Tópicos Avançados em Ciência da Computação I: Introdução à Teoria de Códigos para Criptografia Pós-quântica

Daniel Panario
School of Mathematics and Statistics
Carleton University

IC - Unicamp, Sala 351 do IC-3
das 13:30 as 18:30 em 16-17 de janeiro de 2020,
das 13:30 as 17:30 em 20-24 de janeiro de 2020

Introdução ao Curso

Daniel Panario
School of Mathematics and Statistics
Carleton University

16 de janeiro de 2020

Introdução ao curso

MO850 - Tópicos Avançados em Ciência da Computação I:
Introdução à Teoria de Códigos para Criptografia Pós-quântica

Objetivo:
O objetivo deste curso é o estudo dos principais conceitos, métodos e resultados da Teoria dos Códigos Corretores de Erros para a Criptografia Pós-Quântica baseada em códigos; em especial para as propostas de padronização do NIST (National Institute of Standards and Technology).

Palavras-chave:
Códigos de Goppa; códigos LDPC e QC-LDPC; Tanner graph; algoritmo de decodificação bit-flipping; criptossistemas de McEliece, LEDAcrypt e Bike.

Avaliação: A avaliação é baseada em duas listas de exercícios (40%), apresentação (25%) e projeto final (35%).

- As aulas serão faladas em portunhoglês e escritas em portuglês.

Introdução ao curso

MO850 - Tópicos Avançados em Ciência da Computação I:
Introdução à Teoria de Códigos para Criptografia Pós-quântica

Objetivo:
O objetivo deste curso é o estudo dos principais conceitos, métodos e resultados da Teoria dos Códigos Corretores de Erros para a Criptografia Pós-Quântica baseada em códigos; em especial para as propostas de padronização do NIST (National Institute of Standards and Technology).

Palavras-chave:
Códigos de Goppa; códigos LDPC e QC-LDPC; Tanner graph; algoritmo de decodificação bit-flipping; criptossistemas de McEliece, LEDAcrypt e Bike.

Avaliação: A avaliação é baseada em duas listas de exercícios (40%), apresentação (25%) e projeto final (35%).

- As aulas serão faladas em portunhoglês e escritas em portuglês.

Daniel Panario | Introdução ao Curso | 16 de janeiro de 2020 | 3 / 26

Ementa do curso

Revisão de conceitos matemáticos básicos: corpos finitos, espaços vetoriais sobre corpos finitos. Detalhamento de limites fundamentais e de alguns métodos de codificação: códigos de Hamming, BCH, Goppa, cíclicos, quase cíclicos.

Códigos LDPC (low density parity check), e MDPC (medium density parity check). Métodos de decodificação fundamentais em criptossistemas baseados em códigos: algoritmos de bit flipping e soma-produto.

Aplicação dos conceitos vistos nos métodos criptográficos baseados em códigos inscritos no concurso de padronização do NIST de criptografia pós-quântica, em andamento.

Criptografia pós-quântica

Algoritmos criptográficos tradicionais baseados nos problemas do logaritmo discreto e da fatoração de inteiros se tornam inseguros contra um adversário quântico.

Computadores quânticos quebrariam os mais populares sistemas de chave pública:

- RSA,
- DSA,
- ECDSA,
- ECC,
- HECC
- etc,

Podem ser atacados em tempo polinomial usando o algoritmo de Shor.

Criptografia pós-quântica (cont)

Criptografia pós-quântica procura sistemas que:

- sejam executados em computadores convencionais, e
- que sejam seguros contra ataques usando computadores quânticos.

Métodos principais:

- Hash-based cryptography;
- Code-based cryptography;
- Lattice-based cryptography;
- Multivariate-quadratic-equations cryptography;
- Isogenies-based cryptography;
- Zero-knowledge proofs.

Texto: Bernstein, Buchmann, and Dahmen, eds., Post-Quantum Cryptography, Springer, 2009.

Propostas na segunda rodada do NIST

Na primeira rodada houve 69 submissões. Delas ainda estão na competição 26 propostas: 17 para encriptação e decriptação (das quais 7 são baseadas em códigos), e 9 para assinaturas digitais:

- Hash-based cryptography (1 assinatura);
- Code-based cryptography (7 PKE/KEM);
- Lattice-based cryptography (9 PKE/KEM, 3 assinaturas);
- Multivariate-quadratic-equations cryptography (4 assinaturas);
- Isogenies-based cryptography (1 PKE/KEM);
- Zero-knowledge proofs (1 assinatura).

A visão do NIST

O processo de padronização do NIST (National Institute of Standards and Technology, USA) busca novos algoritmos, em particular para encriptação e decriptação (Key Encapsulation Mechanisms - KEMs), assim como para assinaturas digitais.

A visão do NIST está no seguinte documento:

https://csrc.nist.gov/CSRC/media/Presentations/Round-2-of-the-NIST-PQC-Competition-What-was-NIST/images-media/pqcrypto-may2019-moody.pdf

Os sistemas criptográficos baseados na teoria de códigos são candidatos para criar sistemas de encriptação e decriptação seguros; ver slides 8 (primeira rodada) e slides 28/29 (segunda rodada, proposta ainda sob consideração).

Propostas na segunda rodada (baseadas em códigos)

Ver lâminas 28-29 do documento de NIST; verbatim dessa página:

- BIKE
  Three versions. Based on quasi-cyclic MDPC codes. Ephemeral use only. Similar key size and performance to lattice schemes. More analysis needed of particular security assumption.
  Tweaks: New decoder yielding smaller error rates, new CCA version

- Classic McEliece
  Based on established McEliece cryptosystem (binary Goppa codes). Lots of analysis of security problem. No decryption failures. Short ciphertexts. Okay performance. Very large public keys. Only level 5 parameters given.
  Tweaks: More parameter sets/security levels, future proposal with 2 times faster keygen algorithm

Propostas na segunda rodada (baseadas em códigos)

- HQC
  Low decryption failure rate (necessary for CCA security). As a result, slightly larger key and ciphertext sizes. More analysis needed of particular security assumption.
  Tweaks: dropped some parameter sets, updated implementation

- LEDAcrypt
  Merger. Based on quasi-cyclic LDPC codes, which have more structure than QC-MDPC codes. New parameters with low decryption rates. Needs more analysis.
  Tweaks: Updated parameters, CCA version, better failure rates, new transform

- NTS-KEM
  Very, very similar to Classic McEliece, but with some different design choices. Needs constant time implementation.
  Tweaks: Uses implicit rejection

Propostas na segunda rodada (baseadas em códigos)

- Rollo
  Merger of 3 rank-based schemes using LRPC codes. 2 schemes are ephemeral, 1 targets CCA security. Newer security assumption.
  Tweaks: Uses ideal codes instead of quasi-cyclic ones (Rollo-3), updated parameters

- RQC
  Rank-based scheme. No decryption failures. As a result, slower speeds and ciphertext size. Security problem needs more analysis, as it is newer.
  Tweaks: Uses ideal codes (not quasi-cyclic), updated parameters, updated implementation.

Segunda rodada da competição de NIST

O documento principal para ler é o seguinte:

https://csrc.nist.gov/Projects/post-quantum-cryptography/round-2-submissions

A seguir damos um resumo (bem breve) das informações das sete propostas ainda na competição que usam teoria de códigos. Também damos links para os documentos das sete propostas que usaremos nesse curso.

Páginas das propostas e resumo dos códigos usados

1. BIKE - Bit Flipping Key Encapsulation:
   https://bikesuite.org/
   Specification: BIKE-Spec-2019.06.30.1.pdf
   Slides Round 2: Slides_bike.pdf
   Codes: QC-MDPC; bit flipping.

2. Classic McEliece:
   https://classic.mceliece.org/
   Specification: mceliece-20190331.pdf
   Slides Round 2: Slides_classic-mceliece.pdf
   Codes: Goppa codes.

3. HQC
   http://pqc-hqc.org/
   Specification: hqc-specification_2019-08-24.pdf
   Slides Round 2: Slides_hqc.pdf
   Codes: BCH codes, syndrome decoding of BCH codes.

Páginas das propostas e resumo dos códigos usados

1. LEDAkem Key Encapsulation Module e
   LEDApkc Public Key Cryptosystem
   `https://www.ledacrypt.org/`
   Specification: LEDAkem_spec_latest.pdf e LEDApkc_spec_latest.pdf
   Slides Round 2: Slides_ledacrypt.pdf
   Codes: LDPC codes, Q-decoder bit flipping.

2. NTS-KEM
   `https://nts-kem.io/`
   Specification: nts-kem-20191129.pdf
   Slides Round 2: Slides_nts-kem.pdf
   Codes: McEliece and Niederreiter variant.

Páginas das propostas e resumo dos códigos usados

1. Rollo
   `http://www.pqc-rollo.org/`
   Specification: rollo-specification_2019-08-24.pdf
   Slides Round 2: Slides_rollo.pdf
   Codes: rank metric, LRPC codes, rank syndrome decoding

2. RQC
   `http://pqc-rqc.org/`
   Specification: rqc-specification_2019-08-24.pdf
   Slides Round 2: Slides_rqc.pdf
   Codes: rank metric, QC codes, rank syndrome decoding

Conteúdo das aulas

- Introdução ao curso e propostas da competição do NIST; conteúdo das aulas; passada (aérea) no sistema de McEliece.
- Introdução aos corpos finitos I: noções básicas, propriedades fundamentais, representações e operações. Corpos finitos: estrutura, existência e unicidade dos corpos finitos; subcorpos; elementos primitivos.
- Introdução à teoria de códigos I: noções básicas; distância; códigos lineares; decodificação por síndrome; cotas; códigos de Hamming; introdução aos códigos cíclicos e quase-cíclicos.

Conteúdo das aulas (cont)

- Teoria de códigos II: códigos low density parity check (LDPC) e quase-cíclicos low density parity check (QC-LDPC); grafo de Tanner; algoritmo de decodificação bit-flipping.
  Proposta NIST: LEDAcrypt; seções 1 e 2, páginas 11-26.
- Códigos medium density parity check (MDPC).
  Proposta NIST: Bit Flipping Key Encapsulation (BIKE); seção 1, páginas 1-6.
- Corpos finitos II: polinômios irredutíveis, primitivos e minimais; polinômios linearizados e de permutação.
- Códigos de Goppa; McEliece cryptosystem.
  Propostas NIST: Classic McEliece, seções 1 e 2, páginas 4-10 (primeira submissão); NTS-KEM, seções 2-2.2 (páginas 5-8).

Conteúdo das aulas (cont)

- Teoria de códigos III: códigos cíclicos e ideais de polinômios; códigos BCH que corrigem dois e t erros; códigos de Reed-Solomon (breve). Proposta NIST: Hamming Quasi-Cyclic (HQC); seções 1.5.2-1.5.4 (páginas 16-21). Decodificação de códigos BCH; polinômio localizador de erros.
- Corpos finitos III: fatoração de polinômios e polinômio localizador de erros; métodos baseados em polinômios linearizados, e traço de Berlekamp.
- Códigos low rank parity check (LRPC) e códigos Gabidulin. Propostas NIST: Rollo e RQC; seções 1.1.1-1.1.4 (páginas 3-11) de RQC.

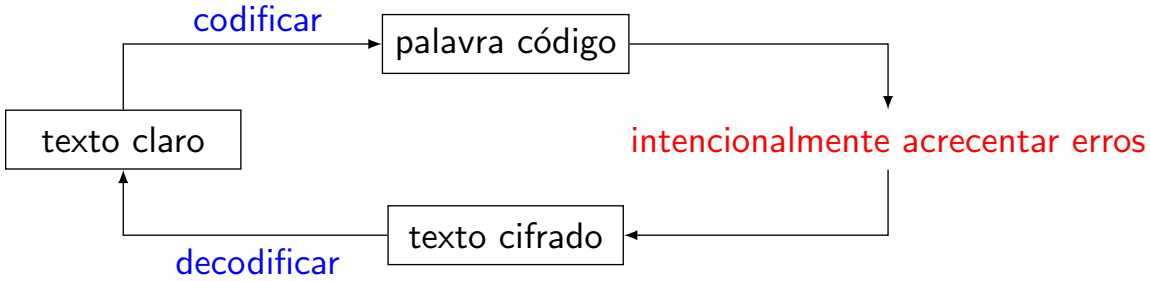
Deadlines para alunos registrados

- Distribuição da primeira lista de exercícios e da lista de projetos: sexta, 17 de janeiro.
- Projeto final decidido: segunda, 20 de janeiro.
- Entrega da primeira lista de exercícios: sexta, 24 de janeiro.
- Entrega da primeira parte do projeto: sexta, 24 de janeiro.
- Distribuição da segunda lista de exercícios: sexta, 24 de janeiro.
- Apresentação oral: sexta, 24 de janeiro.
- Entrega da segunda lista de exercícios: sexta, 31 de janeiro.
- Entrega final do projeto: segunda, 3 de fevereiro.

Breve Introdução ao Sistema de McEliece (1978) e a Criptografia Baseada em Códigos

O sistema de McEliece: primeira visão

Robert McEliece (1978) apresentou o primeiro sistema criptográfico baseado na teoria de códigos. A estrutura desse sistema é usada até hoje por outros métodos baseados na teoria de códigos.



O sistema de McEliece: preparação

Elementos do sistema:

- É dado um código de Goppa (binário) Γ de comprimento 1024, dimensão 524 e que pode corrigir 50 erros. O código Γ é mantido em segredo.
- A chave secreta do sistema de McEliece é a tripla (G, S, P) consistindo de uma matriz geradora G para o código Γ, uma matriz 1024×1024 de permutação P, e uma matriz 524×524 inversível S.
- Os tamanhos 1024, 524 e 50 são parâmetros públicos do sistema.
- A chave pública do sistema de McEliece é a matriz 524×1024 $\widehat{G} = SGP$.

O sistema de McEliece: encriptação e decriptação

Encriptação de uma mensagem $m \in \{0,1\}^{524}$:

- Calcular $m\widehat{G}$ e esconder a mensagem somando um vetor de erro aleatório e de comprimento 1024 e peso 50.
- Enviar $y = m\widehat{G} + e$.

Decriptação do texto y:

- Calcular $yP^{-1} = mSG + eP^{-1}$.
- Temos que mSG é uma palavra de código no código secreto Γ; também temos que o vector de erro permutado eP^{-1} tem peso 50.
- Usar o algoritmo de decodificação para o código Γ para achar mS e, portanto, m.

Segurança dos códigos no sistema de McEliece

A proposta original estava baseada em códigos de Goppa binários irredutíveis para os quais existem algoritmos eficientes de decodificação (por exemplo, o algoritmo de Paterson, 1975).

É crucial que exista um algoritmo eficiente de decodificação de t erros para o código sendo usado!

Um outra aspecto é fundamental: o código de Goppa a ser usado é escolhido aleatoriamente.

A segurança do sistema reside na dificuldade de decodificar erros de um codigo aleatório. Dado um código linear (sem estrutura obvia), Berlekamp, McEliece e van Tilborg (1978) provaram que o problema de decodificar códigos lineares é NP-completo.

Variantes do código no sistema de McEliece

Atualmente decodificar um código binário de comprimento n e dimensão próxima de $n/2$ custa aproximadamente $2^{(0.5+o(1))n/\log_2(n)}$ operações binárias.

É, portanto, importante que o código não possua estrutura. Muitos códigos com algoritmo de decodificação rápida têm estrutura. É, então, necessário esconder essa estrutura. (Voltaremos sobre isto quando falarmos de códigos LDPC.)

Após 42 anos o sistema de McEliece ainda é considerado seguro (após reajuste dos parâmetros).

Exsitem muitas variantes usando outros códigos; a maioria dessas variantes não são seguras. Os métodos ainda na competição do NIST são exceções.

Segurança de sistemas basedos em códigos

Repetindo os aspectos fundamentais desses sistemas:

(1) Deve existir um algoritmo eficiente de decodificação de muitos erros para o código sendo usado.

(2) O código a ser usado deve ser escolhido aleatoriamente.

(3) O código não deve ter estrutura; se tiver, é necessário esconder a estrutura.