Asymptotics of Fingerprinting and Group Testing
(Extended Abstract)

Thijs Laarhoven

Department of Mathematics and Computer Science
Eindhoven University of Technology
5600 MB Eindhoven, The Netherlands
mail@thijs.com

Abstract

This extended abstract provides an overview of the results presented in [8, 9] regarding the large-coalition asymptotics of collusion-resistant fingerprinting and the asymptotics of group testing for large numbers of defectives.

1 Fingerprinting

To protect copyrighted content against unauthorized redistribution, distributors commonly embed fingerprints in the content, uniquely linking copies to individual users. If the distributor then finds an illegal copy of the content online, he can determine which user was responsible. To combat this solution, a group of c colluders might cooperate and perform a collusion attack. By comparing their versions of the content, they will detect differences in their copies, which must be part of the fingerprint. They can then try to create a mixed pirate copy, where the resulting fingerprint matches the fingerprints of different colluders in different segments of the content, making it hard for the distributor to find the responsible users. The goal of collusion-resistant fingerprinting is to assign fingerprints of length ℓ to n users in such a way that, even if c pirates collude, the pirate copy can still be traced back to the responsible users with high probability.

In 1998, Boneh and Shaw [3] were the first to show that one can construct such a scheme with a code length polynomial in c and logarithmic in n. In particular, their construction had a code length of the order $\ell \propto c^4 \ln n$, and they showed that any scheme requires a code of length at least $\ell \propto c \ln n$. In 2003, Tardos [12] proved a stronger lower bound of the order $\ell \propto c^2 \ln n$, and described an improved scheme with $\ell = 100c^2 \ln n$, showing that up to leading constants, his construction is optimal.

Later work on fingerprinting focused on finding the optimal leading constant and finding constructions with shorter code lengths. Huang and Moulin [6] derived explicit expressions for the channel capacities of the related max-min games, and proved that the optimal asymptotic code length is $\ell \sim 2c^2 \ln n$. By providing a construction matching this lower bound, Oosterwijk et al. [10] later showed that this bound can be achieved using a simple decoder.

The fingerprinting game can naturally be generalized to adaptive settings, where the colluders broadcast their pirate copy in real-time. With the construction of [7] one can efficiently convert arbitrary non-adaptive schemes to adaptive schemes, which may be able to compete with the celebrated scheme of Fiat and Tassa [5].

2 Group testing

A different area of research that has received considerable attention over the last few decades is group testing. Suppose a large population of size n contains $c \ll n$ infected people. To identify these people, it is possible to perform blood tests: testing a subset of the population will lead to a positive test result if this subset contains at least one infected person, and a negative result if all tested people are clean. Since the time to run a single test may be long, the subsets to test need to be chosen in advance, after which all tests are performed simultaneously. Then, when the test results come back, the subset of infected people needs to be identified. The goal of group testing is to identify the infected people using as few group tests ℓ as possible.

Already in the 1980s it was known that the optimal code length of such schemes scales as $\ell \sim c \log_2 n$ [11]. Later work focused on slight variations of the classical model such as noisy group testing, where a positive result may not always correspond to the presence of a defective item [2, 4]. For these variants, exact asymptotics on the capacities and constructions achieving these capacities were yet unknown.

3 Contributions

Building upon previous work of Huang and Moulin, in [8] we derived exact asymptotics (for large c) for the capacities of various fingerprinting and group testing models. In almost all cases this led to a lower bound on the code length of the order $\ell \propto c \ln n$. For several models it further turned out that there is a strict gap between the capacities of simple and joint decoders. In particular, for the traditional group testing model this gap is a factor $\ln 2$; an optimal joint decoder asymptotically requires a factor $\ln 2 \approx 0.69$ fewer tests than any simple decoder.

With these results in mind, [9] discusses explicit simple and joint decoders for various fingerprinting and group testing models with provable code lengths matching the capacities derived in [8]. The considered decoders are based on log-likelihood ratios, which are well-known from hypothesis testing literature to be optimally discriminative for deciding between two hypotheses H_0 and H_1 (e.g., distinguishing between guilty and innocent users, or infected and clean persons). Combining these results with [1], this also led to the discovery of a new fingerprinting decoder for arbitrary attacks. This decoder seems to have all the desired properties one can hope for, so maybe, just maybe, this finally ends the search for the optimal decoder in fingerprinting.

References

[1] E. Abbe and L. Zheng, "Linear universal decoding for compound channels," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 5999–6013, 2010.

[2] G. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1880–1901, 2012.

[3] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Transactions on Information Theory*, vol. 44, no. 5, pp. 1897–1905, 1998.

[4] M. Cheraghchi, A. Hormati, A. Karbasi, and M. Vetterli, "Group testing with probabilistic tests: theory, design and application,"
IEEE Transactions on Information Theory, vol. 57, no. 10, pp. 7057–7067, 2011.

[5] A. Fiat and T. Tassa, "Dynamic traitor tracing,"
Journal of Cryptology, vol. 14, no. 3, pp. 354–371, 2001.

[6] Y.-W. Huang and P. Moulin, "On the saddle-point solution and the large-coalition asymptotics of fingerprinting games,"
IEEE Transactions on Information For. and Sec., vol. 7, no. 1, pp. 160–175, 2012.

[7] T. Laarhoven, J. Doumen, P. Roelse, B. Škorić, and B. de Weger, "Dynamic Tardos traitor tracing schemes,"
IEEE Transactions on Information Theory, vol. 59, no. 7, pp. 4230–4242, 2013.

[8] T. Laarhoven, "Asymptotics of fingerprinting and group testing: tight bounds from channel capacities,"
submitted to IEEE Transactions on Information Theory, 2014.

[9] T. Laarhoven, "Asymptotics of fingerprinting and group testing: capacity-achieving log-likelihood decoders,"
submitted to IEEE Transactions on Information Theory, 2014.

[10] J.-J. Oosterwijk, B. Škorić, and J. Doumen, "A capacity-achieving simple decoder for bias-based traitor tracing schemes,"
submitted to IEEE Transactions on Information Theory, 2013.

[11] A. Sebő, "On two random search problems,"
Journal of Statistical Planning and Inference, vol. 11, pp. 23–31, 1985.

[12] G. Tardos, "Optimal probabilistic fingerprint codes,"
ACM Symposium on Theory of Computing, pp. 116–125, 2003.