AICR · OPEN PROBLEM

Public-Key Encryption from Constant-Noise Planted k-XOR

A linear number of equations, constant noise rate

Status. Statement: AI-written from Boaz Barak, *The Complexity of Public-Key Cryptography*, Cryptology ePrint Archive 2017, not yet checked by a human. The question is open in the source and is posed there as a question rather than as a conjecture: no scheme in the survey is based on constant noise at linear density, the survey's own reference point — the Applebaum–Barak–Wigderson linear-function variant — building public-key encryption from planted noisy linear equations only when the noise rate is sub-constant, and Alekhnovich's scheme needing noise about $1/\sqrt{n}$.

Category. *Public-Key Cryptography & Assumptions.*

Conjecture (informal). *Every public-key encryption scheme we actually rely on comes from one of two places: number theory in Abelian groups, or geometry of codes and lattices. There is a third kind of hardness that private-key cryptography uses freely and public-key cryptography cannot: take a system of random XOR equations, each involving a constant number of the variables, plant an assignment that satisfies most of them, and ask an algorithm to find an assignment that does as well. With a linear number of equations and a constant fraction of them deliberately falsified, this looks like a plain, structureless average-case problem, and it immediately gives candidate one-way functions. The conjecture is that it also gives public-key encryption: there is some constant arity, some constant equation-to-variable ratio, and some constant noise rate ε for which one can build an encryption scheme whose semantic security follows from nothing but the hardness of that search problem. What makes this hard is that every non-algebraic scheme we have needs the noise to be sub-constant — for the ABW linear variant, noise rate times arity $O(1/\log n)$; for Alekhnovich, about $1/\sqrt{n}$, which is where the paper locates the structural weakness. That small noise is precisely the feature that lets lattice-based schemes be broken given an oracle for a problem in* **NP** ∩ **coNP** *— not a problem believed to be easy, but a form of computational structure that generic one-way function candidates do not have. So the conjecture asks for a scheme built on noise so large that this structural weakness cannot arise, and there is evidence that noise this sparse may be an inherent limitation of one general family of schemes built from noisy codewords — the family containing Alekhnovich's and Regev's schemes.*

1 The setting

Public-key encryption, unlike private-key cryptography, rests on a very short list of assumptions. The survey this problem is taken from divides the well-studied schemes into two families: an "algebraic" family built on Abelian groups (integer factoring, discrete logarithm, elliptic curves) and a "geometric" family built on codes and lattices, whose cleanest formalisation is Learning With Errors [Reg09]. Private-key cryptography is in a completely different position: candidate one-way functions fall out of almost any planted

average-case problem, including random constraint satisfaction problems with a planted assignment [BKS13] and random local functions [Gol11, App15]. The gap between these two pictures is the gap between Impagliazzo's Minicrypt and Cryptomania [Imp95], and the survey's subject is what, if anything, forces it.

Applebaum, Barak and Wigderson [ABW10] tried, in the survey's own words, to explore the question of whether public-key encryption can be based on the conjectured average-case difficulty of combinatorial problems. Their scheme uses a random bipartite graph with n left vertices, m right vertices and constant right degree d, and rests on two assumptions: that the induced local map is a pseudorandom generator, and that a planted set of $O(\log n)$ right vertices whose neighbourhood is smaller than itself cannot be detected. They also gave a variant in which the local predicate is replaced by a linear function with added noise at rate δ, and that variant carries a hard constraint: decryption requires $\delta k = O(1/\log n)$, so δ is forced to be sub-constant. Driving δ lower buys a larger k — low enough and the expansion assumption can be dropped altogether — but at that point the pseudorandomness assumption itself acquires structure, since it admits a non-constructive short certificate of the kind Feige, Kim and Ofek [FKO06] built for dense random 3CNF formulas.

Alekhnovich's scheme [Ale11] sits in the same trap from the other side: it is a mod-2 analogue of the lattice schemes and uses linear equations with noise rate about $1/\sqrt{n}$. The survey's diagnosis is that the small noise, and not the linearity, is what does the damage — after all, the knapsack and approximate k-XOR problems are **NP**-hard as worst-case problems. A noise rate below $1/\sqrt{n}$ is the analogue of a lattice approximation factor above $\sqrt{n}$, and it is exactly at those approximation factors that lattice-based schemes can be broken given an oracle for a problem in $\mathbf{NP} \cap \mathbf{coNP}$ [AR05], structure that generic one-way function candidates do not have. Ben-Sasson et al. [BBD+16] give evidence that noise this sparse is an inherent limitation of a general family of schemes built from noisy codewords, a family containing both Alekhnovich's scheme and Regev's.

That is what makes the constant-noise regime the one worth asking about. At $m = O(n)$ equations of constant arity k, with a planted assignment satisfying a $1 - \varepsilon$ fraction for a constant ε,

the assumption has no known structure of this kind: it is a plain planted constraint satisfaction problem, of exactly the sort that yields one-way functions with no difficulty at all. The density $O(n)$ is simply what the survey's question fixes. What is missing is a public-key scheme that can run on it: no scheme in the survey is based on constant noise — ABW's linear-function variant needs $\delta k = O(1/\log n)$, and Alekhnovich's uses noise about $1/\sqrt{n}$.

Progress to date. The survey reports two partial results, both negative in flavour. First, the ABW linear-function variant works only with sub-constant noise, since efficient decryption forces the noise rate times the arity to be $O(1/\log n)$; and pushing the noise low enough to drop the expansion assumption makes the remaining pseudorandomness assumption structured, because it then admits a non-constructive short certificate in the style of Feige–Kim–Ofek [FKO06]. Second, Ben-Sasson, Ben-Tov, Damgard, Ishai and Ron-Zewi [BBD+16] define a general family of schemes from noisy codewords, containing Alekhnovich's scheme and Regev's, and show that under a conjecture from additive combinatorics all such schemes need noise patterns satisfying a generalised square-root-sparsity condition; the survey reads this as evidence that sparse noise may be inherent for schemes of this type.

2 Notation and parameters

Write $[n] = \{1, \dots, n\}$, and let $\oplus$ denote addition in $\mathbb{F}_2$. Throughout, $k \geq 3$ is an integer constant, $c > 0$ and $\varepsilon \in (0, 1/2)$ are real constants, all three independent of n, and $m := \lceil cn \rceil$.

A *k-XOR instance* on n variables with m equations is a pair

$$\Phi \;=\; \big((S_1, \dots, S_m), (b_1, \dots, b_m)\big),$$

where $S_i \subseteq [n]$ with $|S_i| = k$, and $b_i \in \{0,1\}$, for every $i \in [m]$. For $y \in \{0,1\}^n$ the *value* of y on Φ is the fraction of equations it satisfies,

$$\mathrm{val}_\Phi(y) \;:=\; \frac{1}{m}\left|\left\{\, i \in [m] \;:\; \bigoplus_{j \in S_i} y_j = b_i \,\right\}\right|.$$

Efficient means probabilistic polynomial time in n, and $\mathrm{negl}(n)$ denotes a function that is eventually smaller than n^{-a} for every constant $a > 0$.

- n: number of variables in the equation system, and the security parameter.
- k: arity of each XOR equation; a constant, at least 3.
- c: equation density; the system has $m = \lceil cn \rceil$ equations, so a linear number of them; a constant.
- ε: noise rate; the planted assignment falsifies an ε fraction of the equations; a constant in $(0, 1/2)$.

3 The conjecture

Definition 1 (Planted k-XOR at linear density). Let $\mathcal{P}_n^{k,c,\varepsilon}$ be the distribution on k-XOR instances obtained as follows: sample $x \leftarrow \{0,1\}^n$ uniformly; sample $S_1, \ldots, S_m$ independently and uniformly from the k-element subsets of $[n]$; sample $T \subseteq [m]$ uniformly among the subsets of size $\lfloor \varepsilon m \rfloor$; and set

$$b_i = \bigoplus_{j \in S_i} x_j \qquad (i \notin T),$$

$$b_i = 1 \oplus \bigoplus_{j \in S_i} x_j \qquad (i \in T).$$

The output is $\Phi = ((S_i)_i, (b_i)_i)$; the planted assignment x is not output, and satisfies exactly a $1 - \lfloor \varepsilon m \rfloor / m \geq 1 - \varepsilon$ fraction of the equations.
The planted k-XOR problem with parameters (k, c, ε) is *hard* if for every efficient algorithm A,

$$\Pr_{\Phi \leftarrow \mathcal{P}_n^{k,c,\varepsilon}} \left[\mathrm{val}_\Phi\big(A(\Phi)\big) \geq 1 - \varepsilon \right] = \mathrm{negl}(n).$$

Definition 2 (Public-key encryption of one bit). A public-key encryption scheme is a triple $\Pi = (\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ of efficient algorithms, where $\mathrm{Gen}(1^n)$ outputs a pair (pk, sk), $\mathrm{Enc}(pk, b)$ takes $b \in \{0,1\}$ and outputs a ciphertext, and

$\mathrm{Dec}(sk,\cdot)$ outputs a bit. It is *correct* if

$$\Pr\big[\,\mathrm{Dec}\big(sk,\mathrm{Enc}(pk,b)\big)=b\,\big]\;\geq\;1-\mathrm{negl}(n)$$

for both $b\in\{0,1\}$, the probability being over $(pk,sk)\leftarrow\mathrm{Gen}(1^n)$ and the coins of Enc and Dec. It is *CPA-secure* if for every efficient A,

$$p_b\;:=\;\Pr\big[A\big(pk,\mathrm{Enc}(pk,b)\big)=1\big]\qquad(b\in\{0,1\}),$$
$$\big|p_0-p_1\big|\;=\;\mathrm{negl}(n).$$

Conjecture 1 (public-key encryption from constant-noise planted k-XOR). *There exist an integer constant $k\geq 3$, real constants $c>0$ and $\varepsilon\in(0,1/2)$, and a correct public-key encryption scheme $\Pi=(\mathrm{Gen},\mathrm{Enc},\mathrm{Dec})$, such that the following holds:*

> *if the planted k-XOR problem with parameters (k,c,ε) is hard, then Π is CPA-secure.*

Here the three algorithms of Π must be efficient and its correctness must hold unconditionally, so that the hardness of planted k-XOR at linear density $m=\lceil cn\rceil$ with constant noise rate ε is the only assumption carrying the security of Π.

Remark 1 (readings fixed by this write-up). Three choices above are not forced by the survey's one-sentence question. First, "base a public-key encryption scheme on the difficulty of" is rendered as the bare implication of Conjecture 1; demanding instead a black-box reduction with polynomial loss would give a stronger statement. Second, "solving" is read as the search problem of Definition 1, that is, finding an assignment of the promised value; substituting the distinguishing version of the assumption would give a weaker statement, and the survey does not say which it means. Third, the noise model of Definition 1 falsifies exactly a $\lfloor\varepsilon m\rfloor$-sized set of equations, matching the survey's "satisfying $1-\varepsilon$ of them"; the independent Bernoulli variant differs only by $o(1)$ in the value and should be considered equivalent here. Likewise $O(n)$ is read as $\lceil cn\rceil$ for a constant c we are free to choose, rather than for a c fixed in advance.

4 Bibliography

[ABW10] B. Applebaum, B. Barak, A. Wigderson. *Public-key cryptography from different assumptions.* Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, pages 171–180, 2010.

[Ale11] M. Alekhnovich. *More on average case vs approximation complexity.* Computational Complexity, 20(4):755–786, 2011; preliminary version in FOCS '03.

[App15] B. Applebaum. *The cryptographic hardness of random local functions — survey.* IACR Cryptology ePrint Archive, 2015:165, 2015.

[AR05] D. Aharonov, O. Regev. *Lattice problems in NP ∩ coNP.* J. ACM, 52(5):749–765, 2005; preliminary version in FOCS '04.

[BBD+16] E. Ben-Sasson, I. Ben-Tov, I. Damgard, Y. Ishai, N. Ron-Zewi. *On public key encryption from noisy codewords.* Public-Key Cryptography — PKC 2016, Part II, pages 417–446, 2016.

[BKS13] B. Barak, G. Kindler, D. Steurer. *On the optimality of semidefinite relaxations for average-case and generalized constraint satisfaction.* Innovations in Theoretical Computer Science, ITCS '13, pages 197–214, 2013.

[FKO06] U. Feige, J. H. Kim, E. Ofek. *Witnesses for non-satisfiability of dense random 3CNF formulas.* 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), pages 497–508, 2006.

[Gol11] O. Goldreich. *Candidate one-way functions based on expander graphs.* Studies in Complexity and Cryptography, pages 76–87, 2011; original version ECCC TR00-090, 2000.

[Imp95] R. Impagliazzo. *A personal view of average-case complexity.* Proceedings of the Tenth Annual Structure in Complexity Theory Conference, pages 134–147, 1995.

[Reg09] O. Regev. *On lattices, learning with errors, random linear codes, and cryptography.* J. ACM, 56(6), 2009; preliminary version in STOC 2005.