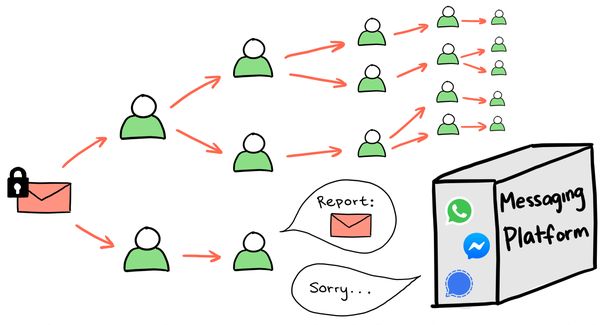
Source-Tracking for Encrypted Messaging Systems

Charlotte Peale, Saba Eskandarian, Dan Boneh
Department of Computer Science, Stanford University

The Problem: How can Messaging Platforms Control the Spread of Viral Misinformation?

Secure messaging platforms are unable to enforce the same type of content moderation employed by sites such as Twitter to prevent the viral spread of misinformation and malicious content via forwards.



Even if a user reveals a message to the platform, there is not much the platform can do in response.

A Solution: Hold the Sources of Misinformation Accountable

Source-tracking for encrypted messages would allow a user to report a particular message and reveal the original source of the forwarding chain to the messaging platform.

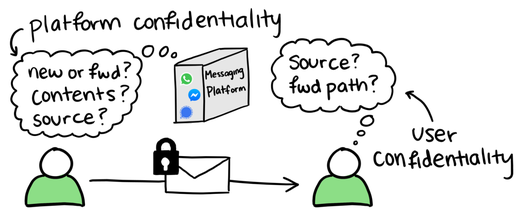
After a report, the source of a forwarded message is revealed.

Related Work

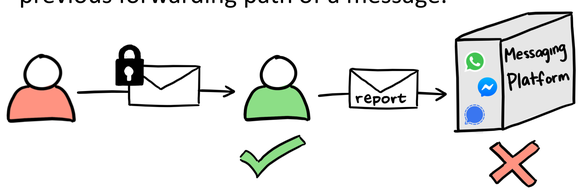
- *Message traceback (*Tyagi et al., 2019): Platform can recover the entire path of a forwarded message.
  - Requires platform storage (linear in total msgs).
  - Revealing an entire message path could unnecessarily expose innocent users.
- *Message franking* (Facebook): Links plaintexts to the previous sender upon report.
  - Only tracks most recent forwarder, not source.

What Properties Should a Source-Tracking Scheme Satisfy?

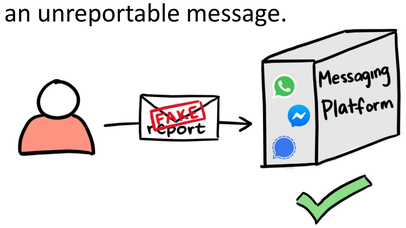
Secure schemes should guarantee *confidentiality, unforgeability, accountability, and deniability:*



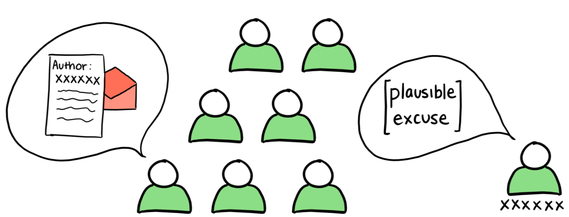
1. **Confidentiality**: The messaging platform should be unable to discern the contents or type of messages being sent, and users should learn nothing about the previous forwarding path of a message.



2. **Accountability**: An attacker shouldn't be able to create an unreportable message.



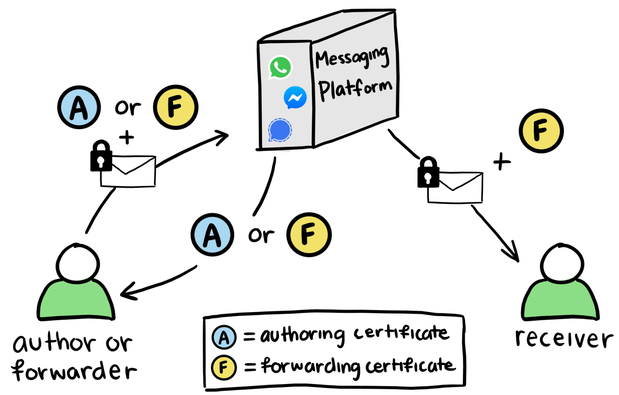
3. **Unforgeability**: An attacker shouldn't be able to frame a user as the source of a message they didn't send.



4. **Deniability**: Only the platform should be able to verify that a report linking a user to a message is valid.

Scheme 1: Source-Tracking from Blind Signatures

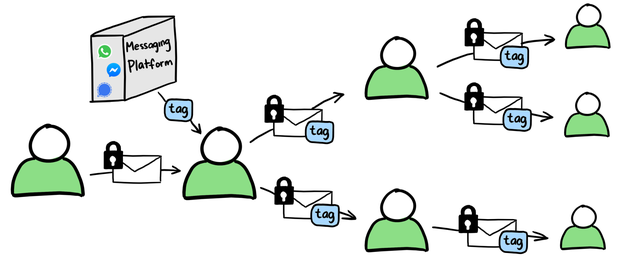
- Built on *blind signatures with attributes* (Baldimtsi and Lysyanskaya, 2013).
  - Allow a user to get a signature on a message linked to some commitment to the user's particular attributes.
  - Blinding ensures valid signatures cannot be linked to the issuing transaction.
- Our scheme:
  - (message, source) pairs are attributes.
  - Blind signatures on pairs are one-time-use 'credits' that a user can present to prove the validity of the message they want to forward.



- Sender and receiver get a fresh signature on the (message, source) pair, used to unlinkably forward the message in the future.
- Reporting:
  - User presents a valid signature and opening for the associated commitment.
  - Reveals to platform an encryption of the source user's identity under the platform's public key.
- An interactive scheme requiring 2 rounds of communication to send or receive a message.

Scheme 2: A Simpler Scheme with Weakened Confidentiality

- Satisfies a weaker, but still practical form of confidentiality.
  - Users can distinguish between different instances of the same message plaintext.
- Also employs commitments and signatures in order to link a source to a message, but these tags stay constant throughout the forwarding path.
- Tags for each message are much shorter and computationally simple.
- Sending, receiving, and reporting transactions can be done noninteractively.



Future Directions

- Work toward more efficient schemes would improve the practicality of implementing source-tracking.
- While source-tracking is designed with good intentions, there is the undeniable possibility that it could be abused to assist in enforcing censorship, etc. Considering methods for abuse mitigation in source tracking schemes is therefore an important research direction.

Acknowledgements

I am incredibly grateful to my mentors Saba Eskandarian and Dan Boneh for their encouragement and guidance this summer. I would also like to thank the applied cryptography and security groups for being so welcoming. Lastly, thank you to the CURIS program for giving me the opportunity to work on this project.

Wedge-Lifted Codes and the Disjoint Repair Group Property

Jabari Hastings, Amy Kanne, Ray Li, Mary Wootters

Stanford
Computer Science

Abstract

In this work, we construct new codes, which we call *wedge-lifted codes,* that satisfy a notion of locality referred to as the t-disjoint-repair group property (t-DRGP). These codes, which can be seen as a variant of *lifted codes* introduced by Guo, Kopparty and Sudan, consist of the evaluations of multivariate polynomials whose restriction to a configuration of intersecting lines is a Reed-Solomon codeword. We show that wedge-lifted codes of length N are less redundant than previously known constructions with small alphabet size that satisfy the $\sqrt[2m]{N}$-DRGP for some $m \in \mathbb{N}$.

Presenter

Jabari Hastings BS '20, MS '21

Background

A *code* of block length N over alphabet Σ is a subset $C \subseteq \Sigma^N$. Codes protect messages from errors or corruption by adding redundancy.

For example, a *grid code* $C \subseteq \{0,1\}^N$ protects a a $\left(\sqrt{N}-1\right) \times \left(\sqrt{N}-1\right)$ message by adding parity checks to each row and column.

1	0	0	1
1	1	0	0
1	1	1	1
1	0	1	0

Fig 1. Binary grid codeword.

If a symbol were to get corrupted, then we can recover it in two separate ways at the same time. We can sum along either the column or the row that contains the corrupted symbol.

1	0	0	1
1	1	0	0
1	1	1	1
1	0	1	0

1	0	0	1
1	1	0	0
1	1	1	1
1	0	1	0

Fig 2. Rows and columns satisfy parity checks.

Generalizing this notion, we say that a codeword $c \in C$ has the *t-disjoint repair group property* (t-DRGP) if for every symbol c_i there exist disjoint subsets $S_1, \ldots, S_t \subseteq [N] \backslash \{i\}$ and functions $f_1, \ldots, f_t$ such that $c_i = f(c|_{S_i})$

Can we get low-redundancy codes with the t-DRGP for other values of t apart from 2 by applying more parity checks?

Construction

Let $\overrightarrow{p} = (x, y) \in \mathbb{F}_q^2$ be a point and $H \subseteq \mathbb{F}_q$ be a set. The $[H, \overrightarrow{p}]$-*wedge* is the set of affine lines with slope in H that pass through $\overrightarrow{p}$:

$$\left\{(T, \alpha(T-x)+y) : T \in \mathbb{F}_q, \alpha \in H\right\}$$

Given a family of disjoint sets $\mathscr{H} = \{H_1, \ldots, H_t\}$ we want to assign values to each point in the plane so that the points on each wedge satisfy a parity check.

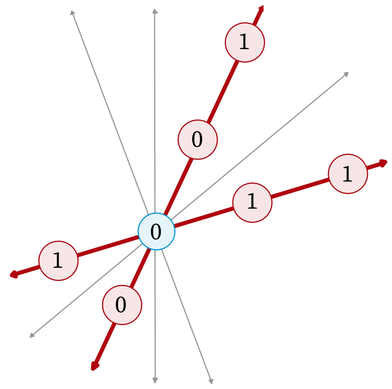


Fig 3. Points on wedge satisfy a parity check.

We use bivariate polynomials to assign the values. We choose $P \in \mathbb{F}_q[X, Y]$ such that for each $H \in \mathscr{H}$ and each $(x, y) \in \mathbb{F}_q^2$, we have

$$\sum_{T \in \mathbb{F}_q} \sum_{\alpha \in H} P(T, \alpha(T-x)+y) = 0 \qquad (\star)$$

When $H < \mathbb{F}_q^\times$ is a *multiplicative subgroup* and $H_1 = H, \ldots, H_t$ are *cosets,* the following hold:

- $\mathscr{H}$ partitions $\mathbb{F}_q^\times$ into t disjoint sets.
- For any point $\overrightarrow{p}$ and distinct sets H_i, H_j, the wedges $[H_i, \overrightarrow{p}]$ and $[H_j, \overrightarrow{p}]$ only share $\overrightarrow{p}$.

This gives us the t-DRGP!

Results

The redundancy of wedge-lifted codes is related to the number of 'good monomials'—those which satisfy ($\star$).

Proposition. *Let* $a + b < 2(q-1)$ *and* $H < \mathbb{F}_q^\times$ *be a subgroup with* $|H| > 1$. *The monomial* $X^a Y^b$ *is good if and only if* $a \vee b = q-1$ *or there does not exist a whole number* i *with* $i \equiv b \mod |H|$ *such that* $i \wedge (a \wedge b) = i$.

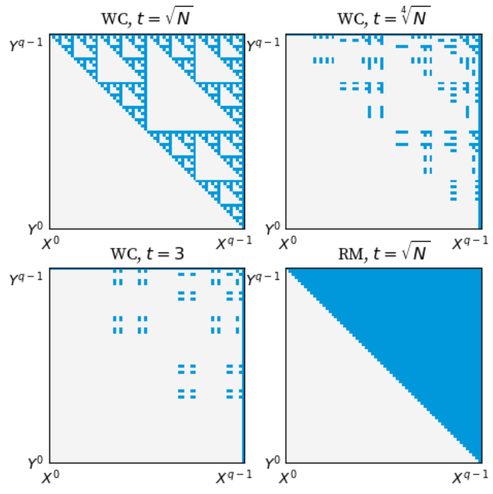


Fig 4. Distribution of good monomials. Gray areas indicate good monomials and blue areas indicate bad monomials.

Theorem. (Main Result) *Let* $q = 2^\ell$ *and* $m \in \mathbb{N}$ *satisfy* $q^{1/m} - 1 \mid q - 1$. *Let* $H < \mathbb{F}_q^\times$ *be a multiplicative subgroup of order* $\frac{q-1}{q^{1/m}-1}$ *and* $\mathscr{H}$ *be the set of its cosets. Let* C *be the corresponding wedge-lifted code.*

- *The block length is* q^2
- *The redundancy is at most* $(2^{m+1}-1)^{\ell/m}$
- *The code has the* $(q^{1/m}-1)$*-DRGP*

Our construction has redundancy at most that of existing codes with small alphabet sizes for the corresponding values of t.

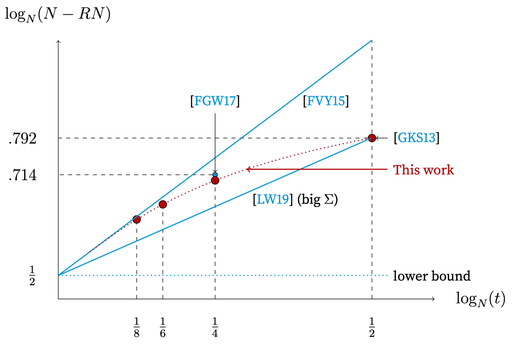


Fig 5. Constructions with the t-DRGP. Points that are lower and further to the right correspond to better codes.

Future Work

The immediate next steps would be to improve our current construction.

- Can we attain low redundancy for other values of t apart from $\sqrt[2m]{N}$?
- What happens when we consider polynomials with more than 2 variables?
- What happens when we try more sophisticated repair functions?

References

[FGW17] S Luna Frank-Fischer, Venkatesan Guruswami, and Mary Wootters. Locality via partially lifted codes. In *Approximation, Randomization, and Combinato- rial Optimization. Algorithms and Techniques (APPROX/RANDOM 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

[FVY15] A. Fazeli, A. Vardy, and E. Yaakobi. Codes for distributed pir with low storage overhead. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 2852–2856, June 2015.

[GKS13] Alan Guo, Swastik Kopparty, and Madhu Sudan. New affine-invariant codes from lifting. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 529–540, 2013.

[LW19] Ray Li and Mary Wootters. Lifted multiplicity codes and the disjoint repair group property. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*. Schloss Dagstuhl- Leibniz-Zentrum fuer Informatik, 2019.

Gravitational Contrastive Learning: Inducing Structure During Instance Discrimination

Jasmine Bayrooti, Mike Wu, Alex Tamkin, Noah Goodman

Computation and Cognition Lab, Computer Science, Stanford University

Motivation

Training supervised image classification models takes lots of labelled data

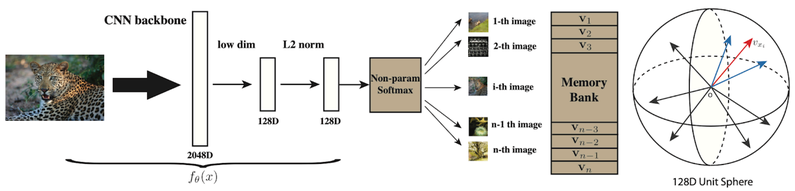
Contrastive learning is a way to learn meaningful image representations **without** explicit labels

An instance-level discrimination (IR) approach pushes representations closer to "positive" views and away from "negative" views

Goal: Improve IR by inducing **topological structure** in the representation space

Related Works

Regular IR distributes representations uniformly over a sphere



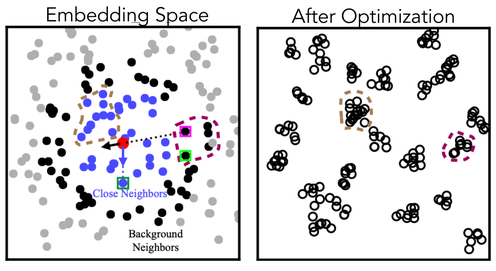
Maybe we want the structure of the representations to reflect a **different** distribution

Local Aggregation (LA) aggregates similar images:

- Positives - close points with same cluster label
- Negatives – other close points

LA induces only fixed structures

Idea: develop a more **flexible** approach



Views

Figure 1a: Sample input image

Figure 1b: A sample IR positive view for the input

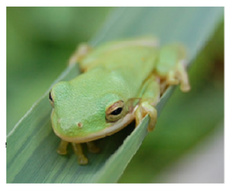


Figure 1c: A sample IR negative view for the input

Methods

Objectives

IR Objective	LA Objective:
$P(i\|v) = \frac{e^{v_i^T v}}{\sum_{j=1}^{N} e^{v_j^T v}}$	$-\log \frac{\sum_{i \in C_i \cap B_i} P(i\|v_i)}{\sum_{i \in B_i} P(i\|v_i)}$

Our Procedure

IR Objective with a **gravity-inspired** kmeans update

- Positives – centroids for close cluster labels
- Negatives – all other points

$$v_i = v + \mu * \sum_{j=1}^{K} \frac{\sum_{k=1}^{N} d(c_k, v) + \epsilon}{d(c_j, v) + \epsilon} * (c_j - v)$$

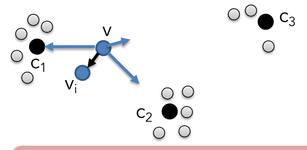


Figure 2: This illustrates the gravity update applied to representation v considering the 3 nearest centroids. This exercises a greater updating force on points that are close to a few centroids and a smaller net update on points between centroids or in less dense regions.

Steps

Training:

1. Train encoder to optimize representations for each image
2. Use IR objective with gravity updates after each epoch

Evaluation:

3. Train a linear classifier over image representations

Results

Model Classification Accuracies

	CIFAR 10	CIFAR 100[1]	ImageNet
IR	0.823	0.656	0.446
LA	0.841	0.652	**0.480**
Gravity	**0.850**	**0.666**	0.452

Table 1: Comparison of classification accuracies from linear classifiers trained over optimized representations. For CIFAR 100, super label classification accuracy was used. For CIFAR 10 and ImageNet, image label classification was used.

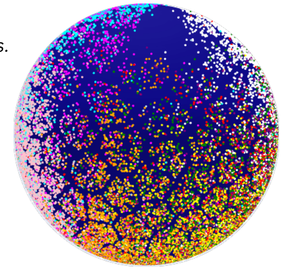


Figure 3: Shows the fully optimized embedding space trained in three dimensions with 300 clusters using gravity updates. Each point corresponds to a unique image in the dataset and the color corresponds to the image's class.

Discussion

Gravity approach **outperforms** IR and LA

- Except on ImageNet likely due to lack of hyperparameter tuning
- Shows that a clustering distribution is helpful!
- We use less noisy positive views than LA
- Found that LA's background set is not critical

Pushing points closer to the kmeans centroid may be equivalent to **Bayesian inference** on the generative model over embeddings

Future Work

Look into choices for μ

Continue experiments on ImageNet

Explore projecting to other manifolds

References

1. Zhirong Wu et al. "Unsupervised feature learning via non-parametric instance discrimination". https://arxiv.org/abs/1805.01978
2. Zhuang, C. et al. "Local aggregation for unsupervised learning of visual embeddings". *https://arxiv.org/abs/1903.12355*
3. *Yonglong Tian et al. "What makes good views for contrastive learning". https://arxiv.org/abs/2005.10243*

Coding Storytime: Bridging Early Computational Thinking and Storytelling Using A Voice-Guided App

STANFORD HUMAN-COMPUTER INTERACTION GROUP

Jimmy Le: jimmyle@stanford.edu

Nadin Tamer: nadint@stanford.edu

LELAND STANFORD JUNIOR UNIVERSITY · DIE LUFT DER FREIHEIT WEHT · 1891

Motivation

Problem

Computational thinking toys are often inaccessible due to cost and literacy requirements.

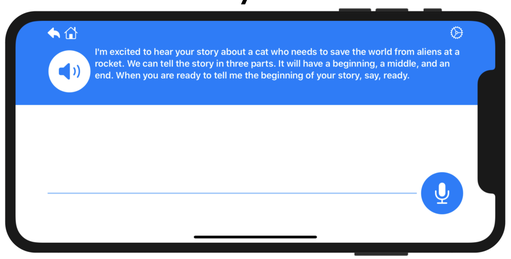
Our Solution

A voice-guided iOS app leveraging storytelling to teach computational thinking to children ages 5–8.

System Features

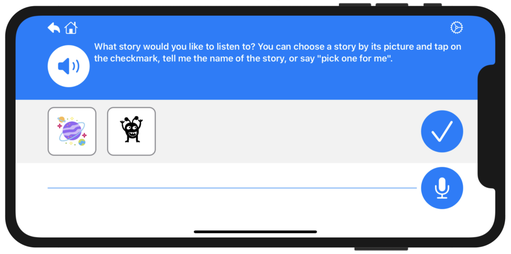
Storytelling

Create a Story



Create your own story with a story plan

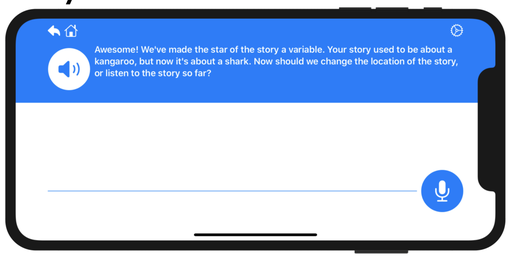
Listen to a Story



Listen to old stories

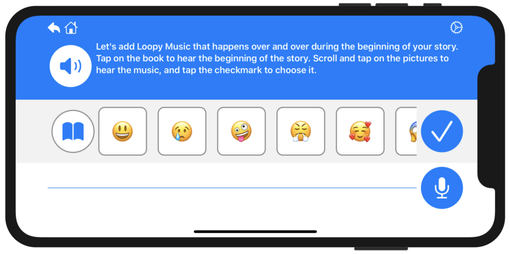
Computing Concepts

Very Variable



Change the star & location of your story

Loopy Music

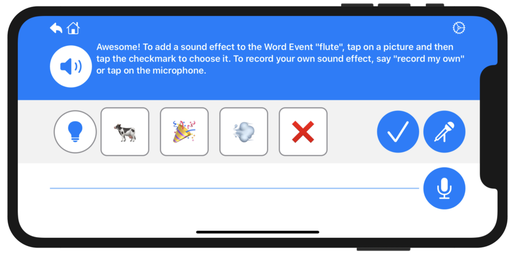


Add background music to your story

Scrambled Sequences

Put your story back in the right order

Word Events



Add sound effects to words in your story

Short-Term Evaluation Metrics

Computing Concepts

- Game usability
- Concept recall and recognition in new contexts

Storytelling

- Game usability
- Narrative quality
- Story structure recall and use in new contexts

Attitudinal

- Engagement
- Perception as computing

Recognition in new contexts task

User Studies

Study Design

- N=22 (11 females, ages 5–8: M=6.85, SD=1.16)
- 4 sessions (pre-test, gameplay, gameplay, and post-test)

Preliminary Results

Children perform above-chance in recognizing computing concepts in new contexts.

Next Steps

- Analysis of short-term evaluation data
- Visual design
- Long-term deployment

Acknowledgements

A special thank you to our mentor Griffin Dietz, Thomas Hsieh, Jenny Han, Dr. Elizabeth Murnane, Prof. James Landay, and the CURIS coordinators.