x^2 \cdot x = x^3$ e reduzir para $x + 1$.

Encontrando o Inverso Multiplicativo em $GF(2^n)$

Assim como o Algoritmo de Euclides pode ser adaptado para encontrar o mdc de dois polinômios, o Algoritmo Euclides Estendido pode ser adaptado para encontrar o inverso de um polinômio.

Especificamente, o algoritmo encontrará o inverso de $b(x)$ módulo $a(x)$ se o grau de $b(x)$ for menor do que o grau de $a(x)$ e $mdc(a(x), b(x)) = 1$. Se $a(x)$ é um polinômio irredutível, então ele não tem nenhum fator diferente de si ou 1, portanto, temos que $mdc(a(x), b(x)) = 1$.

O algoritmo pode ser caracterizado da mesma maneira como fizemos com o Algoritmo de Euclides Estendido (AEE) para os inteiros. Dado os polinômios $a(x)$ e $b(x)$ com o grau de $a(x)$ maior do que o de $b(x)$, nós desejamos resolver a seguinte equação para os valores de $v(x)$, $w(x)$, e $d(x)$, onde $d(x) = mdc(a(x), b(x))$:

$$a(x)v(x) + b(x)w(x) = d(x)$$

Se $d(x) = 1$, então temos que $w(x)$ é um inverso de $b(x)$ módulo $a(x)$. Da Tabela 4.3 temos:

	000	001	010	011	100	101	110	111
+	0	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
000	0	**0**	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
001	1	1	**0**	$x+1$	x	x^2+1	x^2	x^2+x+1	x^2+x
010	x	x	$x+1$	**0**	1	x^2+x+1	x^2+x	x^2+1	x^2
011	$x+1$	$x+1$	x	1	**0**	x^2+x+1	x^2+x	x^2+1	x^2
100	x^2	x^2	x^2+1	x^2+x	x^2+x+1	**0**	1	x	$x+1$
101	x^2+1	x^2+1	x^2	x^2+x+1	x^2+x	1	**0**	$x+1$	x
110	x^2+x	x^2+x	x^2+x+1	x^2	x^2+1	x	$x+1$	**0**	1
111	x^2+x+1	x^2+x+1	x^2+x	x^2+1	x^2	$x+1$	x	1	**0**

(a) Adição

	000	001	010	011	100	101	110	111
·	0	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
000	0	0	0	0	0	0	0	0	0
001	1	0	**1**	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
010	x	0	x	x^2	x^2+x	$x+1$	**1**	x^2+x+1	x^2+1
011	$x+1$	0	$x+1$	x^2+x	x^2+1	x^2+x+1	x^2	**1**	x
100	x^2	0	x^2	$x+1$	x^2+x+1	x^2+x	x	x^2+1	1
101	x^2+1	0	x^2+1	**1**	x^2	x	x^2+x+1	$x+1$	x^2
110	x^2+x	0	6	7	**1**	5	3	2	4
111	x^2+x+1	0	x^2+x+1	x^2+1	x	**1**	x^2+1	x^2	$x+1$

(b) Multiplicação

Tabela 4.3: Aritmética Polinomial Modular $x^3 + x + 1$.

A Tabela 4.4 apresenta o calculo da inversa de $(x^2 + x + 1) \mod (x^8 + x^4 + x^3 + x + 1)$. O resultado é $(x^7 + x + 1)^{-1} = (x^7)$. Logo, temos que $(x^7 + x + 1)(x^7) \equiv 1 \pmod{(x^8 + x^4 + x^3 + x + 1)}$.

Considerações Computacionais

Se o polinômio $f(x)$ pertence a $GF(2^n)$,

$$f(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1x + a_0 = \sum_{i=0}^{n-1} a_i x^i,$$

ele pode ser representado exclusivamente pela sequência de seus n coeficientes $(a_{n-1}, a_{n-2}, \ldots, a_0)$. Então, todo polinômio contido em $GF(2^n)$ pode ser representado por um número $n - bits$.

As Tabelas 4.2 e 4.3 apresentam as operações de adição e multiplicação para $GF(2^3)$ módulo $m(x) = (x^3+x+1)$. A Tabela 4.2 usa a representação binária, e a Tabela 4.3 usa a representação polinomial.

Adição: A soma de polinômios é realizada pela adição dos coeficientes correspondentes, e no caso de polinômios sobre $\mathbb{Z}_p$, além disso, é uma operação XOR. Então, a adição de dois polinômios em $GF(2^n)$ corresponde a uma operação XOR bit-a-bit.

Considere os dois polinômios pertencentes a $GF(2^8)$ em nosso exemplo anterior: $f(x) = x^6 + x^4 + x^2 + x + 1$ e $g(x) = x^7 + x + 1$, além disso, considere as formas de soma: **(1)** polinomial, **(2)** binária e **(3)** hexadecimal [2] respectivamente:

[2] Aqui cada um dos dois grupos de 4 bits em um byte é denominado por um único caractere hexadecimal, e os dois caracteres são incluídos em colchetes.

- **(1)** $(x^6 + x^4 + x^2 + x + 1) + (x^7 + x + 1) = x^7 + x^6 + x^4 + x^2 + 2x + 2$
- **(2)** $(01010111) \oplus (10000011) = (11010100)$
- **(3)** $\{57\} \oplus \{83\} = \{D4\}$

Multiplicação: Não existe uma operação XOR simples que realize multiplicação em $GF(2^n)$, porém, ela é razoavelmente e fácil de implementar com um computador. Vamos discutir a técnica usando o polinômio $m(x) = x^8 + x^4 + x^3 + x + 1$ e corpo finito $GF(2^8)$ que é usado no AES. A técnica é baseada na observação de que

$$x^8 \bmod m(x) = m(x) - x^8 = (x^4 + x^3 + x + 1) \tag{4.2}$$

Inicialização	$a(x) = x^8 + x^4 + x^3 + x + 1; \quad v_{-1}(x) = 1; w_{-1}(x) = 0$ $b(x) = x^7 + x + 1; v_0(x) = 0; \quad w_0(x) = 1$
Iteração 1	$q_1(x)x; \quad r_1(x) = x^4 + x^3 + x^2 + 1$ $v_1(x) = 1; \quad w_1(x) = x$
Iteração 2	$q_2(x) = x^3 + x^2 + 1; \quad r_2(x) = x$ $v_2(x) = x^3 + x^2 + 1; \quad w_2(x) = x^4 + x^3 + x + 1$
Iteração 3	$q_3(x) = x^3 + x^2 + x; \quad r_3 = 1$ $v_3 = x^6 + x^2 + x + 1; \quad w_3(x) = x^7$
Iteração 4	$q_4(x) = x; \quad r_4(x) = 0$ $v_4(x) = x^7 + x + 1; \quad w_4(x) = x^8 + x^4 + x^3 + x + 1$
Resultado	$d(x) = r_3(x) = mdc(a(x), b(x)) = 1$ $w(x) = w_3(x) = (x^7 + x + 1)^{-1} \mod (x^8 + x^4 + x^3 + x + 1) = x^7$

Tabela 4.4: Euclides Estendido $[(x^8 + x^4 + x^3 + x + 1), (x^7 + x + 1)]$

A equação (4.2) pode ser verificada como exemplo de uma relação mais geral que sempre é é verdadeira para $GF(2^n)$; Se $p(x)$ tem grau n, então

$x^n \bmod p(x) = p(x) - x^n$.

Agora, vamos considerar um polinômio em $GF(2^8)$, que tem a forma $f(x) = b_7x^7 + b_6x^6 + b_5x^5 + b_4x^4 + b_3x^3 + b_2x^2 + b_1x + b_0$. Se nós multiplicarmos por x, nós temos

$$\begin{aligned} x \cdot f(x) = \quad & (b_7x^8 + b_6x^7 + b_5x^6 + b_4x^5 + b_3x^4 \\ & + b_2x^3 + b_1x^2 + b_0x) \mod m(x) \end{aligned} \tag{4.3}$$

Se $b_7 = 0$, então o resultado é um polinômio de menor grau do que 8, que já está em forma reduzida, e computacionalmente não faz diferença. Se $b_7 = 1$, então redução do módulo $m(x)$ é obtida usando a equação (4.2):

$$\begin{aligned} x \cdot f(x) = \quad & (b_6x^7 + b_5x^6 + b_4x^5 + b_3x^4 + b_2x^3 + b_1x^2 + b_0x) \\ & + (x^4 + x^3 + x + 1) \end{aligned}$$

Segue-se que a multiplicação por x (isto é, 00000010) pode ser implementada como um deslocamento um 1-bit seguido por um XOR condicional com (00011011), o qual representa $(x^4 + x^3 + x + 1)$. Resumindo,

$$x \cdot f(x) = \begin{cases} (b_6b_5b_4b_3b_3b_1b_00), & \text{se } t = 1 \\ (b_6b_5b_4b_3b_3b_1b_00) \oplus (00011011), & \text{se } b_7 = 1 \end{cases} \tag{4.4}$$

multiplicando pela maior potência de x atingida por repetidas aplicações da equação (4.4) e adicionando os resultados intermediários por qualquer constante em $GF(2^8)$.

Num exemplo anterior, nós vimos que para $f(x) = x^6 + x^4 + x^2 + x + 1$, $g(x) = x^7 + x + 1$, $m(x) = x^8 + x^4 + x^3 + x + 1$, nós temos $f(x) \cdot g(x)$ mod $m(x) = x^7+x^6+1$. Refazendo esta conta utilizando a notação decimal, nós precisamos calcular $(01010111) \cdot (10000011)$. Primeiro, determinamos o resultado da multiplicação para potências de x:

$$
\begin{aligned}
(01010111) \cdot (00000010) &= (10101110) \\
(01010111) \cdot (00000100) &= (01011100) \oplus (00011011) \quad = (01000111) \\
(01010111) \cdot (00001000) &= (10001110) \\
(01010111) \cdot (00010000) &= (00011100) \oplus (00011011) \quad = (00000111) \\
(01011011) \cdot (00100000) &= (00001110) \\
(01010111) \cdot (01000000) &= (00011100) \\
(01010111) \cdot (10000000) &= (00111000)
\end{aligned}
$$

Então,

$$
\begin{aligned}
(01010111) \cdot (10000011) &= (01010111) \oplus (10101110) \oplus (00111000) \\
&= (11000001)
\end{aligned}
$$

que é equivalente a $x^7 + x^6 + 1$.

Usando Gerador

Apresentaremos uma técnica equivalente para construir um corpo finito da forma $GF(2^n)$, usando o mesmo polinômio irredutível, que é mais conveniente. Para começar, precisamos de duas definições:

Um **gerador** g de um corpo finito F de ordem q (contendo q elementos) é um elemento cujas primeiras $q-1$ potências geram todos os elementos não

nulos de F. Isto é, os elementos de F consistem em $\{0, g^0, g^1, \ldots, g^{q-2}\}$.

Seja um corpo F cujas operações são definidas módulo $f(x)$, onde $f(x)$ é um polinômio irredutível como vimos anteriormente. Um elemento b contido em F é chamado de **raiz** do polinômio se $f(b) = 0$.

Por fim, pode ser mostrado que uma raiz g de um polinômio irredutível é um gerador de um corpo finito cujas operações são definidas módulo esse polinômio.

Considere o corpo finito $GF(2^3)$, cujas operações são definidas módulo $f(x) = x^3 + x + 1$. Então, o gerador g deve satisfazer $f(g) = g^3 + g + 1 = 0$. Tenha em mente, como discutido anteriormente, que não precisamos encontrar uma solução numérica para igualdade. Em vez disso, lidamos com aritmética polinomial em que aritmética nos coeficientes é realizada módulo 2. Assim sendo, a solução para seguinte equação é $g^3 = -g - 1 = g + 1$. É possível perceber que g de fato gera todos os polinômios de grau menor que 3, pois:

$$
\begin{aligned}
g^4 &= g(g^3) = g(g+1) = g^2 + g \\
g^5 &= g(g^4) = g(g^2+g) = g^3 + g^2 = g^2 + g + 1 \\
g^6 &= g(g^5) = g(g^2+g+1) = g^3 + g^2 + g = g^2 + 2g + 1 = g^2 + 1 \\
g^7 &= g(g^6) = g(g^2+1) = g^3 + g = 2g + 1 = 1 + g^0
\end{aligned}
$$

Vimos que as potências de g geram todos os polinômios não negativos em $GF(2^3)$. Além disso, deve ser claro que $g^k = g^{k \bmod 7}$ para qualquer inteiro k. A Tabela 4.5 apresenta os resultados do gerador de $GF(2^3)$ usando o polinômio $f(x) = x^3 + x + 1$.

Esta representação por potências torna a multiplicação fácil. Para multiplicar na notação de potência, basta adicionar ao expoente módulo 7. Por exemplo, $g^4 + g^6 = g^{10 \bmod 7} = g^3 = g + 1$. O mesmo resultado é alcançado

Potência	***Polinomial***	***Binaria***	***Decimal (Hex)***
0	0	000	0
$g^0(=g^7)$	1	001	1
g^1	g	010	2
g^2	g^2	100	4
g^3	$g+1$	011	3
g^4	g^2+g	110	6
g^5	g^2+g+1	111	7
g^6	g^2+1	101	5

Tabela 4.5: Gerador de $GF(2^3)$ usando x^3+x+1

usando aritmética polinomial:

Temos que $g^4 = g^2+g$ e $g^6 = g^2+1$, então, $(g^2+g)\cdot(g^2+1) = g^4+g^3+g^2+g$, em seguida precisamos determinar $(g^4+g^3+g^2+g) \mod (g^3+g+1)$. Nós obtemos, então, o resultado $g+1$, o mesmo obtido usando a notação de potência.

A Tabela 4.6 apresenta a adição e a multiplicação para $GF(2^3)$ usando a notação de potência. Note que isso produz os resultados idênticos para a representação polinomial (Tabela 4.3) com algumas linhas e colunas intercambiadas.

	000	001	010	100	011	110	111	101
+	0	1	G	g^2	g^3	g^4	g^5	g^6
000	0	**0**	1	G	g^2	$g+1$	g^2+g	g^2+g+1	g^2+1
001	1	1	**0**	$g+1$	g^2+1	g	g^2+g+1	g^2+g	g^2
010	g	g	$g+1$	**0**	g^2+g	1	g^2	g^2+1	g^2+g+1
100	g^2	g^2	g^2+1	g^2+g	**0**	g^2+g+1	g	$g+1$	1
011	g^3	$g+1$	g	1	g^2+g+1	**0**	g^2+1	g^2	g^2+g
110	g^4	g^2+g	g^2+g+1	g^2	g	g^2+1	**0**	1	$g+1$
111	g^5	g^2+g+1	g^2+g	g^2+1	$g+1$	g^2	1	**0**	g
101	g^6	g^2+1	g^2	g^2+g+1	1	g^2+g	$g+1$	g	**0**

(a) Adição

	000	001	010	100	011	110	111	101
·	0	1	G	g^2	g^3	g^4	g^5	g^6
000	0	0	0	0	0	0	0	0	0
001	1	0	**1**	G	g^2	$g+1$	g^2+g	g^2+g+1	g^2+1
010	g	0	g	g^2	$g+1$	g^2+g	g^2+g+1	g^2+1	**1**
100	g^2	0	g^2	$g+1$	g^2+g	g^2+g+1	g^2+1	**1**	g
011	g^3	0	$g+1$	g^2+g	g^2+g+1	g^2+1	**1**	g	g^2
110	g^4	0	g^2+g	g^2+g+1	g^2+1	**1**	g^2+g+1	g^2	$g+1$
111	g^5	0	g^2+g+1	g^2+1	**1**	g	g^2	$g+1$	g^2+g
101	g^6	0	g^2+1	**1**	g	g^2	$g+1$	g^2+g	g^2+g+1

(b) Multiplicação

Tabela 4.6: $GF(2^3)$ Aritmética usando geradores para o polinômio (x^3+x+1).

4.4 Resumo

Neste capítulo, definimos o que são grupos, anéis e suas propriedades. Definimos o que são corpos, como construir um corpo finito de ordem p, onde p é primo. Especificamente, definimos o conjunto $GF(p)$ com as propriedades:

- **1.** $GF(p)$ consiste em p elementos.
- **2.** As operações + e $\cdot$ são definidas sobre o corpo finito $GF(p)$. As operações de adição, multiplicação podem ser executadas sem sair deste conjunto. Cada elemento do conjunto diferente de 0 tem um inverso.
- **3.** São validas as propriedades de corpo, de domínio de integridade e as propriedades de anel.

Nós vimos que os elementos de $GF(p)$ são as classes de resíduos representadas pelos inteiros $\{0, 1, \ldots, p-1\}$ e que as operações aritméticas são adição e multiplicação mod p.

Além disso, apresentamos como construir um corpo finito de ordem 2^n. Especificamente, definimos o conjunto $GF(2^n)$ com as propriedades:

- **1.** $GF(2^n)$ consiste de 2^n elementos.
- **2.** As operações $+$ e $\cdot$ são bem definidas sobre um corpo finito $GF(2^n)$. As operações de adição, multiplicação podem ser executadas sem sair dele. Cada elemento do conjunto diferente de 0 tem um inverso.
- **3.** São validas as propriedades de corpo, de domínio de integridade e as propriedades de anel.

Vimos que os elementos de $GF(2^n)$ podem ser definidos como o conjunto das classes de resíduos módulo $m(x)$ dos polinômios com coeficientes em $\mathbb{Z}_2$ de grau $n-1$ ou menor e o nulo, onde $m(x)$ é um polinômio irredutível de grau n. E também que eles podem ser representados por um valor com n-bits.

Também vimos que uma definição equivalente de um corpo finito $GF(2^n)$ faz o uso de um gerador e que a aritmética é definida usando potências do gerador.

Capítulo 5

Curvas Elípticas

Neste capítulo, apresentamos definições fundamentais sobre curvas elípticas que servirão de base para o capítulo 6. Mais detalhes podem ser encontrados no livro de Anthony W. Knapp e Neal Koblitz, "Elliptic Curves" [2] e no de William Stallings, "Cryptography and Network Security" [18] que foram utilizados como referências.

Uma **curva elíptica** é uma curva algébrica [1] plana definida por uma equação não-singular da forma:

$$y^2 = x^3 + ax + b. \tag{5.1}$$

Formalmente, seja E uma curva elíptica sobre um corpo K, E é uma curva cúbica não-singular em duas variáveis $f(x, y) = 0$ com um ponto racional k (que deve ser um "ponto no infinito" O) da curva no plano projetivo. O corpo K pode ser, por exemplo, o conjunto dos números reais.

Para compreender melhor esta definição precisamos revisar o conceito de grupo abeliano. Em seguida, examinaremos o conceito de curvas elípticas

[1] ANTHONY W. KNAPP. A Course in Number Theory and Cryptography. Springer, 1994

definidas sobre o conjunto dos números reais. Isto será seguido por um olhar sobre curvas elípticas definidas em corpos finitos. Finalmente, nós seremos capazes de examinar codificadores de curvas elípticas. Na seção 4.1 falamos de **grupos abelianos**, vamos relembrá-los:

Grupos Abelianos

Um **grupo abeliano** é um conjunto, G, juntamente com uma operação [2] $\bullet$ (chamada lei do grupo G) que combina quaisquer dois elementos a e b para formar outro elemento, denotado por $a \bullet b$. Para se qualificar como um grupo, o conjunto e a operação, $\{G, \bullet\}$, devem satisfazer quatro requisitos conhecidos como axiomas do grupo:

- **(A1) Operação fechada em G:** Para todo a e $b \in G$, o resultado da operação, $a \bullet b$, também está em G.
- **(A2) Associatividade:** Para todo a, b e $c \in G$, $a \bullet (b \bullet c) = (a \bullet b) \bullet c$.
- **(A3) Identidade:** Há um elemento $e \in G$ tal que $a \bullet e = e \bullet a = a$ para todo $a \in G$.
- **(A4) Inverso:** Para cada $a \in G$, existe um elemento $a' \in G$, comumente denotado a^{-1} (ou $-a$, se a operação for denotada $+$) tal que $a \bullet a' = a' \bullet a = e$, onde e é o elemento identidade.
- **(A5) Comutatividade:** $a \bullet b = b \bullet a$, $\forall a, b \in G$.

Uma **curva elíptica** é definida por uma equação em duas variáveis e seus respectivos coeficientes como já observamos na equação (5.1). Para criptografia, as variáveis e coeficientes são restritos a elementos em um corpo

[2] O operador $\bullet$ é genérico e pode ser referir a adição, multiplicação, ou qualquer outro operador matemático.

finito. Antes de discutirmos sobre isto, nós vamos olhar as curvas elípticas nas quais as variáveis e coeficientes são números reais. Este caso é mais fácil de compreender por enquanto.

5.1 Curvas Elípticas sobre $\mathbb{R}$

Curvas Elípticas **não são elipses**. Elas são assim chamadas porque eram **descritas por equações cúbicas**, similares àquelas usadas para calcular a comprimento de uma elipse. No geral, equações cúbicas para curvas elípticas assumem a seguinte forma, conhecida como **Equação de Weierstrass**:

$$y^2 + axy + by = x^3 + cx^2 + dx + e$$

onde $a, b, c, d, e \in \mathbb{R}$ e x e y também assumem valores nos números reais [3], para o nosso propósito, é suficiente nos limitarmos a equação (5.1).

Tal equação é dita cúbica, ou de grau 3 porque seu maior expoente é o 3, como já vimos na definição de uma curva elíptica, o O é um elemento chamado de *"ponto no infinito"* ou o *ponto zero*, nós vamos discutir isto em breve. Para desenhar esta curva, precisamos resolver esta equação:

$$y = \sqrt{x^3 + ax + b}$$

[3]Note que x e y são variáveis, que assumem valores. Isto está em contraste com nossa discussão sobre anéis e corpos polinomiais no capítulo 3, em que x foi tratado como indeterminado.

Para valores fixos de a e b, o gráfico consiste em valores positivos e negativos de y para cada valor de x. Então, cada curva é simétrica sobre $y = 0$.

Defina o conjunto dos pontos da curva elíptica $E(a, b)$ consistindo em todos os pontos (x, y) que satisfazem a equação (5.1) juntamente com o elemento O.

Portanto, usar um valor diferente do par (a, b) resulta em um conjunto diferente de $E(a, b)$. Podemos gerar diferentes curvas conforme retratado na figura 5.1. Lembremos que a definição da curva elíptica também requer que a curva seja não-singular.

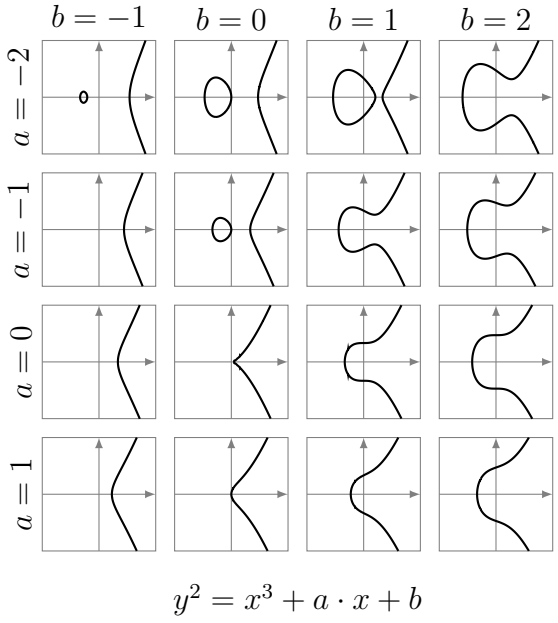


Figura 5.1: Exemplos de curvas elípticas

Descrição Geométrica

Podemos dizer que *um grupo* pode ser definido baseado no conjunto $E(a, b)$ para valores específicos de a e b na equação (5.1), se a seguinte condição for atendida:

$$4a^3 + 27b^2 \neq 0. \tag{5.2}$$

Para definir o grupo, nós devemos definir uma operação, que vamos chamar de adição e denotar por $+$, para o conjunto $E(a, b)$, onde a e b satisfazem a equação (5.2). Em termos geométricos, as regras para adição podem ser declaradas da seguinte forma: Se três pontos em uma curva elíptica estão em linha reta, então sua soma é O. A partir desta definição, nós podemos definir as regras de adição de uma curva elíptica.

- **1.** O serve como identidade aditiva. Portanto, $O = -O$ e para qualquer ponto P na curva elíptica, $P + O = P$. No que segue, nós assumimos que $P \neq O$ e $Q \neq O$.

- **2.** O oposto de um ponto P é o ponto com a mesma coordenada x, porém, o oposto da coordenada y; Isso é, se $P = (x, y)$, então $-P = (x, -y)$. Note que estes dois pontos e ponto no infinito podem ser unidos por uma linha vertical, isto é, $P + (-P) + O = P - P + O = O$.

- **3.** Para somar dois pontos P e Q com diferentes valores da coordenada x, deve-se desenhar uma linha reta entre estes dois pontos e encontrar o terceiro ponto R na intersecção com a curva (a menos que a linha seja tangente à curva em P ou Q, neste caso nós temos que $R = P$

ou $R = Q$, respectivamente). Para formar uma estrutura de grupo, nós definimos a adição como: $P + Q = -R$, ou seja, nós definimos $P + Q$ como a imagem espelhada (com respeito ao eixo x) do terceiro ponto de intersecção.

- **4.** A interpretação geométrica do item anterior também se aplica a dois pontos, P e $-P$, com a mesma coordenada x. Os pontos são unidos por uma linha vertical, que pode ser visto também como a interseção da curva no ponto infinito. Por isso, temos que $P + (-P) = O$, que é consistente com o item (2).

- **5.** Para multiplicar um ponto Q por 2, desenhe a linha tangente e encontre o outro ponto de interseção S. Então $Q + Q = 2Q = -S$.

Com a lista de regras acima, pode-se mostrar que o conjunto $E(a, b)$ é um grupo abeliano.

Descrição Algébrica

Nós apresentamos alguns resultados que nos possibilitam o cálculo da adição sobre curvas elípticas. [4] Para dois pontos distintos, $P = (x_P, y_P)$ e $Q = (x_Q, y_Q)$, que são diferentes de zero, a inclinação da reta r que os une é $\Delta = (y_Q - y_P)/(x_Q - x_P)$. Existe exatamente um outro ponto no qual r intercepta a curva elíptica, e que é o oposto da soma de P e Q. Após algumas manipulações algébricas, podemos expressar a soma $R = P + Q$ da seguinte maneira:

[4]Para se aprofundar nesta explicação veja mais detalhes no livro referência [2] desta seção.

$$\begin{aligned} x_R &= \Delta^2 - x_P - x_Q \\ y_R &= -y_P + \Delta(x_P - x_R) \end{aligned} \tag{5.3}$$

Nós também precisamos ser capazes de somar um ponto a si mesmo: $P + P = 2P = R$. Quando $y_P \neq 0$, as expressões são:

$$\begin{aligned} x_R &= \left(\frac{3x_P^2 + a}{2y_P}\right)^2 - 2x_P \\ y_R &= \left(\frac{3x_P^2 + a}{2y_P}\right)(x_P - x_P) - y_P \end{aligned} \tag{5.4}$$

5.2 Curvas Elípticas sobre $\mathbb{Z}_p$

A **Criptografia de Curva Elíptica** faz uso de curvas elípticas em que as variáveis e coeficientes são todos restritos a elementos de um corpo finito. Duas famílias de curvas elípticas são usadas em aplicações criptográficas: curvas primárias sobre $\mathbb{Z}_p$ e curvas binárias sobre $GF(2^m)$ [6]. Para uma **curva primária** definida sobre $\mathbb{Z}_p$, usamos uma equação cúbica na qual as variáveis e coeficientes todos assumem valores no conjunto de inteiros de 0 a $p-1$ e realizamos cálculos em módulo p. Para uma **curva binária** definida sobre $GF(2^m)$, as variáveis e coeficientes assumem valores em $GF(2^m)$ e os cálculos são realizados sobre $GF(2^m)$, além disso, salientamos que as curvas primárias são melhores para aplicações de software pois as operações estendidas de processamento ("bit-fiddling") não são necessárias, mas são necessárias por curvas binárias; e que as curvas binárias são melhores para aplicações de hardware, nas quais são necessárias pouquíssimas portas lógicas para criar um sistema criptográfico poderoso e rápido. Examinamos essas duas famílias

nesta seção e na próxima.

Existe interpretação geométrica das curvas elípticas sobre os corpos finitos. A interpretação algébrica usada para a aritmética de uma curva elíptica sobre os números reais é facilmente transferida, e esta é a abordagem que tomamos.

Para curvas elípticas sobre $\mathbb{Z}_p$, assim como nos números reais, nos limitamos à equação da forma (5.1), porém neste caso tanto os coeficientes quanto as variáveis limitam-se a $\mathbb{Z}_p$:

$$y^2 \bmod p = (x^3 + ax + b) \bmod p \tag{5.5}$$

Por exemplo, a equação (5.5) é satisfeita para $a = 1$, $b = 1$, $x = 9$, $y = 7$ e $p = 23$

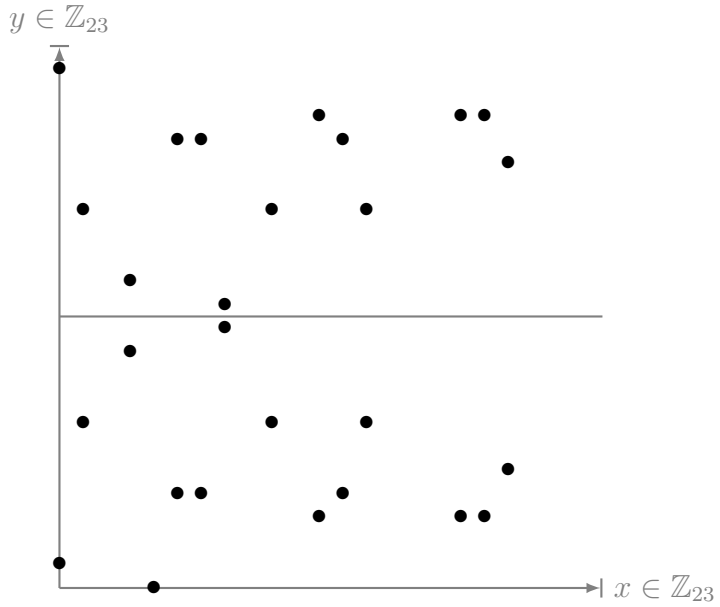
$$7^2 \bmod 23 = (9^3 + 9 + 1) \bmod 23$$
$$49 \bmod 23 = 739 \bmod 23$$
$$3 = 3$$

Agora considere o conjunto $E_p(a, b)$, que consiste de todos os pares $(x, y) \in \mathbb{Z}_p \times \mathbb{Z}_p$ que satisfazem a equação (5.5), junto com um ponto no infinito O, onde os coeficientes a, b também são elementos de $\mathbb{Z}_p$ e, neste caso, dizemos que $E_p(a, b)$ é dado pela equação (5.5).

Por exemplo, seja $p = 23$ e considere uma curva elíptica $y^2 = x^3 + x + 1$. Para o conjunto $E_{23}(1, 1)$, estamos interessados apenas nos inteiros não negativos no quadrante de $(0, 0)$ a $(p - 1, p - 1)$ que satisfazem a equação mod p. A tabela 5.1 lista os pontos (exceto o O) que são parte de $E_{23}(1, 1)$. A figura 5.2 plota os pontos de $E_{23}(1, 1)$; note que os pontos, com uma exceção, são simétricos em relação a $y = 11.5$.

$(0,1)$	$(6,4)$	$(12,19)$
$(0,22)$	$(6,19)$	$(13,7)$
$(1,7)$	$(7,11)$	$(13,16)$
$(1,6)$	$(7,12)$	$(17,3)$
$(3,10)$	$(9,7)$	$(17,20)$
$(3,13)$	$(9,16)$	$(18,3)$
$(4,0)$	$(11,3)$	$(18,20)$
$(5,4)$	$(11,20)$	$(19,5)$
$(5,19)$	$(12,4)$	$(19,18)$

Tabela 5.1: Pontos da curva elíptica $E_{23}(1,1)$



$$y^2 = x^3 + x + 1 \in \mathbb{Z}_{23}$$

Figura 5.2: A curva elíptica $E_{23}(1, 1)$

Pode ser demonstrado que um grupo abeliano finito, pode ser definido com base no conjunto $E_p(a, b)$ dado que $(x^3 + ax + b) \bmod p$ não tem fatores repetidos, isto é equivalente à condição:

$$(4a^3 + 27b^2) \bmod p \neq 0 \bmod p. \tag{5.6}$$

Observe que a equação (5.6) tem a mesma forma que a equação (5.2).

A regra para adição sobre $E_p(a, b)$ corresponde a técnicas algébricas descritas para curvas elípticas definidas sobre os números reais. Para todos os pontos $p, Q \in E_p(a, b)$:

- **1.** $P + O = P$

- **2.** Se $P = (x_P, y_P)$, então $P + (x_P, -y_P) = O$. O ponto $(x_P, -y_P)$ é o oposto de P, denotado por $-P$, por exemplo, em $E_{23}(1,1)$ para $P = (13,7)$, nós temos que $-P = (13,-7)$. Porém, $-7 \bmod 23 = 16$, por isso, $-P = (13,16)$, que também está em $E_{23}(1,1)$.

- **3.** Se $P = (x_P, y_P)$ e $Q = (x_Q, y_Q)$ com $P \neq -Q$, então $R = P + Q = (x_R, y_R)$ é determinado pelas seguintes regras:

$$x_R = (\lambda^2 - x_P - x_Q) \bmod p$$
$$y_R = (\lambda(x_P - x_R) - y_P) \bmod p$$

onde

$$\lambda = \begin{cases} \left(\frac{y_Q - y_P}{x_Q - x_P}\right) \bmod p, & \text{se } P \neq Q \\ \left(\frac{3x_P^2 + a}{2y_P}\right) \bmod p, & \text{se } P = Q \end{cases}$$

- **4.** Multiplicação é definida por repetidas somas, por exemplo, $4P = P + P + P + P$. Seja $P = (3,10)$ e $Q = (9,7)$ em $E_{23}(1,1)$, então,

$$\lambda = \left(\frac{7-10}{9-3}\right) \bmod 23 = \left(\frac{-3}{6}\right) \bmod 23 = \left(\frac{-1}{2}\right) \bmod 23 = 11$$
$$x_R = (11^2 - 3 - 9) \bmod 23 = 109 \bmod 23 = 17$$
$$y_R = (11(3-17) - 10) \bmod 23 = -164 \bmod 23 = 20$$

Logo, $P + Q = (17,20)$. Para encontrar $2P$, temos que

$$\lambda = \left(\frac{3(3^2) - 1}{2 \times 10}\right) \bmod 23 = \left(\frac{5}{20}\right) \bmod 23 = \left(\frac{1}{4}\right) \bmod 23 = 6$$

Para encontrar o inverso multiplicativo no passo 4 na equação anterior em $\mathbb{Z}_{23}$, podemos usar o Algoritmo Euclides Estendido definido na seção 4.1. Para confirmar, note que $(6 \cdot 4) \bmod 23 = 24 \bmod 23 = 1$.

$$x_R = (6^2 - 3 - 3) \bmod 23 = 30 \bmod 23 = 7$$
$$y_R = (6(3 - 7) - 10) \bmod 23 = (-34) \bmod 23 = 12$$

e $2P = (7, 12)$. Para determinar a segurança das várias cifras de curvas elípticas, é interessante saber a quantidade de pontos em um grupo abeliano finito definido a partir de uma curva elíptica. No caso do grupo finito $E_p(a, b)$, o número de pontos N é delimitado por

$$p + 1 - 2\sqrt{p} \leq N \leq p + 1 + 2\sqrt{p}$$

Observe que o número de pontos em $E_p(a, b)$ é, aproximadamente, igual ao números de elementos em $\mathbb{Z}_p$, isto é, p elementos.

5.3 Curvas Elípticas sobre $GF(2^m)$

Lembre-se de que no capítulo 4 falamos que um **corpo finito** da forma $GF(2^m)$ consiste de 2^m elementos, juntamente com operações de adição e multiplicação que podem ser definidas sobre polinômios. Para curvas elípticas sobre $GF(2^m)$, nós usamos uma equação cúbica na qual as variáveis e os coeficientes assumem valores em $GF(2^m)$ para algum número m no qual os cálculos são realizados usando as regras de aritmética em $GF(2^m)$.

Acontece que a forma da equação cúbica apropriada para aplicações criptográficas para curvas elípticas é um pouco diferente em $GF(2^m)$ do que a

forma usada em $\mathbb{Z}_p$. A forma apropriada é esta:

$$y^2 + xy = x^3 + ax^2 + b, \tag{5.7}$$

$(0, 1)$	(g^5, g^3)	(g^9, g^{13})
$(1, g^6)$	(g^5, g^{11})	(g^{10}, g)
$(1, g^{13})$	(g^6, g^8)	(g^{10}, g^8)
(g^3, g^8)	(g^6, g^{14})	$(g^{12}, 0)$
(g^3, g^{13})	(g^9, g^{10})	(g^{12}, g^{12})

Tabela 5.2: Pontos na Curva Elíptica $E_{2^4}(g^4, 1)$

onde se compreende que as variáveis x e y e os coeficientes a e b são elementos de $GF(2^m)$ e que os cálculos são realizados em $GF(2^m)$.

Agora vamos considerar o conjunto $E_{2^m}(a, b)$ consistindo de todos os pares de inteiros (x, y) que satisfazem a equação (5.7), junto com os pontos no infinito O.

Por exemplo, vamos usar o corpo finito $GF(2^4)$ com o polinômio irredutível $f(x) = x^4 + x + 1$. Isso produz um gerador g que satisfaz $f(g) = 0$ com o valor de $g^4 = g + 1$, ou em binário, $g = 0010$. Podemos desenvolver as as potências de g da seguinte maneira:

$g^0 = 0001$	$g^4 = 0011$	$g^8 = 0101$	$g^{12} = 1111$
$g^1 = 0010$	$g^5 = 0110$	$g^9 = 1010$	$g^{13} = 1101$
$g^2 = 0100$	$g^6 = 1100$	$g^{10} = 0111$	$g^{14} = 1001$
$g^3 = 1000$	$g^7 = 1011$	$g^{11} = 1110$	$g^{15} = 0001$

Por exemplo, $g^5 = (g^4)(g) = g^2 + g = 0101$.

Agora considere a curva elíptica $y^2 + xy = x^3 + g^4x^2 + 1$. Neste caso, $a = g^4$ e $b = g^0 = 1$. Um ponto que satisfaz esta equação é (g^5, g^3):

$$(g^3)^2 + (g^5)(g^3) = (g^5)^3 + (g^4)(g^5)^2 + 1$$
$$g^6 + g^8 = g^{15} + g^{14} + 1$$
$$1100 + 0101 = 0001 + 1001 + 0001$$
$$1001 = 1001$$

A tabela 5.2 lista os pontos (com exceção do O) que são parte de $E_{2^4}(g^4, 1)$, a figura 5.3 plota os pontos de $E_{2^4}(g^4, 1)$.

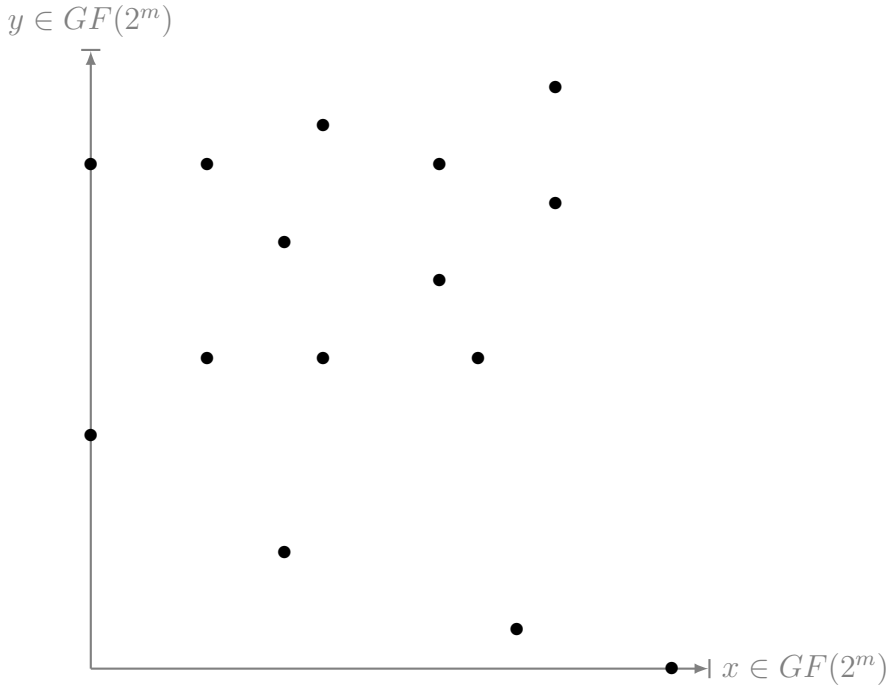


Figura 5.3: A Curva Elíptica $E_{2^4}(g^4, 1)$

Pode ser mostrado que um grupo abeliano finito podem ser definido com base no conjunto $E_{2^m}(a, b)$, desde que $b \neq 0$. As regras para adição pode ser

declaradas da seguinte forma. Para todos os pontos P, $Q \in E_{2^m}(a, b)$:

- **1.** $P + O = P$.
- **2.** Se $P = (x_P, y_P)$, então $P + (x_P, x_P + y_P) = O$. O ponto $(x_P, x_P + y_P)$ é o oposto de P, isto é, denotamos por $-P$.
- **3.** Se $P = (x_P, y_P)$ e $Q = (x_Q, y_Q)$ com $P \neq -Q$, então $P \neq Q$ e $R = P + Q = (x_R, y_R)$ é determinado pelas seguintes regras:

$$x_R = \lambda^2 + \lambda + x_P + x_Q + a$$
$$y_R = \lambda(x_P - x_R) + x_R + y_P$$

  onde

$$\lambda = \left(\frac{y_Q + y_P}{x_Q + x_P}\right)$$

- **4.** Se $P = (x_P, y_P)$, então $R = 2P = (x_R, y_R)$ é determinado pelas seguintes regras:

$$x_R = \lambda^2 + \lambda + a$$
$$y_R = x_2^P(\lambda + 1x)_R$$

  onde

$$\lambda = x_P + \left(\frac{y_P}{x_P}\right)$$

5.4 Resumo

Vimos neste capítulo que a CCE é uma alternativa em relação ao padrão RSA. Definimos as curvas elípticas e que elas pertencem aos corpos finitos, além disso, a CCE é baseada em curvas elípticas, e esta segue as propriedades de um grupo abeliano, por fim, demos uma explicação algébrica e geométrica das curvas elípticas.

Capítulo 6

Aplicação à Criptografia

Neste capítulo abordaremos os conceitos tratados nos capítulos anteriores aplicados à Criptografia de Curva Elíptica.

A maioria dos produtos e padrões que usam *criptografia de chave pública*, para *criptografar assinaturas digitais* usam RSA [16]. O comprimento da chave para a segurança e o uso da RSA aumentou nos últimos anos, e isso colocou uma carga de processamento mais "pesada" em aplicativos que usam este padrão.

Este fardo tem ramificações, especialmente, para sites de comércio eletrônico que realizam um grande número de transações seguras. Um sistema concorrente desafia a RSA: Criptografia de Curva Elíptica (CCE).

A principal atração da CCE, em comparação com a RSA, é que parece oferecer segurança igual para um tamanho de chave muito menor, reduzindo assim a sobrecarga de processamento. Por outro lado, mesmo que a teoria da CCE tenha sido escrita faz algum tempo, apenas recentemente o seu uso começou a aparecer com mais frequência, e despertou maior interesse dos cripto-analíticos, paralelo a isto surgiu também sondagens sobre a fraquezas da CCE. Assim, o nível de confiança na CCE tem sido aprimorado, frequen-

temente, para se tornar padrão em muitas aplicações.

CCE é, fundamentalmente, mais difícil de explicar do que RSA ou Diffie-Hellman [4], e uma descrição matemática completa está além do escopo deste trabalho de conclusão de curso.

Uma série de *cifras de chave pública* baseia-se no uso de grupo abeliano. Por exemplo, em Diffie-Hellman a troca de chaves envolve multiplicar pares de inteiros diferentes de zero módulo um número primo q. Chaves são geradas por exponenciação sobre grupos. Para atacar as cifras construídas pelo algoritmo Diffie-Hellman [4], o invasor deve encontrar k dado a e a^k; note que isto é um problema logarítmico.

Em Criptografia de Curva Elíptica, uma operação de multiplicação é usada e, como vimos, a multiplicação neste conjunto é equivalente a várias adições dos pontos da curva elíptica. Portanto, a criptoanálise tem como uma das suas funções determinar o valor de k dado a e $(a \cdot k)$.

6.1 Criptografia de Curva Elíptica

A operação de adição em Criptografia de Curva Elíptica (CCE) é a contrapartida da multiplicação modular no RSA, e adição múltipla é a contrapartida de exponenciação modular. Para formar um sistema criptográfico usando curvas elípticas, precisamos encontrar um "problema difícil" correspondente à fatoração do produto de dois primos ou tomar o logaritmo discreto.

Considere a equação $Q = kP$ onde $Q, P \in E_p(a, b)$ e $k < p$. É relativamente fácil calcular Q dado k e P, porém é relativamente difícil determinar k dado Q e P. Isto é chamado de problema de logaritmo discreto para curvas elípticas.

Damos um exemplo tirado do site (`www.certicom.com`). Considere o

grupo $E_{23}(9, 17)$. Este é um grupo definido pela equação $y^2 \mod 23 = (x^3 + 9x + 17) \mod 23$. Qual é o logaritmo discreto k de $Q = (4, 5)$ na base $P = (16, 5)$?

Podemos usar o método de força bruta, para responder esta questão, que consiste em calcular múltiplas vezes o valor de P até encontrar Q temos que:

$$P = (16, 5); 2P = (20, 20); 3P = (14, 14); 4P = (19, 20); 5P = (13, 10);$$
$$6P = (7, 3); 7P = (8, 7); 8P = (12, 17); 9P = (4, 5),$$

portanto, como $9P = (4, 5) = Q$, o logaritmo discreto $Q = (4, 5)$ na base $P = (16, 5)$ é $k = 9$. Em uma aplicação real, k **seria tão grande que tornaria o método de forçar bruta inviável**.

Análogo a troca de chaves Diffie-Hellman

Troca de chaves usando curvas elípticas pode ser feito da seguinte maneira. Primeiro escolha um inteiro grande q, que seja um número primo p ou um inteiro de forma 2^m, e parâmetros de curva elíptica a a b para equação 5.5 ou equação 5.7. Isto define o grupo elíptico dos pontos $E_q(a, b)$. Em seguida, escolha um *ponto base* $G = (x_1, y_1)$ em $E_p(a, b)$ cuja ordem é um valor muito grande n. A **ordem** n de um ponto G em uma curva elíptica (ou seja, o menor inteiro positivo n tal que $n \cdot G = O$) e G são parâmetros do sistema criptográfico conhecido por todos os participantes.

Um troca de chaves entre os usuários Alice e Bob pode ser realizada segundo a Tabela 6.1.

Elementos Públicos Globais	
$E_q(a,b)$	curva elíptica com parâmetros a, b e q, onde q é um um primo ou um inteiro da forma 2^m
G	ponto na curva elíptica cujo a ordem é um valor n grande

	Geração da Chave Pública de Alice
Selecionar n_A privada	$n_A < n$
Calcule P_A pública	$P_A = n_A \cdot G$

	Geração da Chave Pública de Bob
Selecionar n_B privada	$n_B < n$
Calcule P_B pública	$P_B = n_B \cdot G$

Cálculo da Chave Secreta de Alice
$k = n_A \cdot P_B$

Cálculo da Chave Secreta de Bob
$k = n_B \cdot P_A$

Tabela 6.1: CCE troca de chaves

- **1.** Alice escolhe um inteiro n_A menor que n, ou seja, ela seleciona uma das chaves privadas que foram criadas por ela. Então, ela gera uma chave pública $P_A = n_A \cdot G$; A chave pública de Alice é um ponto em $E_q(a, b)$.

- **2.** Bob similarmente seleciona uma chave privada n_B e calcula uma chave pública P_B.

- **3.** Alice gera a chave secreta $k = n_A \cdot P_B$; Bob também gera uma chave secreta $k = n_B \cdot P_A$.

Os dois cálculos na **etapa 3** produzem o mesmo resultado porque

$$n_A \cdot P_B = n_A \cdot (n_B \cdot G) = n_B \cdot (n_A \cdot G) = n_B \cdot P_A$$

para quebrar este esquema, um invasor precisaria ser capaz de computar k dado G e $k \cdot G$, que é assumido como sendo difícil.

Por exemplo [1], dado $p = 211$; $E_p(0, -4)$, que é equivalente à curva $y^2 = x^3 - 4$; e $G = (2, 2)$. Primeiro podemos calcular $240 \cdot G = O$. A chave privada de Alice é $n_A = 121$, então a chave pública de Alice é gerada por $P_A = 121 \cdot (2, 2) = (115, 48)$. A chave privada Bob é $n_B = 203$, então a chave pública de Bob é dada por $203 \cdot (2, 3) = (130, 203)$. O compartilhamento da chave secreta é $121 \cdot (130, 203) = 203 \cdot (115, 48) = (161, 69)$.

Perceba que a chave secreta é um par de números. Se esta chave for usada como uma chave de sessão para criptografia convencional, então um único número deve ser gerado. Nós poderíamos simplesmente usar as coordenadas de x ou alguma função simples das coordenadas de x.

Codificação e Decodificação de Curvas Elípticas

Várias abordagens para codificar/decodificar usando curvas elípticas foram analisadas na literatura. Nesta subseção, nós vamos olhar para talvez a mais simples. A primeira tarefa neste sistema é codificar a mensagem de texto simples m para ser enviada como um ponto P_m na curva. É o ponto P_m que será criptografado como um texto cifrado e subsequentemente decodificado.

Note que não podemos simplesmente codificar a mensagem como a coordenada x ou y de um ponto porque nem todas essas coordenadas estão em $E_q(a, b)$; por exemplo, como podemos perceber da tabela 5.1.

Novamente, existem várias abordagens para essa codificação, as quais não abordaremos aqui, mas basta dizer que existem técnicas relativamente diretas que podem ser usadas.

Como no sistema de troca de chaves um sistema de codificação/decodificação requer um ponto G e um grupo elíptico $E_q(a, b)$ como parâmetros.

[1] Fornecido por Ed Schaefer da Universidade Santa Clara

Todo usuário, seleciona uma chave privada e gera uma chace pública, vamos pegar dois usuários quaisquer que chamaremos de B e A. Se o usuário A seleciona uma chave privada n_A e gera uma chave pública $P_A = n_A \cdot G$.

Para *codificar* e enviar um *texto claro* ou *mensagem clara*, P_m, para B, A escolhe um inteiro positivo aleatório k, com isso, podemos produzir (gerar) o *texto cifrado* ou *mensagem cifrada*, C_m, consistindo do par de pontos:

Codificação

$$C_m = \{k \cdot G, P_m + k \cdot P_B\}$$

Note que A usou a chave pública, P_B, do usuário B para codificar o texto claro.

Agora para B *decodificar* a mensagem cifrada, ele multiplica o primeiro ponto do par da *chave secreta* dele e subtrai o resultado do segundo ponto:

Decodificação

$$P_m + k \cdot P_B - n_B \cdot (k \cdot G) = P_m + k \cdot (n_B \cdot G) - n_B \cdot (k \cdot G) = P_m$$

Com isso, ele decodifica a mensagem, por fim, note que para cifrar a mensagem clara, nós adicionamos $k \cdot P_B$ a P_m, portanto, ninguém a não ser A ou B, sabem o valor de k, mesmo que P_B seja uma chave pública e, portanto, o invasor tenha acesso, ninguém poderá decodificar a cifra $k \cdot P_B$ a não ser B.

Contudo, A acaba incluindo uma "pista" neste processo de codificação obrigatoriamente, que é suficiente para decifrar a mensagem, se alguém souber a chave privada n_B de B. Para um invasor recuperar a mensagem, ele

(invasor) teria que computar k com valor de G e $k \cdot G$, o que é considerado difícil.

Vamos pegar um exemplo deste processo de criptografia, que pode ser encontrado em Koblitz [13], pegue $p = 751$; $E_p(-1, 188)$, que é equivalente à curva $y^2 = x^3 - x + 188$; e $G = (0, 376)$. Suponha que Alice deseja enviar uma mensagem para Bob que esteja codificada no ponto $P_m = (562, 211)$ e que Alice selecione o número aleatório $k = 386$, a chave pública de Bob é $P_B = (201, 5)$.

Temos que $386 \cdot (0, 376) = (676, 558)$ e $(562, 201) + 386 \cdot (201, 5) = (385, 328)$. Assim, Alice envia o texto cifrado $C_m = \{(676, 558), (385, 328)\}$ para Bob de forma codificada (criptografada).

Segurança da Criptografia de Curvas Elípticas

A segurança da CCE depende de quão difícil é determinar k dado $k \cdot P$ e P. Isto é chamado de problema do logaritmo da curva elíptica. Uma técnica eficiente para obter o logaritmo da curva elíptica é conhecida como método Pollard Rho [12].

Vamos usar este método para comparar vários algoritmos mostrando tamanhos de chaves comparáveis em termos de esforço computacional para criptoanálise (veja a Tabela 6.2). Como pode ser visto, um tamanho de chave consideravelmente menor pode ser usado para a CCE em comparação com a RSA.

Além disso, para comprimento de chaves iguais, o esforço computacional necessário para CCE e RSA é comparável segundo Jurisc [10]. Assim, há uma vantagem computacional no uso de CCE como um comprimento de chave menor que uma RSA comparativamente seguro.

Esquema Simétrico (tamanho da chace em bits)	*Esquema Baseado CCE* (tamanho do n em bits)	*RSA* (tamanho do módulo em bits)
56	112	512
80	160	1024
112	224	2048
128	256	3072
192	384	7680
256	512	15360

Tabela 6.2: Tamanho de chaves comparáveis em termos de esforço computacional

6.2 Resumo

Neste capítulo falamos sobre o uso de Criptografia de Curva Elíptica e também dissemos que a troca de chaves em CCE é semelhante ao Diffie-Hellman e explicamos como codificar e decodificar uma mensagem usando esse método. Por fim, falamos sobre a confiabilidade da segurança da CCE.

Capítulo 7

Conclusão

Este trabalho teve como pretensão apresentar os conceitos fundamentais para compreendermos as bases das cifras largamente usadas em sistemas que aplicam criptografia em seus ambientes de produção, quando se pretende construir uma comunicação confiável entre dispositivos, segura e confidencial. Portanto, teoria dos números e curvas elípticas são a base para compreendermos o funcionamento dos algoritmos criptográficos modernos.

Referências Bibliográficas

[1] Secure Hash Standard (SHS) - NIST.

[2] Anthoy W. Knapp. *Elliptic Curves*. Princeton University Press, 1999.

[3] Joan Daemen and Vincent Rijmen. *The design of Rijndael AES Advanced Encryption Standard*. Springer-Verlag, 2002.

[4] Whitfield Diffie and Martins E. Hellman. New directions in cryptography, 1976.

[5] Christof Paar e Jan Pelzl. *Understanding Cryptography*. Springer, 2010.

[6] A. Fernandes. Elliptic curve cryptography, 1999.

[7] IBM. Data Encryption Standard (DES). Federal Information Processing Standard Publication 46, 1977.

[8] Marcos Antonio S. J. Algoritmos criptográficos para redes de sensores, 2008.

[9] Williamson M. J. Non-secret encryption using a finite field., 1974.

[10] Jurisic, A., and Menezes, A. Elliptic curves and cryptography, 1997.

[11] David Kahn. The codebreakers, Dec 1996.

[12] Jonathan Katz and Yehuda Lindell. Introduction to modern cryptography, 2007.

[13] Neal Koblitz. *A Course in Number Theory and Cryptography*. Springer, 1994.

[14] Calvin T. Long. *Elementary Introduction to Number Theory*. Waveland, 1995.

[15] Francisco Cesar Polcino .M. Números, 2001.

[16] R.L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public, key cryptography, 1977.

[17] Claude E. Shannon. Communication theory of secrecy systems. *Bell System Technical*, 28:656–715, 1949.

[18] William Stallings. *Cryptography and Network Security*. Pearson, 2011.

[19] Joan Daemen Vincent Rijmen. Specification for the Advanced Encryption Standard (AES). Federal Information Processing Standards Publication 197, 2001.

[20] Andrew Wiles. Modular Elliptic Curves and Fermat's Last Theorem, 1995.