Report from Dagstuhl Seminar 18511

Algebraic Coding Theory for Networks, Storage, and Security

Edited by

Eimear Byrne[1], Martin Bossert[2], and Antonia Wachter-Zeh[3]

1 **University College Dublin, IE,** ebyrne@ucd.ie
2 **Universität Ulm, DE,** martin.bossert@uni-ulm.de
3 **Technical University of Munich, DE** antonia.wachter-zeh@tum.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 18511 "Algebraic Coding Theory for Networks, Storage, and Security". The ever increasing traffic in networks and the growth of distributed storage systems require advanced techniques based on algebraic coding to meet user demand. Private access to such services is a major concern for consumers and is still a new field in the context of distributed storage. The topics of this workshop are very relevant for a number of emerging industrial research fields concerned with efficient and reliable storage and transmitting large files through large networks.

Seminar December 16–21, 2018 – http://www.dagstuhl.de/18511
2012 ACM Subject Classification Mathematics of computingSecurity and privacy
Keywords and phrases Coding theory, infomation theory, distributed storage, cryptography, error-correction, private information retrieval, private computation, adversarial channel
Digital Object Identifier 10.4230/DagRep.8.12.49
Edited in cooperation with Giuseppe Cotardo

1 Executive Summary

Eimear Byrne
Martin Bossert
Antonia Wachter-Zeh

Algebraic Coding Theory for Networks, Storage, and Security was the fourth in a series of seminars exploring applications of coding theory in modern communications theory (see also Dagstuhl Seminars 16321 (2016), 13351 (2013) and 11461 (2011)). The seminar brought together 50 mathematicians, engineers and computer scientists with expertise in coding theory, network coding, storage coding, cryptography and code-based security to participate in dissemination and collaboration within the seminar themes.

The main focus of this workshop was to explore novel results in coding theory for application in data storage management, cryptography and privacy. The impact of novel coding techniques across these domains was discussed and explored. Particular emphasis was placed on new applications of coding theory in public key cryptography, coding techniques for privacy in distributed storage and on practical schemes using coding theory for content delivery. These novel coding applications continue to have a significant impact on changing focus and broadening of coding theory fundamentals.

Algebraic Coding Theory for Networks, Storage, and Security, *Dagstuhl Reports*, Vol. 8, Issue 12, pp. 49–67
Editors: Eimear Byrne, Martin Bossert, and Antonia Wachter-Zeh

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Overview talks were given by Philippe Gaborit (*Recent Results for Cryptography Based on Rank Metric*), Emina Soljanin, (*Service Rates of Codes*), Eitan Yaakobi, (*Private Proximity Retrieval*), Sacha Kurz, (*Multisets of Subspaces and Divisible Codes*), Heide Gluesing-Luerssen (*On Ferrers Diagram Codes*) and Salim El Rouayheb (*GASP Codes for Secure Distributed Matrix Multiplication*). In addition, several short talks were given by other participants based on current research interests with a view to stimulating collaboration. Presentations on system cybersecurity, private information retrieval, locally recoverable codes, adversarial channels and various aspects of rank metric codes were given. The remaining seminar time was allocated to discussion groups, including those in code-based cryptography, private computation, service rates of codes, algebraic geometry codes and adversarial channels. Aside from the working group discussions, particpants took the opportunity to engage in specific collaborations with co-authors.

We summarize some of the content of the working group discussions below. It has been well documented that redundancy is a basic requirement for stability of distributed data storage systems. Algebraic codes have been identified as having applications in providing efficiency in this domain far exceeding replication. Coding theory methods allow information retrieval minimizing disc access, storage size, local recoverability, data repair and data retrieval. Consequently, the area of storage coding has seen an exponential growth. An important aspect of user access in distributed storage is privacy of information retrieval so that users who are remotely accessing files can do so without storage servers knowing what they have accessed. Attempts to efficiently solve this problem come from coding theory.

An important application of secret sharing schemes is distributed storage of private data, where each party is a storage node and all parties wish to store a secret securely and reliably. Secret sharing is a fundamental cryptographic primitive and is used as a building block in numerous secure protocols. In our discussions we focussed on secret sharing schemes for the threshold access structure and on secret sharing with errors/attacks in a broader context. Fuzzy vaults and secret sharing over networks were discussed. A motivation for this area is for example biometric authentication in the presence of adversaries.

Another aspect of distributed storage is the service rate of codes. Emerging applications, such as distributed learning and fog computing, add yet another use for coding. In these applications, the goal is to maximize the number of users that can be simultaneously served by the system. One such service is simultaneous download of different jointly coded data blocks by many users competing for the system's resources. Here, coding affects the rates at which users can be served. The achievable service rate region is the set of request rates for each file that can be supported by the system. A variety of approaches to open problems about service rate were discussed. In particular, we addressed the question of code constructions that serve all requests for fixed rate constraints on file and the problem of how to determine the achievable service rate region for certain families of codes.

Privacy and security present formidable challenges in our modern connected world. Public-key cryptography is the foundation of multi-party communication as well as for key exchange of symmetric cryptosystems. With the increasing likelihood of a capable quantum computer, post-quantum secure systems have recently turned into the research focus, especially for devices that are hard to update and have very long life cycles. Code-based cryptography provides post-quantum secure public-key systems.

The working group on code-based-cryptography discussed the importance of security reduction arguments and went through several examples of these in relation to coding theory in cryptography. This discussion was a great benefit to the participants, many of whom have expertise in coding theory and keen to broaden their understanding of cryptography.

The group also focussed on McEliece-like systems based on quasi-cyclic moderate density parity-check (QC-MDPC) codes and on low-rank parity-check (LRPC) codes. Distinguisher attacks were discussed, as well as possible modifications to the broken Gabidulin based cryptosystem.

Reliable communication across a channel in the presence of an adversary is a very general channel model that arises in many applications. Coding strategies for data transmission and authentication across the arbitrarily varying channel (where an adversary may alter the channel statistics) and for covert communication were discussed. A framework for linear systems under attack, such as the scenario where a restricted number of sensor measurements is vulnerable to adversarial attacks, was introduced and coding theoretic arguments used for attack detection and correction strategies.

There were about 20 PhD and postdoctoral researchers in attendance, who reported a very positive experience and satisfaction at being give the opportunity to explore new collaborations with more senior researchers and to get exposure to new problems in coding theory. All participants welcomed the time made available to them to take part in discussion groups and in more focussed collaborations. All were very pleased with the quality of the facilities and administrative support offered by staff at Schloss Dagstuhl, which made for a very productive meeting. Andreas Lenz and Rawad Bitar organised an afternoon excursion to Trier for the group. Giuseppe Cotardo collected and compiled data for the final published report.

2 Table of Contents

3 Overview of Talks

3.1 Recent results on rank based cryptography

Philippe Gaborit (University of Limoges, France)
gaborit@unilim.fr

We propose an overview of recent results in rank based cryptography.

3.2 The Covering Radius of Rank-Metric Codes

Alberto Ravagnani (University College Dublin, Ireland)
alberto.ravagnani@ucd.ie

The covering radius of a rank-metric code is the maximum distance between the code and a matrix from the ambient space. In this talk, I will discuss some structural properties of matrix codes with the rank metric, and relate them to the covering radius. In particular, I will present new bounds on this parameter obtained with different combinatorial methods.

3.3 How general is the rank decoding algorithm based on rank one elements?

Anna-Lena Trautmann-Horlemann (University of St. Gallen, Switzerland)
anna-lena.horlemann@unisg.ch

The covering radius of a rank-metric code is the maximum distance between the code and a matrix from the ambient space. In this talk, I will discuss some structural properties of matrix codes with the rank metric, and relate them to the covering radius. In particular, I will present new bounds on this parameter obtained with different combinatorial methods.

3.4 Achievable Service Rates in Coded Distributed Storage

Emina Soljanin, Carolyn Mayer (Rutgers University - Piscataway, US)
emina.soljanin@rutgers.edu, cdmayer@wpi.edu

Coding has traditionally been used in transmission and storage of data to provide reliability in a more efficient way than simple replication. The traditional performance indicators of codes are the minimum distance and the code rate. More recently, special codes have

been developed that also provide efficient maintenance of storage under node failures. In addition to the traditional metrics, the properties of codes that matter in such scenarios are the code locality and availability. Emerging applications, such as distributed learning and fog computing, are adding yet another use for coding. In these applications, the goal is to maximize the number of users that can be simultaneously served by the system. One such service is simultaneous download of different jointly coded data blocks by many users competing for the system's resources. Here, coding affects the rates at which users can be served. Interestingly, the best schemes often combine replication and coding. This talk will define the service rates of codes as new performance metrics, survey the existing literature, and show a connection between optimizing the code service rates to a graph vertex cover problem.

3.5 New Bounds and Generalizations of Locally Recoverable Codes with Availability

Alexey Frolov (Skolkovo Institute of Science and Technology (Skoltech) – Moscow, Russia)
al.frolov@skoltech.ru

We investigate the distance properties of linear locally recoverable codes (LRC codes) with all-symbol locality and availability. New upper and lower bounds on the minimum distance of such codes are derived. The upper bound is based on the shortening method and generalized Hamming weights (GHWs) that are fundamental parameters of any linear codes with many useful applications. This bound improves existing upper bounds. To reduce the gap in between upper and lower bounds we do not restrict the alphabet size and propose explicit constructions of codes with locality and availability via rank-metric codes. The first construction relies on expander graphs and is better in low rate region, the second construction utilizes LRC codes developed by Wang et al. as inner codes and is better in high rate region. We also suggest one possible generalization of LRC codes in which the recovering sets can intersect in a small number of coordinates. This feature allows us to increase the achievable code rate and still meet load balancing requirements. We derive upper and lower bounds on the parameters of such codes and present explicit constructions of codes with such a property.

3.6 General adversarial channels

Sidharth Jaggi (The Chinese University of Hong Kong, China)
jaggu@ie.cuhk.edu.hk

The question of when communication is possible in an adversarial jamming context is intimately connected to the question of high-dimensional packings – for instance, communication over a binary-input binary-output channel where the adversary can flip up to pn bits is equivalent to designing packings of pn-radius Hamming balls in n-dimensional Hamming space.

We consider a fairly general class of adversarial channels, and:

- show that each adversarial channel has a bijection with a certain "confusability polytope" embedded in the simplex of all distributions of joint-types of pairs of inputs to the channel;
- precisely characterize when a positive rate is possible (i.e. exponential-size packings are possible). Sufficiency is characterized in terms of codes where each pair of codewords has joint-type given by a "completely positive distribution" outside the confusability polytope. Necessity follows by a Ramsey theoretic argument showing that each large code must have a sufficiently large subcode where each pair of codewords has roughly the same type-class, followed by a Plotkin-type argument, and a separate Fourier analytic argument to handle asymmetric type-classes.

3.7 Private Proximity Retrieval

Eitan Yaakobi (Technion – Haifa, Israel)
yaakobics.technion.ac.il

A private proximity retrieval (PPR) scheme is a protocol which allows a user to retrieve (an approximation of) the set of indices of all records in a database that are within some distance r from the user's record x. The privacy at each server is given by the fraction of the record x that is kept private. The distortion of a PPR scheme measures how accurately the user can calculate this set of indices. We assume that each servers stores a copy of the database. While it is possible to achieve perfect privacy by studying a related private information retrieval problem, it is unclear how to do so without pre-computing the answers for every possible user record x and radius r. We therefore focus on protocols that trade perfect privacy for a massive reduction in computation and storage.

In this paper, we initiate this study while focusing on case when the records are binary vectors together with the Hamming distance. In particular, for a given privacy level, we investigate the minimum number of servers that guarantee a prescribed distortion value. We also consider collusions of pairs of servers and investigate other distance measures.

3.8 Private Function Computation for Coded Databases

Jörg Kliewer (New Jersey Institute of Technology - Newark, US)
jkliewer@njit.edu

We consider the problem of private computation in a distributed storage system. In private computation, a user wishes to compute a function of f messages stored in non-colluding databases while revealing no information about the computation result to the databases. We first employ computation of a linear function of the messages, where linear codes are used to encode the information on the databases. We show that this private linear computation capacity, which is the ratio of the desired linear function size and the total amount of downloaded information, matches the maximum distance separable (MDS) coded capacity of private information retrieval for a large class of linear codes that includes MDS codes.

Our converse result is valid for any number of messages and linear combinations, and the capacity expression depends on the rank of the coefficient matrix obtained from all linear combinations. Finally, we also present initial results how our linear computation approach can be extended to computing arbitrary multivariate polynomials of the messages.

3.9 On decoding Folded and/or Derivative Codes over any field

Daniel Augot (INRIA Saclay - Île-de-France, France École Polytechnique, France)
daniel.augot@inria.fr

Sudan [5] and Guruswami-Sudan [3] algorithms are very general algorithms for decoding Reed-Solomon and algebraic-geometry codes, using the interpolation approach. It can be shown that the algorithms can recover the list of codewords in a ball of relative radius $1-\sqrt{R}$, for Reed-Solomon codes of rate R. One can see that the finiteness of the field does not play any more in these decoding algorithms (although, admittedly, there are variants of these algorithms, adapted to the finite field case, which give an even better decoding radius).

A breakthrough was obtained by Guruswami and Rudra [2] for decoding folded Reed-Solomon, for which, when the parameters are properly set, a decoding radius of $1-R-\varepsilon$ can be achieved (for long codes, and large alphabet). Crucially, this methods relies on a relation

$$X^q \equiv \gamma X \mod E(X)$$

for well chosen $\gamma \in GF(q)$ and $E(X) \in GF(q)[X]$. Using this trick, the root-finding algorithm has to deal with a polynomial of degree q^{s-1}, where s is the interpolation order. Later, Guruswami [1] proposed another algorithm for decoding folded Reed-Solomon codes, where an exhaustive search in an affine space of dimension $s-1$ has to be done, leading to a q^{s-1} factor in the complexity. Also, Guruswami and Wang consider derivate codes, but still face a q^{s-1} complexity wall [4].

It is striking that, for decoding these codes there is no "algebraic algorithm" which would work over any field, finite or not. This talk will discuss this issue.

References

1 "Linear-algebraic list decoding of folded Reed-Solomon codes", Venkatesan Guruswami, *Computational Complexity (CCC), 2011 IEEE 26th Annual Conference on*, 77–85, IEEE, 2011.
2 "Explicit capacity-achieving list-decodable codes", Venkatesan Guruswami and Atri Rudra, "Proceedings of the thirty-eighth annual ACM symposium on Theory of computing", 1–10, ACM, 2006.
3 "Improved decoding of Reed-Solomon and algebraic-geometric codes", Venkatesan Guruswami and Madhu Sudan, *Foundations of Computer Science, 1998. Proceedings. 39th Annual Symposium on*, 28–37, IEEE, 1998.
4 "Optimal rate list decoding via derivative codes", Venkatesan Guruswami and Carol Wang, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 593–604, Springer, 2011.
5 "Decoding of Reed Solomon codes beyond the error-correction bound", Madhu Sudan, *Journal of complexity*, 13(1):180–193, 1997.

3.10 (Multi-)Sets of subspaces and divisible codes

Sascha Kurz (University of Bayreuth, Germany)
sascha.kurz@uni-bayreuth.de

A multi set of subspaces in $GF(q)^v$ gives rise to a q^r divisible linear code if the dimensions of the subspaces are at least $r+1$. This connection has implications e.g. for the existence of vector space partitions, packing or covering problems for subspaces, or subspace codes. Several optimal linear codes in the Hamming metric are divisible. The cylinder conjecture of Ball is actually a classification statement for divisible codes. Extendability results for partial spreads concluded from minihypers can be obtained via divisible codes. Generalizations also permit extendability results for codes in the rank metric or subspace codes.

The aim of this talk is to give a brief introduction to the connection between multisets of subspaces and divisible codes, survey some results and applications of q^r divisible codes, and, most importantly, to encourage collaboration from other participants.

3.11 A new rank metric codes based encryption scheme

Pierre Loidreau (DGA MI, France Université de Rennes 1, France)
pierre.Loidreau@univ-rennes1.fr

We present new approaches for the design of code-based rank metric public-key encryption schemes based on Gabidulin codes. Compared to the rank metric code based NIST proposals using probabilistic decoding of LRPC codes, the use of algebraic codes enables to design schemes no decoding errors. Hence as in the original McEliece cryptosystem but contrarily to MDPC and lattice based cryptography, an attacker does not get any statistical advantage of a decoding failure to attack the scheme.

3.12 Minimal linear codes from functions over finite fields

Sihem Mesnager (Universities of Paris XIII (LAGA), France Paris VIII, France)
smesnager@univ-paris8.fr

Minimal linear codes have significant applications in secret sharing schemes and secure two-party computation. In the literature, there are two generic constructions of linear codes. The first one is based on functions over finite fields and second one is based on defining sets (which could be the support or the complementary of the support of functions over finite fields). We shall review some recent constructions of minimal linear codes based on the generic constructions and present those obtained from weakly regular bent functions (2017) and from weakly regular plateaued functions (2018).

3.13 On Ferrers Diagram Codes

Heide Gluesing-Luerssen (University of Kentucky, US)
heide.gl@uky.edu

Ferrers diagram codes are rank-metric codes where all matrices have support in a given Ferrers diagram. With the aid of a lifting construction such codes can be used to construct large subspace codes. In this talk I will report on progress in the construction of maximal Ferrers diagram codes (for any given Ferrers diagram and rank distance). While various methods exists, the general conjecture about the maximum possible dimension of a Ferrers diagram codes with given rank distance remains widely open. In the second part of the talk I will discuss the proportion of maximum Ferrers diagram codes in the space of all Ferrers diagram codes with the same shape and the according dimension.

3.14 Constructions and Classifications of MRD codes

John Sheekey (University College Dublin, Ireland)
john.sheekey@ucd.ie

We will present the known constructions and classifications for maximum rank distance codes, and survey their connections to objects such as spreads and semifields. We will discuss recent results which show that MRD codes consisting of square matrices over the binary field are surprisingly rare in some cases.

3.15 Resilient LTI systems with redundant signals

Margreta Kuijper (University of Melbourne, Australia)
mkuijper@unimelb.edu.au

This talk looks at linear time-invariant (LTI) systems, such as autopilot systems. Such systems involve input, output and state signals. In recent years there have been several events where such systems have been cyber attacked, resulting in malfunctioning and damage. There is currently a need for an automated response as part of the resilience of the system.

In this talk I will explore several fundamental ideas around linear systems under attack. One of these involves the scenario where a restricted number of system outputs (sensor measurements) is vulnerable to adversarial attacks that take place over time. I will introduce the fundamental notion of a system's "security index" as an analogon of "minimal distance" in linear coding theory. I will show how ideas from coding theory can be used for the purpose of attack detection and automated attack correction.

3.16 Parametrizing Systematic Gabidulin Codes and Applications

Alessandro Neri (University of Zurich, Switzerland)
alessandro.neri@math.uzh.ch

Rank metric codes have been introduced in 1978 by Delsarte, and independently by Gabidulin in 1985 and Roth in 1991. These codes are linear subspaces of the space of $n \times m$ matrices over a finite field F_q, but they can also be seen as sets of vectors of length n over an extension field F_{q^m}. Codes that are optimal in this metric are called Maximum Rank Distance (MRD) codes. The first and most studied family of MRD codes is given by the so-called generalized Gabidulin codes, and they represent the analogue of generalized Reed-Solomon (GRS) codes for the rank metric.

In the talk we examine the structure of generalized Gabidulin codes, in comparison with GRS codes. We focus in particular on their encoders. In analogy with GRS codes, we show that the systematic generator matrices of Gabidulin codes have a Cauchy-like structure. Finally, some possible research directions are presented.

3.17 Spectral Methods for Coding in a Non-Commutative Setup

Marcus Greferath (Aalto University, Finland)
marcus.greferath@aalto.fi

In algebraic coding theory, the BCH bound is probably one of the most impressive examples, showing how a Fourier transform can be used in order to construct codes of prescribed minimum distance. So far, this spectral technique is basically restricted to cyclic codes, however there is no strict reason to keep it limited to this case. This talk is particularly interested in the scenario, where a non-commutative finite group describes the co-ordinate domain. We will sketch the successful development of a Fourier theory for this setting and observe a few strange facts.

3.18 Functional PIR and batch codes

Yiwei Zhang (Technion – Haifa, Israel)
ywzhangcs.technion.ac.il

Codes with locality and/or availability have been extensively studied in recent years, including locally repairable codes (LRC), PIR codes, batch codes, etc.

Usually in such a code of dimension s, we focus on the recovering sets only for the s information symbols. We propose a natural generalization of PIR and batch codes, named functional PIR codes and functional batch codes, by analyzing the recovering sets for arbitrary vectors of length s. In this talk we present some bounds and constructions for functional PIR and batch codes. This is a joint work with Prof. Tuvi Etzion and Prof. Eitan Yaakobi.

3.19 Coding for Stochastic Gradient Descent

Rawad Bitar (Rutgers University – Piscataway, US)
rawad.bitar@rutgers.edu

We tackle the problem of stragglers, slow or unresponsive machines, in distributed machine learning algorithms. We consider the setting in which a Master wants to run a machine learning algorithm on tremendous amount of information. The Master offloads the computational tasks to worker machines. Straggler workers are the bottleneck of such systems. Gradient Coding has been recently proposed to mitigate stragglers in distributed gradient descent. We propose stochastic gradient coding that allows graceful performance degradation with the number of stragglers when using stochastic gradient descent.

3.20 The problem of constructing complete MDP convolutional codes over small fields

Julia Lieb (University of Aveiro, Portugal)
jlieb@ua.pt

It has been shown that, transmitting over an erasure channel, maximum distance profile (MDP) convolutional codes have optimal recovery rate for windows of a certain length. Additionally, the subclass of complete MDP convolutional codes has the ability to reduce the waiting time during decoding.

The existence of (complete) MDP convolutional codes for arbitrary parameters has been shown for sufficiently large field sizes. Moreover, there exist basically two general construction techniques for these codes. However, these constructions require very large field sizes.

In this talk, I will show that it turns out to be hard to find constructions over small fields even for quite small parameters although we could show that these codes exist over much smaller fields than those of the known constructions. Finally, some very particular construction examples for moderate field sizes should be presented.

4 Working groups

4.1 Secret sharing in the presence of adversaries

4.1.1 Secret sharing

Consider the scenario that n parties wish to store a secret securely and reliably. To this end, a dealer distributes the secret into n shares, i.e. one share for each party, such that 1) (reliability) a collection $\mathcal{A} \in 2^{\{1,\ldots,n\}}$ of "authorized" subsets of the parties can decode the secret, and 2) (secrecy) a collection $\mathcal{B}$ of blocked subsets of the parties cannot collude to deduce any information about the secret. A scheme to distribute the secret into shares with respect to access structure $(\mathcal{A}, \mathcal{B})$ is called a secret sharing scheme, initially studied in the seminal works by Shamir and Blakeley. An important application of secret sharing schemes

is distributed storage of private data, where each party is a storage node. Besides, secret sharing is a fundamental cryptographic primitive and is used as a building block in numerous secure protocols. In our discussions we focus on secret sharing schemes for the threshold access structure, i.e. $\mathcal{A}$ contains all subsets of $\{1, \ldots, n\}$ of size at least $n - r$ and $\mathcal{B}$ contains all subsets of $\{1, \ldots, n\}$ of size at most z. In other words, the secret can be decoded in the absence of any r parties, and any z parties cannot collude to deduce any information about the secret. We define the threshold t of the scheme as $t = n - r - 1$, i.e. any set of at least $t + 1$ parties are authorized and any set of less parties are blocked.

4.1.2 Initial thoughts

In our initial discussions we looked at the 1981 McEliece & Sarwate paper on secret sharing via Reed-Solomon codes. Without errors/adversaries the secret sharing is accomplished as erasure decoding. However, some of the shares may be in error. The main result of the paper is that errors-and-erasures decoding can be used to still do secret sharing in the presence of errors. After discussing this paper we decided to focus on secret sharing with errors/attacks in a broader context. Fuzzy vaults and secret sharing over networks were discussed. A motivation for this area is for example biometric authentication in the presence of errors/attacks.

We then looked at a number of other papers. The main ones being

1. the 2016 Eurocrypt paper *Essentially optimal robust secret sharing with maximal corruptions* by Bishop, Pastro, Rajaraman and Wichs.
2. the 2015 ITW paper *Talking reliably, secretly, and efficiently: a "complete" characterization* by Zhang, Kadhe, Bakshi, Jaggi and Sprintson. This paper uses hash functions to identify which shares were corrupted by the adversary, and then erasure decoding in the valid shares to recover the secret. Compared to straight forward error decoding this trick allows for smaller rate codes, and hence to reach the channel capacity asymptotically.

4.1.3 Secret sharing over a network with adversaries

Problem Setting: A dealer S wants to share a secret among n parties, such that any(?) k out of these n can recover it. The source uses a random linear network channel to each participant. The adversarial has access to at most t links of the network, which he can both eavesdrop and corrupt and introduce adversarial errors into the network during the distribution of the shares. A similar model was considered in Distributed Secret Dissemination Across a Network by Nihar B. Shah, K. V. Rashmi and Kannan Ramchandran. They developed a protocol to distribute shares efficiently, assuming the network is sufficiently well connected (k-propagating). Their scheme uses predefined channel coefficients to distribute the shares amongst the parties, but assumes that all nodes in the network are honest-but-curious, i.e. they have access to all passing messages, but do not corrupt the messages.

We want to extend their model in two important ways (see Fig. 1):

1. Instead of assuming a given/known topology and precomputing coefficients for each node, we want to assume that each inner node forwards a random linear combination of their incoming information. As known from classical random network coding theory, in this setting errors can/will spread and possibly affect all shares. To be able to detect and correct these adversarial types of errors, we want to use rank-metric or subspace codes.
2. A limited number of links are assumed to be malicious and will introduce errors into the system.

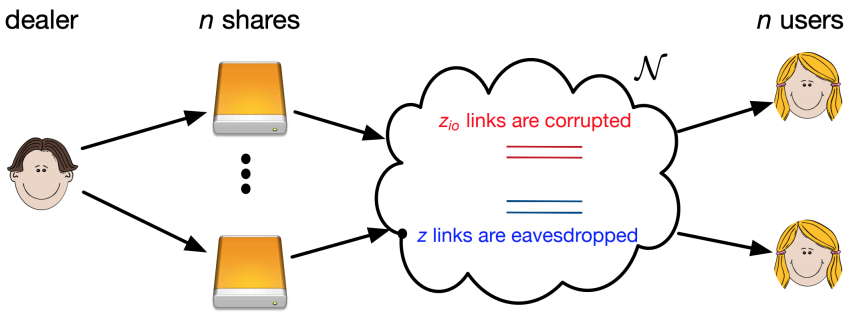

Figure 1 Proposed system model.

Things to be done:

- Establish sufficient conditions on the network to allow for a k-out-of-n secret sharing protocol.
- Find suitable codes for this setting, such that the shares are (related to) the codewords.
- Find means of identifying corrupted shares (and possibly do an analogue of erasure decoding in the network coding setting).
- Find a way to add randomness to each share, to weaken the adversary and to prevent any eavesdropper to gain information about the secret.
- ...

4.2 Code-Based-Cryptography

In this working group, we discussed cryptography based on quasi-cyclic moderate density parity-check (QC-MDPC) codes and based on low-rank parity-check (LRPC) codes. There are a several important considerations about these systems which may be interesting from a coding-theoretic perspective.

- The parity-check matrix for a QC-MDPC code of length $2n$ is defined by two polynomials, modulo $x^n - 1$. In the event that n is not chosen as a prime and $x^n - 1$ can be factored, is it possible to leverage this factorization in reducing the complexity/search space for low-weight codewords?
- MDPC codes are not interesting for classical coding purposes, and for this reason, relatively little analysis of decoding properties of these codes has occurred. Typically, a simple bit flip algorithm is implemented, and there is some risk of decoding failure. Simulation results for current parameters of interest indicate an error rate which is considered too high a rate for serious consideration in cryptographic application. Are there better decoders that can be easily implemented? Are there theoretic results to support improved performance or offer performance guarantees?
- Do there exist quantum algorithms which might break quasi-cyclic syndrome decoding (QCSD)?
- For certain parameters, there exists a distinguisher for LRPC codes based on a containment in (sums of) Gabidulin codes. Much like the Schur square distinguisher in the case of generalized Reed-Solomon codes, this distinguisher reveals a space of dimension much lower than expected for a random code. Is it possible to extend this distinguisher to other parameters using a strategy of puncturing or shortening?

- Fundamentally, the application of lattices, MDPC codes, or LRPC codes in cryptography represent the same framework under three different metrics: Euclidean, Hamming, and rank metric. Taking this global view is helpful in illustrating the advantages and disadvantages of each metric. We considered the specific application of hash functions, wherein the Euclidean system has many good parameters, the Hamming system has some good parameters, and the rank system has no good parameters. What other conclusions can we reach based on these differences? While theory may support the possibility of each perspective, what happens when we push to establish actual security parameters?
- What new types of codes are promising in cryptographic application? Consider bivariate Gabidulin "Reed-Muller" codes over general finite fields, as discussed in the working group on algebraic geometry codes.
- What algorithms exist for decoding random linear rank-metric codes?
- For serious consideration in post-quantum standardization, it is critical going forward to work toward security reductions for code-based cryptography.

4.3 Adversarial channel

Many areas of research fall under the category of adversarial channels. In this working group, we discussed several adversarial settings and interesting problems within each. Specifically, the following topics were identified:

- Reliable communication over the arbitrarily varying channel (AVC): In this setting, channel statistics may change according to the state imposed by an adversary. There are many variations in which the adversary has a varying amount of information regarding the transmitted signal. In particular, coding strategies for authentication over the AVC, and coding strategies to gain information about adversarial power were considered.
- Covert communication: Here, we wish to communicate without an adversary identifying that a transmission is being sent. The signal should approximately mimic the noise of the channel. One potential direction is to adapt the code construction techniques of Ahlswede and Dueck in [1].
- Confusability graphs and zero-error capacity: We briefly discussed the confusability graph framework for studying the zero-error capacity of adversarial channels. Several resources discussed are [3], [4], and[2].
- Network reconstruction: We wish to reconstruct a network given some information from each of the nodes (e.g. the structure of a network of friends on social media), where a node or some subset of nodes are either malicious or (unintentionally) untrustworthy. Several parameters to consider are the nature of the shared information and a metric for measuring the accuracy of a reconstruction. Given precise definitions for each of these things, the question is whether we can characterize the amount of damage an untrustworthy node is capable of.

References

1 "Good codes can be produced by a few permutations", Rudolf Ahlswede and Gunter Dueck, *IEEE transactions on information theory*, 28(3), 1982 .
2 "The zero error capacity of a noisy channel", Claude Shannon, *IRE Transactions on Information Theory*, 2(3):8–19, 1956.
3 "On the Shannon capacity of a graph", László Lovász, *IEEE Transactions on Information theory*, 25(1):1–7, 1979.
4 "Adversarial network coding", Alberto Ravagnani and Frank R. Kschischang, *IEEE Transactions on Information Theory*, 65(1):198–219, 2019.

4.4 Algebraic geometry codes

The first two days we reviewed algebraic plane curves over finite fields, the notions of divisors, Riemann-Roch space and all the tools needed to construct algebraic geometry codes (AG codes for short). We pointed out the connection between the discrete logarithm problem and the elliptic curves. Finally, as an exercise, we found the equivalent version of the Berlekamp-Welch decoding algorithm for Reed-Solomon codes in the AG setting.

The last day we dealt with the case of a Galois group which is the product of cyclic groups, like $C_l \times \ldots \times C_l$ and to try to define Reed-Muller codes, by defining multivariate skew polynomial rings. The exercise was to try to define the codes and compute their minimum distance.

4.5 Service Rates of Codes

Consider a system storing K files $(f_1, \ldots, f_K)$ over N nodes (labelled $1, \ldots, N$) using an (N, K) code. Suppose that file f_i has t_i recovery sets, $R_1^{(i)}, \ldots, R_{t_i}^{(i)}$, and let μ_l be the service rate of node l (i.e. the average rate at which node l resolves received file request). Denote by λ^i the rate of requests for file f_i and $\lambda_j^{(i)}$the portion of requests for file f_i that are assigned to $R_j^{(i)}$.

Then the achievable service rate region of such a system is the set of vectors $(\lambda^1, \ldots, \lambda^K)$ such that, for every $1 \leq i \leq K$, there exist $\lambda_j^{(i)}$, $1 \leq j \leq t_i$, satisfying the following:

$$\sum_{j=1}^{t_i} \lambda_j^{(i)} = \lambda^i, \qquad 1 \leq i \leq K \tag{1}$$

(The demand for file f_i is served.)

$$\sum_{i=1}^{K} \sum_{j : l \in R_j^{(i)}} \lambda_j^{(i)} \leq \mu_l, \qquad 1 \leq i \leq N \tag{2}$$

(No node is sent requests in excess of its service rate.)

$$\lambda_j^{(i)} \geq 0, \qquad 1 \leq i \leq K, 1 \leq j \leq t_i \tag{3}$$

(Requests send to repair groups are nonnegative.)

Given $K-1$ arrival rates $\lambda^1, \ldots, \lambda^{K-1}$ we would like to find the maximum

$$\lambda^K = \sum_{j=1}^{t_K} \lambda_j^{(K)}$$

subject to the constraints described in (1)-(3).

In this group, we discussed a variety of approaches to open problems about service rate. Two problems that arose are introduced below.

1. Given constraints on file request rates, how can we construct a code serving all requests in the described region?

When constructing a code, we may be interested in minimizing the number of nodes used. Note that a minimal length code serving a given set of requests does not always have the minimum length. Other possible considerations include the number of coded nodes and the utilization of each node in a coding scheme.

1. How can we determine the achievable service rate region for given families of codes?

Results about the service rate regions for Simplex Codes and MDS codes rely on the structure of the repair groups. Classifying the structure of repair groups of other families of codes may help determine the service rate region.

These problems can be considered in various settings. Settings proposed within our group use methods or structures including (i) linear optimization, (ii) batch codes, (iii) graph covers, or (iv) generator matrices.

Developing tools to investigate service rate regions from several perspectives will make this emerging area accessible to a wider research community.

Participants

- Gianira Nicoletta Alfarano
Universität Zürich, CH
- Iryna Andriyanova
University of Cergy-Pontoise, FR
- Daniel Augot
INRIA Saclay – Île-de-France, FR
- Allison Beemer
Arizona State University – Tempe, US
- Rawad Bitar
Rutgers University – Piscataway, US
- Jessalyn Bolkema
SUNY – Oswego, US
- Martin Bossert
Universität Ulm, DE
- Eimear Byrne
University College Dublin, IE
- Giuseppe Cotardo
University College Dublin, IE
- Marwa Dammak
University of Cergy-Pontoise, FR
- Salim El Rouayheb
Rutgers University – Piscataway, US
- Tuvi Etzion
Technion – Haifa, IL
- Alexey Frolov
Skoltech – Moscow, RU
- Philippe Gaborit
University of Limoges, FR
- Heide Gluesing-Luerssen
University of Kentucky, US
- Oliver Wilhelm Gnilke
Aalborg University, DK
- Marcus Greferath
Aalto University, FI
- Daniel Heinlein
Aalto University, FI
- Lukas Holzbaur
TU München, DE
- Anna-Lena Horlemann-Trautmann
Universität St. Gallen, CH
- Sidharth Jaggi
The Chinese University of Hong Kong, HK
- Jörg Kliewer
NJIT – Newark, US
- Stanislav Kruglik
Skoltech – Moscow, RU
- Margreta Kuijper
The University of Melbourne, AU
- Sascha Kurz
Universität Bayreuth, DE
- Julien Lavauzelle
Ecole Polytechnique – Palaiseau, FR
- Hunter Lehmann
University of Kentucky, US
- Andreas Lenz
TU München, DE
- Julia Lieb
University of Aveiro, PT
- Pierre Loidreau
University of Rennes, FR
- Felice Manganiello
Clemson University, US
- Carolyn Mayer
Worcester Polytechnic Institute, US
- Sihem Mesnager
University of Paris VIII, FR
- Alessandro Neri
Universität Zürich, CH
- Cornelia Ott
Universität Ulm, DE
- Mario Osvin Pavcevic
University of Zagreb, HR
- Sven Puchinger
TU München, DE
- Alberto Ravagnani
University College Dublin, IE
- Julian Renner
TU München, DE
- Joachim Rosenthal
Universität Zürich, CH
- Ronny Roth
Technion – Haifa, IL
- John Sheekey
University College Dublin, IE
- Carmen Sippel
Universität Ulm, DE
- Emina Soljanin
Rutgers University – Piscataway, US
- Razane Tajeddine
Aalto University, FI
- Violetta Weger
Universität Zürich, CH
- Eitan Yaakobi
Technion – Haifa, IL
- Yiwei Zhang
Technion – Haifa, IL
- Jens Zumbrägel
Universität Passau, DE