Combinatorics in Information Security
PiWORKS talk

Sophie Huczynska
University of St Andrews

Work is joint with Maura Paterson, Birkbeck (University of London)

September 2021

Information security

Information security is concerned with the safe and private transmission and storage of data.

Motivating questions include:

- How can a message be sent so that we can detect whether it has been changed during transmission?
- If we detect that a change has occurred, can we recover the original message - and if so, how?
- How can we encrypt messages/data so that they cannot feasibly be decrypted by anyone other than the intended recipient?
- . . . and many more.

Manipulation detection

In this talk, we consider an encoding system and how to design it to minimise the chances that an undetected change can occur.

Applies to various situations:

- message transmission which is subject to attack
- storage device which is subject to tampering

We will be thinking in terms of the message-sending scenario.

It is helpful to model the situation as a "game" between an encoder and an adversary who is trying to "cheat" the encoder.

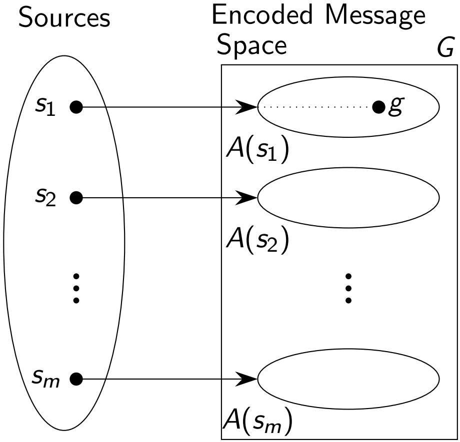
Our focus is on algebraic manipulation detection (AMD) codes.

AMD code model

We will have:

- Set S of plaintext sources (the messages)
- Encoded message space G (finite group, written additively)
- Encoding function E (possibly randomized) maps source $s \in S$ to some $g \in G$
- For each source $s \in S$, subset $A(s)$ of G is the set of valid encodings of s
- Unique decodability: $A(s) \cap A(s') = \emptyset$ if $s \neq s'$, i.e. the sets of encodings do not overlap

Diagram



The "game"

AMD code

Adversary: chooses a value $\delta \in G \setminus \{0\}$ (their "manipulation")

Encoder: chooses source $s \in S$

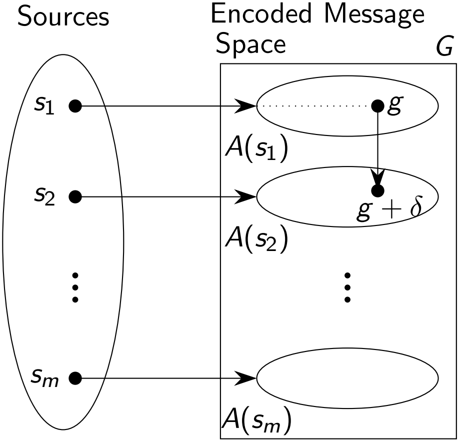
Encoder: s encoded by E to some $g \in A(s)$

Adversary: g is replaced by $g' = g + \delta$

Adversary wins if $g' \in A(s')$ for some $s' \neq s$

"The adversary wins if they succeed in shifting the group element g into an element $g + \delta$ that's an encoding of a different source"

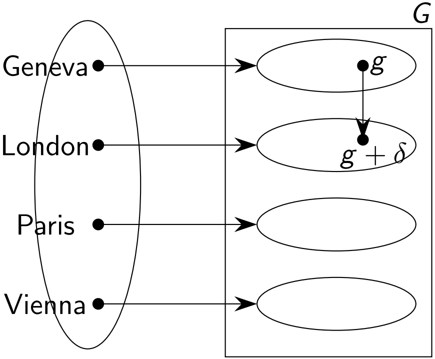
Diagram



If message s_1 is sent and encoded to g, it will be incorrectly decoded to s_2 after this manipulation. In this case, adversary wins!

Imaginary real-life example!

Kim sends message to Robin saying where to meet.
Adversary manipulates the encoded message by adding δ.



Kim sends "Geneva" which is encoded to g, but after manipulation Robin will receive $g + \delta$ and decode this to "London".
Adversary wins: Kim and Robin don't meet!

Combinatorial model

The AMD "game" can be modelled as a set-up in combinatorics.

We model the sender's choice of message probabilistically.

- Adversary chooses $\delta \in G \setminus \{0\}$
- Pick a set A_i uniformly at random (source)
- Then pick an element $d_i \in A_i$ uniformly at random (encoding)
- Adversary "wins" if $d_i + \delta \in A_j$ for some $j \neq i$

Adversary wins if δ occurs as a difference between our element in A_i and some element in A_j.

Need to look at the differences between elements of A_i and A_j.

Difference notation

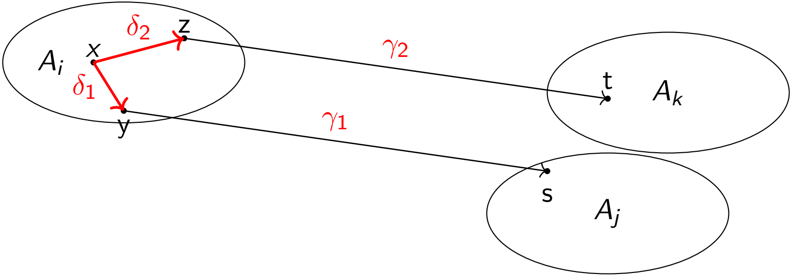
Suppose we have a disjoint family of subsets $A_1, \ldots, A_m$ of G

- For a fixed i, the differences between the elements of A_i are called internal differences:

$$I(A_i) := \{x - y : x, y \in A_i, x \neq y\}$$

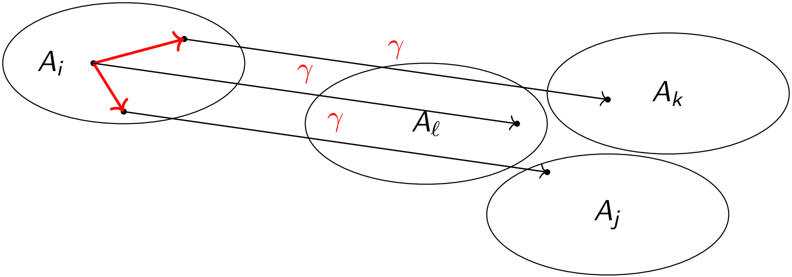
- For $i \neq j$, the differences between the elements of A_i and A_j are called external differences:

$$E(A_i, A_j) := \{x - y : x \in A_i, y \in A_j\}$$



In this diagram,

- δ_1 and δ_2 are internal differences in A_i
  $(x - y = \delta_1,\ x - z = \delta_2)$
- γ_1 and γ_2 are external differences out of A_i (to A_j, A_k resp.)
  $(y - s = \gamma_1,\ z - t = \gamma_2)$



For a disjoint family of sets $A_1, \dots, A_m$, define the number of times a non-zero element γ occurs as an external difference out of A_i by

$$N_i(\gamma) = |\{(x, y) : x - y = \gamma, x \in A_i, y \in A_j, j \neq i\}|$$

In the example above, we show all occurrences of γ as an external difference out of A_i, so $N_i(\gamma) = 3$ here.

Success probability

Returning to our AMD code:

The probability that an adversary succeeds when they pick δ is

$$e_\delta = \frac{1}{m}\left(\frac{N_1(\delta)}{|A_1|} + \cdots + \frac{N_m(\delta)}{|A_m|}\right) \tag{1}$$

where $|A_i|$ is the size of set A_i.

- Source i picked with probability $\frac{1}{m}$
- $N_i(\delta)$ of the possible $|A_i|$ encodings will lead to success for an adversary who picks δ

Which codes are best?

We are seeking AMD codes which are optimal (as good as possible from the sender's point of view).

We want the adversary's chance of success to be as low as possible.

Optimality corresponds to: probability that an adversary succeeds when they pick δ, is constant for all $\delta \in G \setminus \{0\}$.

For these: adversary's maximum success probability is equal to their average success probability.
No choice of δ is better than any other!

Our combinatorial problem

We have translated our requirements for an optimal AMD code, into a combinatorial problem.

We would like:

- a group G
- a set $\mathcal{A}$ of disjoint subsets $A_1, \ldots, A_m$ of G
- such that the following property holds:

$$\frac{1}{m}\left(\frac{1}{|A_1|}N_1(\delta) + \cdots + \frac{1}{|A_m|}N_m(\delta)\right) = \text{constant} \qquad (2)$$

for every $\delta \in G \setminus \{0\}$

RWEDFs

- Surprisingly, the set of combinatorial objects with this property has not previously been named or characterised.
- People have, however, looked at certain special cases.

We have called these objects reciprocally-weighted external difference families (RWEDFs).

Definition

An $(n, m; k_1, \ldots, k_m; \ell)$-RWEDF is a collection of disjoint subsets $A_1, \ldots, A_m$ of an abelian group G , where $|A_i| = k_i$ for all $i \in \{1, \ldots, m\}$, with the property that:

$$\frac{1}{k_1} N_1(\delta) + \cdots + \frac{1}{k_m} N_m(\delta) = \ell$$

for all non-zero $\delta \in G$.

Challenge: how to obtain such objects?

- Consider which already-studied objects in combinatorics may be useful
- Develop new existence/non-existence results of our own

A special case:

- If all the sets A_i have the same size, then the requirement becomes
$$N_1(\delta) + \cdots + N_m(\delta) = \text{ constant}$$
These have been studied: external difference families (EDFs).

EDF example

- Let $G = (\mathbb{Z}_{10}, +)$; take $A_1 = \{4, 7, 9\}$ and $A_2 = \{0, 2, 5\}$
- Differences from A_1 to A_2 are
  $\{4 - 0 = 4, 4 - 2 = 2, 4 - 5 = -1 = 9,$
  $7 - 0 = 7, 7 - 2 = 5, 7 - 5 = 2,$
  $9 - 0 = 9, 9 - 2 = 7, 9 - 5 = 4\}$, ie $\{2, 2, 4, 4, 5, 7, 7, 9, 9\}$.
- Differences from A_2 to A_1 are their negatives, i.e. $\{1, 1, 3, 3, 5, 6, 6, 8, 8, 8\}$.
- Union of all external differences=each nonzero element twice!
- For $\delta = 1$, the adversary's success probability is

$$\frac{1}{2}\left(\frac{N_1(\delta)}{|A_1|} + \frac{N_2(\delta)}{|A_2|}\right) = \frac{1}{2}\left(\frac{0}{3} + \frac{2}{3}\right) = \frac{1}{3}$$

- Same probability for any choice of $\delta \neq 0$.

Theoretical EDF construction

Construction

Let G be the additive group of $GF(q)$, the finite field of order q, where q is a prime power congruent to 1 mod 4.
Let $A_1 = \{$the set of squares in $GF(q)^*\}$.
Let $A_2 = \{$the set of non-squares in $GF(q)^*\}$.
Then $\{A_1, A_2\}$ form a $(q, 2; \frac{q-1}{2}, \frac{q-1}{2}; 1)$-RWEDF.

This is a special case of cyclotomic method - using multiplicative subgroups of a finite field to make EDFs in its additive group.

Examples of squares/nonsquares construction in $GF(q)$

- Let $q = 5$; take $A_1 = \{1, 4\}$ and $A_2 = \{2, 3\}$.
- Differences from A_1 to A_2 are $\{1 - 2 = 4, 1 - 3 = 3, 4 - 2 = 2, 4 - 3 = 1\} = \{4, 3, 2, 1\}$.
- Differences from A_2 to A_1 are their negatives, i.e. also $\{1, 2, 3, 4\}$.
- Each nonzero element of $(\mathbb{Z}_5, +)$ occurs twice as an external difference.
- So for any non-zero $\delta \in G$, adversary's success probability equals

$$\frac{1}{2}\left(\frac{N_1(\delta)}{|A_1|} + \frac{N_2(\delta)}{|A_2|}\right) = \frac{1}{2}\left(\frac{1}{2} + \frac{1}{2}\right) = \frac{1}{2}$$

What about RWEDFs which are not EDFs?

Now we would like to construct examples of RWEDFs which have genuinely different set-sizes (i.e. are not EDFs).

- Q: Do such things exist?
- A: Yes!

Examples of RWEDFs which are not EDFs

Example

Let $G = \mathbb{Z}_{k_1 k_2 + 1}$. The sets

$$A_1 = \{0, 1, \ldots, k_1 - 1\} \text{ and } A_2 = \{k_1, 2k_1, \ldots, k_1 k_2\}$$

form a $(k_1 k_2 + 1, 2; k_1, k_2; \frac{1}{k_1} + \frac{1}{k_2})$-RWEDF.

Can prove: this give an AMD code whose success probability is as small as possible (smallest ℓ) for $m = 2$.

Example

Take $k_1 = 3$ and $k_2 = 4$.
Then $G = \mathbb{Z}_{13}$, $A_1 = \{0, 1, 2\}$ and $A_2 = \{3, 6, 9, 12\}$.

Differences out of A_2:
$3 - 0 = 3, 3 - 1 = 2, 3 - 2 = 1,$
$6 - 0 = 6, 6 - 1 = 5, 6 - 2 = 4,$
$9 - 0 = 9, 9 - 1 = 8, 9 - 2 = 7,$
$12 - 0 = 12, 12 - 1 = 11, 12 - 2 = 10$; i.e. $N_2(\delta) = 1$ for all δ.

Differences out of A_1 are negatives of these: $N_1(\delta) = 1$ for all δ.
For each non-zero $\delta_\in G$, adversary's success probability is

$$\frac{1}{2}\left(\frac{N_1(\delta)}{|A_1|} + \frac{N_2(\delta)}{|A_2|}\right) = \frac{1}{2}\left(\frac{1}{3} + \frac{1}{4}\right) = \frac{1}{2}\frac{7}{12} = \frac{7}{24}$$

Difference sets

Difference sets have been much-studied by mathematicians.

Definition

A difference set in a group G is a set $D \subseteq G$ such that, when we take all pairwise internal differences between the elements of D, every non-identity group element occurs a fixed number λ of times.

Example: $\{1, 2, 4\}$ is a difference set in $\mathbb{Z}_7$ with $\lambda = 1$ - each non-zero element of $\mathbb{Z}_7$ occurs once as a difference.
To see this: $4 - 1 = 3, 4 - 2 = 2, 2 - 1 = 1, 2 - 4 = -2 = 5, 1 - 2 = -1 = 6, 1 - 4 = -3 = 4$.

RWEDFs from difference sets

Theorem

Let G be a group of order n, and let $\mathcal{A} = \{A_1, A_2\}$ partition G. Then $\mathcal{A}$ is an RWEDF $\Leftrightarrow$ A_1 and A_2 are difference sets.

Example: Let $G = \mathbb{Z}_7$.
Let $A_1 = \{1, 2, 4\}$ and $A_2 = \{0, 3, 5, 6\}$.
Then $\{A_1, A_2\}$ is a $(7, 2; 3, 4; \frac{7}{6})$-RWEDF.
For any δ, adversary's success probability is $\frac{7}{12}$.

New families of RWEDFs

Observe that all the examples we have seen so far have 2 sets, i.e. $m = 2$.

Q: Can we get examples with $m > 2$?

Also, notice that the constant ℓ in the definition is in $\mathbb{Q}$ but not necessarily $\mathbb{Z}$.

Q: Can we obtain new constructions for RWEDFs, which have integer ℓ?

Some group theory

Definition

If a finite group G has subgroups $S_1, \ldots S_m$ with the property that $S_1 \setminus \{0\}, \ldots, S_m \setminus \{0\}$ partition G, then the collection of subgroups is called a partition of G.

Groups which have a partition include:

- elementary abelian p-groups of order $\geq p^2$, for p prime
- Frobenius groups (eg dihedral group D_{2n} with n odd)
- groups of Hughes-Thompson type
- groups isomorphic to $PGL(2, p^h)$ with p an odd prime

New group theoretic construction

We can prove the following:

Theorem

Any partition of a finite group G forms an RWEDF with integer ℓ.

Construction: take $A_1 = S_1 \setminus \{0\}, \ldots, A_m = S_m \setminus \{0\}$.

Interestingly, this holds for any group, not just abelian; so we can begin to study RWEDFs in non-abelian situations.

Although motivated by finding non-EDF RWEDFs, this also gives previously-unknown constructions for new EDFs!

RWEDF example

Example of group partition construction:

- Let $G = \mathbb{Z}_3 \times \mathbb{Z}_3$.
- Let $A_1 = \{(1,1),(2,2)\}$, $A_2 = \{(0,1),(0,2)\}$, $A_3 = \{(1,2),(2,1)\}$ and $A_4 = \{(1,0),(2,0)\}$.
- Note these are subgroups with $\{(0,0)\}$ removed in each case.
- For non-zero $\delta \in G$, $N_i(\delta) = 2$ for $\delta \notin A_i$ and $N_i(\delta) = 0$ for $\delta \in A_i$ (for each $1 \leq i \leq 4$).
- $\mathcal{A}$ forms a $(9,4;2,2,2,2;3)$-RWEDF (indeed, this is an EDF).

We can explore different choices of groups to fine-tune success probability.

Open questions

There are many avenues to explore further in this area.

- New constructions for RWEDFs which are not EDFs
- Partitioned external difference families - intermediate case
- Fine-tune our constructions to yield smallest possible success probabilities.

Some key messages before thinking about a PhD

- Enjoying Maths and being qualified to do a PhD is enough!
- Maths can be a great part of your life but need not be your whole life.
- Don't compare yourself to fellow students who may project an image of greater knowledge/certainty..
- PhDs give the opportunity to both research and teach, and subsequent academic jobs also involve both research and teaching.

Choosing a PhD

- Your supervisor and your working relationship with them is crucial.
- Make sure your supervisor will have time for you and that you feel comfortable with them.
- Make sure you are happy with the location and that it works with the non-work parts of your life.
- Choosing an area of maths that suits you is important, but there will be some flexibility to move sideways later.

During your PhD

- Try to develop the ability to speak out/ask when something isn't clear to you.
- Many people around you are bluffing and also don't know.
- Be aware: there is a culture of brevity in maths papers/talks which can make straightforward things unclear by removing intermediate steps.
- There is a culture of removing the process from proof write-ups - don't compare your process with other people's final product.
- Remember that you know more about what you are working on than anyone else.

General observations

- There are many different ways of being a mathematician - find the maths/life balance which works for you.
- Everyone benefits when universities contain practising mathematicians with different approaches and styles.
- Don't let yourself be intimidated.
- Be flexible and remember there are many different possible routes.