Optimal Amplification of Noisy Leakages*

Stefan Dziembowski[1], Sebastian Faust[2], and Maciej Skórski[1]

[1] University of Warsaw
[2] Ruhr University Bochum

Abstract. During the last 15 years there have been intensive research efforts in constructing cryptographic algorithms resilient to the side-channel leakage. The most fundamental part of every such construction are the *leakage-resilient encoding schemes*. Usually the cryptographic secrets encoded by them are assumed to belong to some finite group $(\mathbb{G}, +)$. The most common encoding scheme is the n-out-of-n additive secret sharing: a secret X is encoded as $(X_1, \ldots, X_n)$ such that $X_1 + \cdots X_n = X$. Intuitively, if an adversary receives only small partial independent information about each X_i then his information about X should be even smaller, and should decrease (i.e. the noise should *amplify*) when n grows. However, of course, the concrete parameters (the amount of leakage that can be tolerated, and the number of shares needed to achieve a given level of security) depend on the exact model that is used.

One of the most prominent models used in this area is the so-called *noisy-leakage model* (Chari et al, CRYPTO'99, and Prouff and Rivain, EUROCRYPT'13), which is believed to correspond well to the real-life engineering experience, where the information that the adversary receives is always "noisy". In the Prouff and Rivain model the amount of information that the noise provides is measured using a parameter δ that is equal to 0 when the noise is "full", and equal to 1 when there is no noise. It is natural to ask how small δ needs to be to achieve the amplification of noise (in the additive encoding scheme described above). Until now it was known that such amplification can be achieved for $\delta < 1/16$. In this paper we show that:

- in the prime order groups $\mathbb{G}$ it suffices that $\delta < 1 - 1/|\mathbb{G}|$,
- in general it suffices that $\delta < 1/2$.

We also prove that these bounds are optimal. We then analyze the number n of shares needed to achieve security ϵ of the encoded value X (where ϵ is also defined in terms of "noisy information" that the adversary obtains about X). We give close lower and upper bounds on this value (that differ only in factor polylogarithmic in $|\mathbb{G}|$). We achieve our results using techniques from the additive combinatorics, the harmonic analysis, and the convex optimization.

* The first and the last author were partly supported by the WELCOME/2010-4/2 grant founded within the framework of the EU Innovative Economy Operational Programme.

1 Introduction

Leakage-resilient cryptography [15,19,11,1,24,2,20,6,3,5,12,9,14] aims at constructing cryptographic schemes that are secure against side-channel leakages of secret information Leakage-resilient encoding schemes are important building blocks for constructing such algorithms. They allow to encode a secret message X with a randomized encoding function $\mathsf{Enc}(X) = (X_1, \ldots, X_n)$ such that leakage from the codeword does not help the adversary to recover the secret message X. The simplest encoding function $\mathsf{Enc}(.)$ uses an additive secret sharing scheme, where the shares X_i are chosen uniformly at random from a group $\mathbb{G}$ such that $X := X_1 + \ldots + X_n$. Of course, the leakage from $\mathsf{Enc}(X)$ cannot be arbitrary as otherwise no security is possible. A common assumption that has been studied both in theoretical and practical works [4,24,12,22,8] is to assume that the leakage from the encoding is a "noisy" function. It has been shown in the work of Chari et al. [4] that the encoding scheme described above amplifies security when the leakage is "sufficiently" noisy. While several recent works improve the quantitative bounds on the amount of noise needed [22,10], the current best bounds require the leakage to be very noisy – in particular far away from the level of noise that is typically available in physical measurements [8]. The main contribution of our work is to give an optimal characterization of the noisy leakage model. We develop optimal bounds for the amount of noise that is needed to amplify security, and give matching upper bounds by showing that above this threshold amplification is impossible. Our results are particular important for the security analysis of masking schemes, which we further describe below.

Leakage resilient encodings for masking schemes. Leakage resilient encodings are prominently used to build masking schemes [4,15]. Masking schemes are widely used in practice to protect cryptographic implementations against side-channel leakage – in particular against leakage from the power consumption [16]. The basic idea of a masking scheme is simple: instead of computing directly on the sensitive information (e.g., the secret key), a cryptographic algorithm protected with the masking countermeasure computes on encoded values thereby concealing sensitive information, which makes it harder to extract relevant information from the leakage. The simplest and most widely used masking scheme is the Boolean masking scheme. The Boolean masking scheme introduced in the important work of Ishai et al. [15] uses the simple encoding function from above when $\mathbb{G} = \mathsf{GF}(2)$, but can easily be extended to work in larger groups. For instance, to protect an implementation of the AES algorithm we typically use $\mathbb{G} = \mathsf{GF}(2^8)$ as the AES algorithm can be implemented in a particular efficient way using operations in the Galois field. Another example is a protected implementation of discrete-log based crypto schemes that work in prime-order fields.

The noisy leakage model. While there are several variants of the "noisy leakage model" [20,12,22] most works that consider the security of masking schemes use the model of Chari et al. [4] and its generalization by Prouff and Rivain [22]. Informally, a noisy leakage function $f : \mathbb{G} \to \mathcal{Y}$ is called δ-noisy if the statistical

distance between the uniform distribution X over $\mathbb{G}$ and the conditional distribution of X *given* $f(X)$ is bounded by a parameter $\delta \in [0, 1 - 1/|\mathbb{G}|]$. Here, $\mathcal{Y}$ is the domain of the leakage, which in general can be an infinite set. To better understand the Prouff-Rivain noise model, let us consider the two extreme cases. First, when δ is close to 0 then f is assumed to be very noisy, and hence the noisy leakage reveals little about the underlying sensitive information. On the other extreme, when δ is close to $1-1/|\mathbb{G}|$ then there is almost no noise in the leakage (i.e., the function f is "almost" deterministic). The Prouff-Rivain noise model is believed to model well real-world physical side-channel leakages. Moreover, as shown in [7,10] it is also a robust noise measure as it is equivalent to various other ways of describing the noise present in a leakage function.

Masking schemes in the noisy leakage model. A masking function is said to be secure in the noisy leakage model if for any $X, Y \in \mathbb{G}$, we have that noisy leakage from an encoding of $(X_1, \ldots X_n) \leftarrow \mathsf{Enc}(X)$ is statistically close to noisy leakage from $(Y_1, \ldots, Y_n) \leftarrow \mathsf{Enc}(Y)$. As discussed above recent works have significantly improved the Prouff-Rivain δ-bias for which security of the encoding function can be shown. While initial works [22,7] require that $\delta = O(1/|\mathbb{G}|)$, the recent work of Dziembowski et al. [10] show that noise amplifies security already when $\delta < 1/16$ (and, hence in particular independent of the size of the underlying group $\mathbb{G}$). In this work, we can show that for prime-order groups masking amplifies the noise for $\delta \leq 1 - 1/p$. Notice that in case when p is super-polynomial in the security parameter (as in discrete-log based cryptosystems), then we achieve security under the optimal assumption of $1 - \mathit{negl}(n)$. For general groups (in particular, groups with small factorization) we show that amplification is possible when $\delta < 1/2$. We also show that both our bounds are optimal as for values above the threshold amplification is not possible. We provide further details on our contributions and techniques in the next two sections.

1.1 Our Contributions

We analyze the amplification of noisy leakage for the simple additive encoding function Enc. The quality of the amplification is measured by the ϵ-security of the encoding. We say that an encoding is ϵ-secure if the statistical distance between the δ-noisy leakage of $\mathsf{Enc}(x)$ and $\mathsf{Enc}(x')$ for two elements $x, x' \in \mathbb{G}$ is upper bounded by ϵ. We characterize how many shares n we need in order to amplify the noise available in the leakage from the shares. To this end we derive a value $\delta_{\max}$ which is the maximal value for δ until which we can still amplify the noise for sufficiently large n. Of course, as we show the number of share n needs to increase the closer we set δ to $\delta_{\max}$. One interesting observation arising from our analysis is that the value of $\delta_{\max}$ depends on the structure of the underlying group $\mathbb{G}$. We summarize our results regarding the upper bound until which amplification is possible in the following informal theorem.

Theorem (Noise amplification, informal version of Theorem 1). *Let $\mathbb{G}$ be a group (either prime order, or arbitrary) and let the adversary obtain δ-noisy*

leakage from the shares of the encoding. Define the maximal noise parameter $\delta_{\max}$ *as*

$$\delta_{\max} = \begin{cases} 1 - \frac{1}{p}, & \textit{when } \mathbb{G} \textit{ is of prime order} \\ \frac{1}{2}, & \textit{when } \mathbb{G} \textit{ is arbitrary.} \end{cases} \tag{1}$$

Then for any $\delta < \delta_{\max}$ *and any* $\epsilon > 0$ *we have that the encoding* Enc *is* ϵ*-secure for*

$$n = \text{poly}\left(\log|\mathbb{G}|, \log(\epsilon^{-1}), (\delta_{\max} - \delta)^{-1}\right) \tag{2}$$

(where poly *is some polynomial).*

Let us explain the interplay of the parameters used in the above theorem. The number of shares grows polynomially with two parameters: (a) the logarithm of ϵ^{-1}, i.e., the target security we aim for, and (b) the gap $\theta = \delta_{\max} - \delta$ between the maximal possible noise value and the actual chosen noise level δ. The dependency on (a) is as expected: if we aim for better security meaning a smaller value of ϵ we require a larger security parameter n. The reason for the dependency stated in (b) is more technical, and essentially comes from a bias convergence in the harmonic analysis (when $\mathbb{G}$ has prime order), or in the XOR Lemma (when $\mathbb{G}$ has non-prime order), see Section 3 for details.[3]

Dependence on the group order. As already noted, our Theorem 1 distinguishes between two cases. In the first case the group $\mathbb{G}$ is of prime order. Interestingly, in this case it turns out that arbitrarily small noise (i.e. δ close to 1) can be amplified. Informally, this is thanks to the fact that prime-order groups have no non-trivial sub-groups. On the other hand, if a group has a non-prime order, i.e., it contains non-trivial subgroups, then we require a higher noise (more precisely: $\delta < 1/2$). On a practical level this means that in some sense the prime order groups "provide a better leakage resilience" than the general groups. This may be useful for discrete-log based cryptosystems that typically work in such groups.

Lower bounds on the necessary noisy level via homomorphic attacks. We show that the maximal noise parameter $\delta_{\max}$, as defined in Equation (1), is optimal in the following sense. We show in Proposition 1 that δ has to be less than $1 - \frac{|\mathbb{H}|}{|\mathbb{G}|}$, where $\mathbb{H}$ is the largest proper subgroup of $\mathbb{G}$. An intuitive explanation for why the group structure is relevant here, is the existence of "homomorphic attacks", when given $X_1, \ldots, X_n$ being shares of X and their evaluations $\phi(X_1), \ldots, \phi(X_n)$ under a homomorphism $\phi : \mathbb{G} \to \mathbb{H}$, we can compute $\phi(X) = \phi(X_1) + \ldots + \phi(X_n)$.

The implications of Proposition 1 are two-fold. Firstly, as long as the general groups are considered (and hence no assumptions on the order can be made), then we prove that one needs to assume that the δ is less than $1/2$. This is because, since the largest proper subgroup $\mathbb{H}$ of $\mathbb{G}$ can be of size $|\mathbb{G}|/2$, thus

[3] We show also that the dependency on θ is necessary as otherwise we can provide attacks against the encoding.

$1 - |\mathbb{H}|/|\mathbb{G}|$ can be as small as $1/2$. Secondly, is $\mathbb{G}$ has prime order then $|\mathbb{H}| = 1$ and therefore $1 - |\mathbb{H}|/|\mathbb{G}| = 1 - 1/p$. Hence in this case the lower bound matches the upper bound from Theorem 1.

It is natural to ask if our results can be more fine-grained, and fully characterize the noise requirements in terms of $|\mathbb{H}|/|\mathbb{G}|$. We conjecture that in fact the upper bound $1 - |\mathbb{H}|/|\mathbb{G}|$ can also be always achieved (not only in the cases when $|\mathbb{H}| = |\mathbb{G}|/2$ and $\mathbb{H} = |1|$). We leave proving it as an open problem. We summarize the upper bounds and the relation to the number of shares in Table 1 below.

Group	Necessary noise	The gap	Necessary number of shares
$\lvert\mathbb{G}\rvert$ is even	$\delta < \delta_{\max} = \frac{1}{2}$	$\theta = \frac{1}{2} - \delta$	$n = \text{poly}\left(\log(\epsilon^{-1}), \theta^{-1}\right)$
$\mathbb{G} = \mathbb{Z}_p$	$\delta < \delta_{\max} = 1 - \frac{1}{p}$	$\theta = 1 - \frac{1}{p} - \delta$	

Table 1: Matching bounds for the necessary noise amount and the necessary number of shares

Applications of our techniques outside of masking schemes. We show that our techniques also have applications outside of the domain of leakage resilient cryptography. In particular, using our results we can extend the following product theorem, due to Maurer et al. [17].

Lemma 1 (Product Theorem [17]). *Let $\mathbb{G}$ be a group, $d(\cdot)$ denote the distance from uniform and X_1, X_2 be arbitrary independent random variables on $\mathbb{G}$. Then we have*

$$d(X_1 + X_2) \leqslant 2 \cdot d(X_1) \cdot d(X_2)$$

We give a different proof and calculate optimal constants for any group.

Theorem (Product Theorem with optimal constants, informal version of Theorem 2). *Let $\mathbb{G}$ be a group, $d(\cdot)$ denote the distance from uniform and X_1, X_2 be arbitrary independent random variables on $\mathbb{G}$. Then we have*

$$d(X_1 + X_2) \leqslant c(\mathbb{G}) \cdot d(X_1) \cdot d(X_2)$$

where $c(\mathbb{G}) \leq 2$ is a constant depending on the structure of the underlying group $\mathbb{G}$.

For the exact value of $c(\mathbb{G})$ we refer the reader to Appendix A.6.

Comparing our results to previous works. Our work improves the previous state-of-the-art in the following aspects:

(a) For general groups the best upper bound was given in [10], where it was shown that $\delta < 1/16$. We improve this bound to $\delta < 1/2$, which is optimal for groups with even order. Moreover, for groups of prime-order p we give a novel bound of $1 - \frac{1}{p}$. Notice that for primes that are super-polynomial in the security parameter we achieve security for δ arbitrary close to 1. We notice that in contrast to earlier works our proof techniques also have the advantage of being more modular.

(b) We provide matching lower bounds showing that the noise threshold as well as the growth rate of the number of shares are optimal, which was not known before.

(c) Our proof techniques may be of independent interest and have not been used previously for analyzing the security of noisy leakages. In particular, our analysis uses techniques from convex optimization, additive combinatorics and harmonic analysis.

In Table 2 below we compare our results with related works.

	Proof techniques	Sufficient noise	Minimal Noise	Sufficient n	Minimal n
[22]	direct information theoretic analysis	$O\left(\lvert\mathbb{G}\rvert^{-1}\right)$	not discussed	$\mathrm{poly}\left(\log(\lvert\mathbb{G}\rvert),\ \log(\epsilon^{-1}), (\delta_{\max} - \delta)^{-1}\right)$	not discussed
[7]	reduction to random probing	$O\left(\lvert\mathbb{G}\rvert^{-1}\right)$			
[10]	reduction to random walks, amplifying indistinguishability	$\frac{1}{16}$			
here	optimal reduction to random walks, harmonic analysis, additive combinatorics, convex optimization	$\frac{1}{2}$ for any $\mathbb{G}$ $1 - \frac{1}{p}$ for $\mathbb{Z}_p$	$\frac{1}{2}$ if $\lvert\mathbb{G}\rvert$ even $1 - \frac{1}{p}$ for $\mathbb{Z}_p$	$\mathrm{poly}\left(\log(\lvert\mathbb{G}\rvert),\ \log(\epsilon^{-1}), (\delta_{\max} - \delta)^{-1}\right)$	$\mathrm{poly}\left(\log(\epsilon^{-1}), (\delta_{\max} - \delta)^{-1}\right)$

Table 2: The initial amount of noise needed for the security of Enc.

Comparison with the binomial noise model and the XOR Lemma. We stress that the noise model of Prouff and Rivain [22] we consider is significantly more general than the binomial noise model considered by Faust et al. in [12] (even in the binary case) and considers many other types of noisy functions. In particular, our noisy function f maps elements from the group $\mathbb{G}$ to a possibly infinite set Y.

If we restrict in Theorem 1 the noise model to the special case of binomial noise (i.e., the leakage function f is the binomial noise function), then we obtain comparable parameters with e.g., in [12] (cf. Lemma 4 in [12]).[4]

[4] The main restriction when comparing our result with a direct application of the XOR Lemma, is that we require the probability p of flipping the shares to be $> 1/4$ (in contrast to [12] where $p > 0$ is sufficient). The later restriction stems from the fact that leakages in the binomial noise model with parameter p are transformed in our noise model to a requirement of $delta = 1 - 2p$ noisy-function. We emphasize that for the general type of "noisy leakage" we consider, our bounds are optimal as shown by our lower bounds.

1.2 A high-level proof outline

We now give an overview of our proof techniques (the details appear in Section 3). We prove our result in five steps illustrated on Figure 1. We first show that it is enough to consider uniform secrets X. The proof appears in Section 3.1.1. The cryptographic interpretation of this claim is that it suffices to consider only *random-plaintext attacks*, instead of *chosen-plaintext attacks*.

Then, in Section 3.1.2, we consider the distance between a uniform secret X given δ-noisy leakages $f_1(X_1), \ldots, f_n(X_n)$ from its encoding and a uniformly and independently chosen X'. We show that bounding this distance is equivalent to bounding the distance of a random sum of the form $Z = \sum_{i=1}^{n} Z_i$ with independent summands Z_i, conditioned on noisy information $\{f_i(Z_i)\}_{i=1}^{n}$, from the uniform distribution U. Here we use the fact that X is uniform (guaranteed by Step 1). The fact that leakage functions are δ-noisy guarantees that Z_i is δ-close to uniform given $f_i(Z_i)$, for $i = 1, \ldots, n$. Intuitively it is clear that if independent random variables on a group are close to the uniform distribution, then their sum is even closer to uniform (cf. XOR-lemmas [13], see also Appendix A.8). The main issue here is that our summands are conditional distributions, which means that Z_i is close to U only in average conditioned on concrete values of the leakage $f_i(Z_i) = y_i$.

Next in Section 3.1.3 we get rid of the conditional part $\{f_1(Z_1), \ldots, f_n(Z_n)\}$. This step is accomplished by considering concrete leakage values $f_1(Z_i) = y_i$ for $i = 1, \ldots, n$ and noticing that for most of them the distance from uniform is not much bigger than δ, which we conclude by the Chernoff Bound. This step results into an error term, which is exponential in $n\theta^2$ where $\theta = \delta_{\max} - \delta$ is the gap-parameter defined above. Informally speaking, the gap θ is what allows to further reduce the problem to study only the distance of sums of the form $Z = Z_1 + \ldots + Z_n$ from the uniform distribution.

Later, in Section 3.1.4, we characterize the distributions Z_i for which the distance between $Z = Z_1 + \ldots + Z_n$ and U is *maximal*. It turns out that they have a *simple shape*: they are a combination of a mass-point and a distribution uniform over a set of size $(1-\delta)|\mathbb{G}| - 1$. A description of how these "worst-case" distributions look like, enables us to come up with concrete estimates for the statistical distance in the next step.

Finally, in Section 3.1.5 we prove concrete facts about the convergence speed of cumulative sums of random variables that are sufficiently close to the uniform distribution. Depending on the technique used and the assumption on the group structure imposed, we obtain different bounds in Theorem 1.

We note that for the case when we make no assumptions on the structure of $\mathbb{G}$, steps from Sections 3.1.4 and 3.1.5 (but not from Sections 3.1.1—3.1.3) could be replaced by a product theorem due to Maurer et al. [18]. Our technique allow us to extend this theorem, taking the group structure into account. In the last step we split the proof depending on the technique and the assumption about $\mathbb{G}$. The quantitative comparison of different bounds we get is given in Table 3. Note that the number n of shares is asymptotically larger when the additive combinatorics is used (second row of Table 3) than when the harmonic analysis

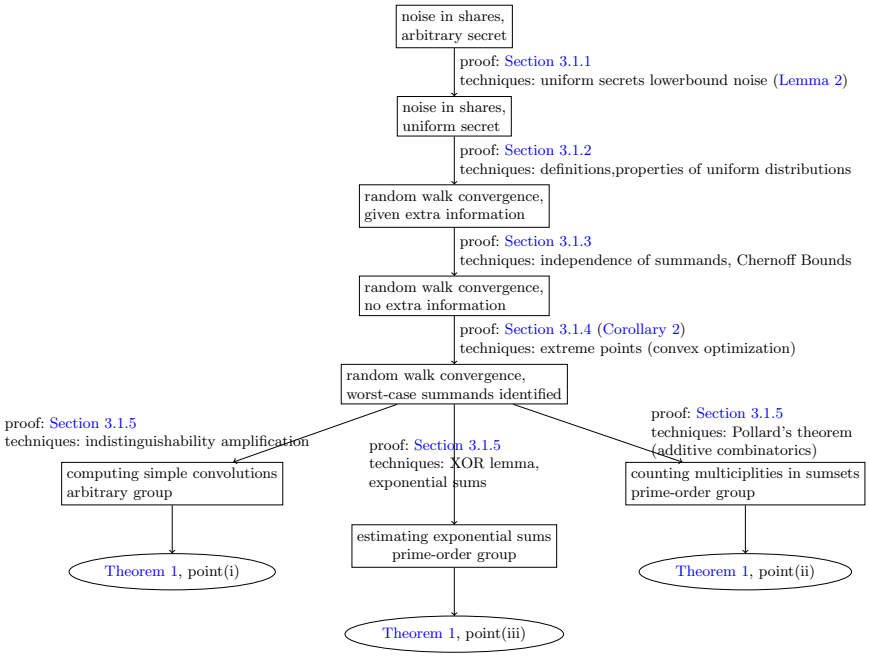

Fig. 1: An overview of the proof of Theorem 1 and applied techniques.

is used (the third row). Nevertheless we think it is instructive the present both results since the proof techniques are different, and both can be of independent interest.

Domain	Proof technique	Number n of shares	Assumption
$\mathbb{G}$ arbitrary	amplifying indistinguishability	$O\left(\log(\lvert\mathbb{G}\rvert/\epsilon)/\theta^2\right)$	$\delta \leqslant \frac{1}{2} - \theta$
$\mathbb{G} = \mathbb{Z}_p$	additive combinatorics	$O\left(\log(\lvert\mathbb{G}\rvert/\epsilon) \cdot 2^{12\log(1/\theta)/\theta^2}\right)$	$\delta \leqslant 1 - \frac{1}{p} - \theta$
$\mathbb{G} = \mathbb{Z}_p$	harmonic analysis	$O\left(\theta^{-4}\log(\lvert\mathbb{G}\rvert/\epsilon)\log(1/\theta)\right)$	$\delta \leqslant 1 - \frac{1}{p} - \theta$

Table 3: The amount of shares needed to mask the secret state below the advantage ϵ, depending on the assumed group structure and proof technique. In the last column, θ is an arbitrarily small positive number.

1.3 Our techniques

In this section we summarize our main techniques used in the proof of Theorem 1.

Convex analysis. We use the convex analysis in Section 3.1.4 to deal with the problem of determining how fast the sum of independent components $Z_1 + \ldots +$

Z_n on a group $\mathbb{G}$ converges towards the uniform distribution. Our assumption on the noise guarantees that Z_i are at most δ-close to the uniform distribution with some parameter δ. Since we can think of distributions over $\mathbb{G}$ as vectors in $\mathbb{R}^{|\mathbb{G}|}$, we observe that the mapping

$$(Z_1, \ldots, Z_n) \longrightarrow \mathsf{SD}\left(Z_1 + \ldots + Z_n; U\right)$$

is a convex mapping, with respect to the distribution of any Z_i when the remaining components are fixed. Since the restrictions on the distance of Z_i's from uniform are also convex, we conclude that the maximum is achieved for one of the *extreme points* from the set of feasible Z_i. As a consequence we observe that they have a surprisingly *simple shape* (see Lemma 5). That simple structure will play an important role in the very last step of the proof. Also, it allows us to derive a general product theorem for groups, with an explicit expression with tight constants.

Additive combinatorics. In Section 3.1.5, when studying the convergence of random sums over a general group $\mathbb{G}$, we can find a proper subgroup $A \triangleleft \mathbb{G}$ which is by definition an *additive set*, that is

$$A + A = A.$$

Such a set may constitute a *trap* for our random walk $Z_1 + \ldots + Z_n$. If $Z_i \in A$ for all i, then $Z_1 + \ldots + Z_n \in A$. When $\mathbb{G} = \mathbb{Z}_p$ such a trap does not exist, so intuitively the sum takes all the elements with similar probability when n is large (because even one non-zero point generates the group when added multiple times). This is where we use some basic facts from additive combinatorics. The first result of this sort is the Cauchy-Davenport theorem which states that the sumset $A + B$, where A, B are arbitrary subsets of $\mathbb{Z}_p$ must be substantially bigger than A and B alone. More precisely

$$|A + B| \geqslant |A| + |B| - 1.$$

This result does not help us much because it gives no estimate on *repetitions* in the sumset $A + B$, that is how many of the expressions $a + b$ where $a \in A, b \in B$ hit the same place. To get more information about the distribution of repetitions in the sumset we use a more refined result due to Pollard. Combining it with the explicit form of Z_i (developed in the previous steps) we obtain a non-trivial upper bound on $\mathsf{SD}(Z_1 + Z_2; U)$ in terms of $\mathsf{SD}(Z_1; U), \mathsf{SD}(Z_2; U)$, which is then extended to the sum of n elements.

Harmonic analysis. Also, in Section 3.1.5, having reduced our problem to the convergence of a random walk with independent increments, we can use techniques from Fourier analysis. Recall that a character is a complex-valued function ϕ which is additive on $\mathbb{G}$, that is $\phi(x + y) = \phi(x) \cdot \phi(y)$. The expectations of characters on independent sums are especially easy to evaluate, because

$$\mathbb{E}\left[\phi\left(Z_1 + \ldots + Z_n\right)\right] = \prod_{i=1}^{n} \mathbb{E}\left[\phi(Z_i)\right].$$

For $\mathbb{Z}_p$ expectations are easier to calculate, because any character ϕ is of the simple form $\phi(x) = \exp(2\pi k\mathrm{i}/p)$ for some k. Since we know the shape of the worst Z_i, we can obtain a concrete estimate for a nontrivial character ϕ, namely,

$$|\mathbb{E}\left[\phi\left(Z_i\right)\right]| < c \ll 1.$$

Intuitively, this comes from the fact that Z_i "contains" a large uniform component which doesn't allow the mass to concentrate at one point. Using a bound on geometric sums over unity roots, we conclude that $\mathbb{E}\left[\phi\left(Z_1 + \ldots + Z_n\right)\right] < c$. Finally we apply the XOR lemma which states that characters are "representative" distinguishers: if two distributions have close expectations under every character, they indeed are statistically close. In our case we apply this claim to $Z = Z_1 + \ldots + Z_n$ and U and the result follows since we have shown that $\mathbb{E}[\phi(Z)]$ is small and trivially we have $\mathbb{E}[\phi(U)] = 0$ (for non-trivial ϕ).

2 Preliminaries

If X and Y are random variables over the same set $\mathcal{X}$ then the statistical distance between X and Y is denoted as $\mathsf{SD}(X;Y)$, and defined as $\mathsf{SD}(X;Y) = \frac{1}{2}\sum_{x\in\mathcal{X}} |\Pr[X = x] - \Pr[Y = x]|$. If Z is a random variable then by $\mathsf{SD}(X;Y|Z)$ we mean $\mathsf{SD}((X,Z);(Y,Z))$, i.e., the statistical distance of the two joint distributions. If two distributions X and Y are equivalent, then we write $X \stackrel{d}{=} Y$. Below we formally define the encoding and decoding of a secret $X \in \mathbb{G}$.

Definition 1 (Encoding and Decoding). *Let $(\mathbb{G}, +)$ be a fixed group and $n > 1$ be a fixed natural number. For any $X \in \mathbb{G}$, we define the encoding function* Enc *by*

$$\mathsf{Enc}^n_{\mathbb{G}}(X) = (X_1, \ldots, X_{n-1}, X - (X_1 + \ldots + X_{n-1}))$$

where $X_1, \ldots, X_{n-1}$ are independent and uniform over $\mathbb{G}$, and the decoding function Dec *by*

$$\mathsf{Dec}^n_{\mathbb{G}}(X_1, \ldots, X_n) = X_1 + \ldots + X_n.$$

We will typically omit n and $\mathbb{G}$ and simply write $(\mathsf{Enc}, \mathsf{Dec})$ *when clear from the context.*

Noisy leakages. The noise in the observed version Y of a real distribution X, denoted by $\beta(X|Y)$, is measured by comparing how close is the product distribution $\mathbf{P}_X \cdot \mathbf{P}_Y$ to the joint distribution $\mathbf{P}_{(X,Y)}$. More formally, we have the following definition, which comes from [22] (see also [7]), where it was argued that it models physical noise in a realistic way.

Definition 2 (Noisy observations and noisy functions). *A random variable $Y \in \mathcal{X}$ is called a* δ-noisy observation of X *if*

$$\beta(X|Y) \stackrel{def}{=} \mathsf{SD}(X';X|Y) \leqslant \delta$$

where X' is an independent copy of X. A function f is called δ-noisy if $f(U)$ is a δ-noisy version of U, where U is uniform over U.

Notice that [22] defined the noisy function as the $\sum_y \Pr[Y = y] \cdot \mathsf{SD}(X'; (X|Y = y))$. This is equivalent to the above when X and X' are independent and Y is the leakage from X. Note that this definition may seem a bit counterintuitive as more noise means a bias value closer to 0, while a value closer to 1 means that less noise is present in the leakage observations.

3 Our Main Result

We are now ready to present our main result that was already informally described in Section 1.1.

Theorem 1. *Let X be a random variable on a group $\mathbb{G}$ and let $Enc(X) = (X_1, \ldots, X_n)$ be its encoding. Suppose that f_i for $i = 1, \ldots, n$ are all δ-noisy functions, i.e. $\beta(X_i|f_i(X_i)) \leqslant \delta$. Then we have the following bounds*

(i) For arbitrary $\mathbb{G}$, if $\delta \leqslant \frac{1}{2} - \theta$, then $\beta(X|f_i(X_1), \ldots, f_i(X_n)) \leqslant \epsilon$ provided that

$$n > 8\theta^{-2} \log(5|\mathbb{G}|\epsilon^{-1})$$

(ii) If $\mathbb{G} = \mathbb{Z}_p$ and $\delta \leqslant 1 - \frac{1}{p} - \theta$ then $\beta(X|f_i(X_1), \ldots, f_i(X_n)) \leqslant \epsilon$, provided that

$$n > \log(3|\mathbb{G}|\epsilon^{-1}) \cdot 2^{12\log(1/\theta)/12\theta}$$

(iii) If $\mathbb{G} = \mathbb{Z}_p$ and $\delta \leqslant 1 - \frac{1}{p} - \theta$ then $\beta(X|f_i(X_1), \ldots, f_i(X_n)) \leqslant \epsilon$, provided that

$$n > 2\theta^{-4} \log(|\mathbb{G}|\epsilon^{-1})$$

3.1 Proof of Theorem 1

The proof of Theorem 1 was already outlined in Section 1.2. In the next sections we describe it in more detail. The final proof appears in Section 3.1.5.

3.1.1 Reducing to uniform secrets.

Below we show that it sufficient to consider only uniform secrets X.

Lemma 2. *Suppose that X is uniform over $\mathbb{G}$ with the encoding $\mathsf{Enc}(X) = (X_1, \ldots, X_n)$. Let X' be an arbitrary distribution over $\mathbb{G}$ with the encoding $\mathsf{Enc}(X') = (X'_1, \ldots, X'_n)$. Then for arbitrary functions $f_1, \ldots, f_n$ we have*

$$\beta(X'|f_1(X'_1), \ldots, f_n(X'_n)) \leqslant 3|\mathbb{G}| \cdot \beta(X|f_1(X_1), \ldots, f_n(X_n)) \tag{3}$$

The proof appears in Appendix A.3. Note that we lose a factor of $|\mathbb{G}|$ in this transformation. However this does not actually affect the bound we want to prove, because we show that the main part $\beta(X|f_1(X_1), \ldots, f_n(X_n))$ converges to 0 exponentially fast with n.

3.1.2 Reducing to random walks conditioned on noisy information.

We now show that the noise in a uniform secret X given δ-noisy leakages $f_1(X_1), \ldots, f_n(X_n)$ is equal to the distance of a random sum of the form $Z = \sum_{i=1}^{n} Z_i$ with independent summands Z_i, conditioned on noisy information $\{f_i(Z_i)\}_{i=1}^{n}$, from the uniform distribution U.

Lemma 3. *For X uniform on a set $\mathcal{X}$ and any functions f_i the following equality holds*

$$\beta(X|(f_i(X_i))_{i=1}^n) = \mathsf{SD}\left(\sum_{i=1}^{n} Z_i;\ U \,\middle|\, (f_i(Z_i))_{i=1}^n\right) \tag{4}$$

where U and Z_i for $i = 1, \ldots, n$ are uniform and independent over $\mathcal{X}$.

The proof appears in Appendix A.4. Justifying the title, we note that we can think of the sum $\sum_{i=1}^{n} Z_i$ as a random walk which starts at 0, with increments Z_i.

3.1.3 Reducing to unconditional random walks.

The following lemma is an easy consequence of the Chernoff Bound.

Lemma 4. *Let $(Z_i, Y_i)_i$ for $i = 1, \ldots, n$ be independent random variables such that $\Delta(Z_i; U|Y_i) \leqslant \delta$. Then, for any $\gamma > 0$, $\delta' = \delta + 2\gamma$ and $n' = \gamma n$*

$$\mathsf{SD}\left(\sum_{i=1}^{n} Z_i; U \,\middle|\, (Y_i)_i\right) \leqslant \max_{(Z'_i)_i:\ \mathsf{SD}(Z'_i;U)\leqslant\delta'} \mathsf{SD}\left(\sum_{i=1}^{n'} Z'_i; U\right) + \mathrm{e}^{-2n\gamma^2} \tag{5}$$

where the maximum is taken over all independent random variables Z'_i.

The proof appears in Appendix A.5. The immediate corollary below shows that we can get rid of the conditinal part in the right-hand side of Equation (4) in Lemma 3.

Corollary 1. *For $n' = \frac{\theta}{4} \cdot n$ we have*

$$\beta(X|(f_i(X_i))_{i=1}^n) \leqslant \max_{(Z'_i)_i:\ \mathsf{SD}(Z'_i;U)\leqslant\delta+\frac{\theta}{2}} \mathsf{SD}\left(\sum_{i=1}^{n'} Z'_i; U\right) + \mathrm{e}^{-\frac{1}{8}n\theta^2}$$

where the maximum is taken over all independent random variables Z'_i.

Note that by combining Step 1, Step 2 and Step 3 with Lemma 1 we can already conclude part (i) of Theorem 1 (see Section 3.1.5).

3.1.4 Worst-case summands.

We prove the following geometrical fact:

Lemma 5 (Shape of extreme points of distributions close to uniform). *Let $\mathcal{X}$ be a finite set and U be uniform over $\mathcal{X}$. Any distribution $X \in \mathcal{X}$ such that $\mathsf{SD}(X;U) \leqslant \delta$ can be written as a convex combination of "extreme" distributions X' of the following form: with probability $p = \delta + |\mathcal{X}|^{-1}$ the distribution X' takes a value a and with probability $q = 1 - p$ the distribution X' is uniform over a set A, where $a \notin A$, of size $|A| = (1-\delta)|\mathcal{X}| - 1$[5]. Equivalently, each of these distributions X' is of the following form:*

$$\mu_{X'} = \mu_U + \delta\mu_b - \delta\mu_B \tag{6}$$

for some B such that $|B| = \delta|\mathbb{G}|$ and $b \notin B$.

Note that we always have $(1-\delta)|\mathcal{X}| - 1 \geqslant 0$, as the range of the noise parameter is $0 \leqslant \delta \leqslant 1 - \frac{1}{\mathcal{X}}$, when we consider secrets over $\mathcal{X}$.

Corollary 2. *Let Z_i, for $i = 1, \ldots, n$ be independent random variables such that $\mathsf{SD}(Z_i;U) \leqslant \delta$ for every i. Then $\mathsf{SD}\left(\sum_i Z_i;U\right)$ is maximized for Z_i as in Lemma 5*

Proof (of Corollary 2). Note that the distribution of $\sum_i Z_i$ is a convolution of individual distributions $\mathbf{P}_{Z_i}$, and therefore it is multilinear in $\mathbf{P}_{Z_i}$. It follows that $\mathsf{SD}\left(\sum_i Z_i;U\right) = \frac{1}{2}\left\|\mathbf{P}_{\sum_i Z_i} - \mathbf{P}_U\right\|_1$ is convex in $\mathbf{P}_{Z_i}$ and the claim follows by the extreme point principle.

Using Lemma 5 we derive the following generalization of Lemma 1

Theorem 2. *Let Z_1, Z_2 be independent random variables on a group $\mathbb{G}$. Then we have*

$$\mathsf{SD}(Z_1 + Z_2;U) \leqslant c_{\max}(\mathbb{G}) \cdot \mathsf{SD}(Z_1;U) \cdot \mathsf{SD}(Z_2;U) \tag{7}$$

Where the constant is given by

$$c_{\max}(\mathbb{G}) = \frac{1}{2} \max_{A,B:|A|=\delta_1|\mathbb{G}|,|B|=\delta_2|\mathbb{G}|} \|\mu_B + \mu_A - \mu_A * \mu_B - \mu_0\|_{\ell_1(G)} \tag{8}$$

where μ_0 is the point mass at 0, $\delta_i = \mathsf{SD}(Z_i;U)$, and μ_A, μ_B are uniform over the sets A and B. Moreover, the sharp constant is achieved for the following random variables: Z_i is constant with probability $\delta_i + \frac{1}{|\mathbb{G}|}$ and with probability $1 - \delta_i - \frac{1}{|\mathbb{G}|}$ is uniform on some set of size $(1-\delta_i)|\mathbb{G}| - 1$.

Lemma 5 is a corollary from Theorem 2, whose proof appears in Appendix A.6. Note that Lemma 1 follows from Theorem 2, since $c_{\max}(\mathbb{G}) \leqslant 2$ trivially, since

$$\|\mu_B + \mu_A - \mu_A * \mu_B - \mu_0\|_1 \leqslant \|\mu_B\|_1 + \|\mu_B\|_1 + \|\mu_A * \mu_B\|_1 + \|\mu_0\|_1 = 4$$

by the triangle inequality and the fact that the total mass of a probability measure is 1.

[5] If $\delta|\mathcal{X}|$ is not an integer, then instead of a uniform distribution we consider the distribution flat over a set A such that $|A| = \lceil (1-\delta)|\mathcal{X}| - 1 \rceil$, which assigns the mass of $\frac{1}{(1-\delta)|\mathcal{X}|-1}$ to all but one points, and the mass of $\frac{\lceil (1-\delta)|\mathcal{X}|-1 \rceil - ((1-\delta)|\mathcal{X}|-1)}{(1-\delta)|\mathcal{X}|-1}$ to the ramining point.

3.1.5 Concrete bounds. In view of Corollary 1 it remains to give an upper bound on the distance between sums of independent random variables which are not too far from uniform and the uniform distribution, i.e., on:

$$\mathsf{SD}\left(\sum_{i=1}^{n} Z_i; U\right).$$

To this end, we split our analysis depending on the structure of $\mathbb{G}$ and chosen technique.

Case: $\mathbb{G}$ is arbitrary. From Corollary 1 and Lemma 1 applied $(n-1)$ times it follows that

$$\beta(X|(f_i(X_i))_{i=1}^n) \leqslant \frac{1}{2}\left(2\delta+\theta\right)^{\frac{\theta}{4}n} + \mathrm{e}^{-\frac{1}{8}n\theta^2}.$$

From the assumption $\delta < \frac{1}{2} - \theta$ and the elementary inequality $1 - u \leqslant \mathrm{e}^{-u}$ we obtain $(2\delta+\theta)^{\frac{\theta}{4}n} \leqslant \mathrm{e}^{-\frac{1}{4}\theta^2 n}$, which gives us

$$\beta(X|(f_i(X_i))_{i=1}^n) < \frac{3}{2}\cdot \mathrm{e}^{-\frac{1}{8}n\theta^2}$$

for uniform X. Taking into account Step 1, we finally obtain

$$\beta(X|(f_i(X_i))_{i=1}^n) < 5|\mathbb{G}|\cdot \mathrm{e}^{-\frac{1}{8}n\theta^2}.$$

for any X, which is equivalent to part (i) of Theorem 1.

Case: $\mathbb{G} = \mathbb{Z}_p$, for p prime (by additive combinatorics). When $\mathbb{G} = \mathbb{Z}_p$, we improve Lemma 1 in the following way, using Corollary 2 and some tools from additive combinatorics (see Appendix A.7 for a proof).

Lemma 6. *Let Z_1, Z_2 be independent random variables on $\mathbb{Z}_p$ such that $\mathsf{SD}(Z_i; U) \leqslant \delta_i$. Then*

$$\mathsf{SD}(Z_1 + Z_2, U_G) \leqslant h(\delta_1, \delta_2) \tag{9}$$

where

$$h(\delta_1,\delta_2) = \begin{cases} 2\delta_1\delta_2, & \phi(\delta_1,\delta_2) \leqslant 0 \\ 2\delta_1\delta_2 - \frac{1}{4}\phi(\delta_1,\delta_2)^2 + \frac{1}{4p^2}, & \phi(\delta_1,\delta_2) > 0 \end{cases} \tag{10}$$

and $\phi(\delta_1,\delta_2) \stackrel{def}{=} \delta_1 + \delta_2 + \min(\max(\delta_1,\delta_2), 1 - |\delta_1 - \delta_2|) - 1$.

We will use only the following consequence of Lemma 6.

Corollary 3. *Let Z_1, Z_2 be independent random variables on $\mathbb{Z}_p$ such that $\mathsf{SD}(Z_i; U) \leqslant \delta_i \leqslant \delta$. Suppose that $\delta > \frac{1}{3}$. Then we have*

$$\mathsf{SD}(Z_1 + Z_2; U_G) \leqslant 2\delta^2 - \frac{(3\delta-1)^2}{4} + \frac{1}{4p^2} \tag{11}$$

Using recursively Corollary 3 and applying Corollary 1 we obtain the following bound for uniform X

$$\beta\left(X|(f_i(X_i))_{i=1}^n\right) \leqslant 2^{-n/\left(2^{16/\theta}\cdot 12\theta\right)} + \mathrm{e}^{-\frac{1}{8}n\theta^2}$$

which, by Step 1, implies part (ii) of Theorem 1. For a detailed derivation, see Lemma 8 in Appendix A.7.

Case $\mathbb{G} = \mathbb{Z}_p$, for a prime number p (by harmonic analysis). We start by obtaining the following auxiliary estimate on trigonometric sums, valid for any $A \subset \mathbb{Z}_p$ such that $|A| = \theta p$ and $k \in \{1, 2, \ldots, p-1\}$:

$$\left|\sum_{x\in A} \exp\left(\frac{2k\pi i x}{p}\right)\right| \leqslant \frac{\sin \pi\theta}{p \sin\frac{\pi}{p}}$$

The proof uses a geometrical argument and some trigonometric identities and is given inside the proof of Lemma 10 in Appendix A.8. Based on Corollary 2 and the XOR lemma (see Lemma 9 in Appendix A.8), we prove that

$$\beta\left(X|(f_i(X_i))_{i=1}^n\right) \leqslant 3|\mathbb{G}|^{\frac{3}{2}} \cdot \mathrm{e}^{-\frac{1}{8}\theta^3} + \mathrm{e}^{-\frac{1}{8}n\theta^2}$$

for uniform X, which by Step 1 implies part (iii) of Theorem 1; the details are given in Lemma 10 in Appendix A.8.

4 Lower bounds

Proposition 1 (The noise threshold (1) is optimal). *For any group $\mathbb{G}$, there exist a δ-noisy function f where*

$$\delta = 1 - \frac{|\mathbb{H}|}{|\mathbb{G}|},$$

$\mathbb{H}$ being the biggest proper subgroup of $\mathbb{G}$, such that for every n we have

$$\beta\left(X|f(X_1), \ldots, f(X_n)\right) \geqslant \delta$$

.

The proof appears in Appendix A.1

Proposition 2 (The growth rate in (2) is optimal). *For $\mathbb{G} = \mathbb{Z}_2$, X being uniform on $\mathbb{G}$, and any $\theta < \frac{1}{2}$ there exists a $\left(\frac{1}{2} - \theta\right)$-noisy leakage function f such that for every n satisfying*

$$n < \log\left((2\epsilon)^{-1}\right) / \log\left((1-2\theta)^{-1}\right) \tag{12}$$

we have

$$\beta\left(f(X)|f(X_1), \ldots, f(X_n)\right) \geqslant \epsilon.$$

In turn, for $\mathbb{G} = \mathbb{Z}_p$ where p is prime, X being uniform on $\mathbb{G}$, and any $\theta < 1 - \frac{1}{p}$ there exists a $\left(1 - \frac{1}{p} - \theta\right)$-noisy leakage function f such that for every n satisfying

$$n < \log(2\epsilon^{-1}) / \log(1-\theta) \tag{13}$$

we have

$$\beta\left(f(X) | f(X_1), \ldots, f(X_n)\right) \geqslant \epsilon.$$

The proof appears in Appendix A.2. Note that for $\theta < \frac{1}{4}$ we have $\log((1-2\theta)^{-1}) < \frac{1}{4}\theta^{-1}$, and then Equation (12) can be replaced by

$$n < 4\theta^{-1} \log\left((2\epsilon)^{-1}\right)$$

Similarly, for $\theta < \frac{1}{2}$ we have $\log((1-\theta)^{-1}) < \frac{1}{2}\theta^{-1}$, and then Equation (13) can be replaced by

$$n < 2\log(2\epsilon^{-1}) \cdot \theta^{-1} \tag{14}$$

References

1. Adi Akavia, Shafi Goldwasser, and Vinod Vaikuntanathan. Simultaneous hardcore bits and cryptography against memory attacks. In Omer Reingold, editor, *TCC 2009*, volume 5444 of *LNCS*, pages 474–495. Springer, Berlin, Germany, March 15–17, 2009.
2. Joël Alwen, Yevgeniy Dodis, and Daniel Wichs. Leakage-resilient public-key cryptography in the bounded-retrieval model. In Shai Halevi, editor, *CRYPTO 2009*, volume 5677 of *LNCS*, pages 36–54, Santa Barbara, CA, USA, August 16–20, 2009. Springer, Berlin, Germany.
3. Zvika Brakerski, Yael Tauman Kalai, Jonathan Katz, and Vinod Vaikuntanathan. Overcoming the hole in the bucket: Public-key cryptography resilient to continual memory leakage. In *51st FOCS*, pages 501–510, Las Vegas, Nevada, USA, October 23–26, 2010. IEEE Computer Society Press.
4. Suresh Chari, Charanjit S. Jutla, Josyula R. Rao, and Pankaj Rohatgi. Towards sound approaches to counteract power-analysis attacks. In Michael J. Wiener, editor, *CRYPTO'99*, volume 1666 of *LNCS*, pages 398–412, Santa Barbara, CA, USA, August 15–19, 1999. Springer, Berlin, Germany.
5. Francesco Davì, Stefan Dziembowski, and Daniele Venturi. Leakage-resilient storage. In Juan A. Garay and Roberto De Prisco, editors, *SCN 10*, volume 6280 of *LNCS*, pages 121–137, Amalfi, Italy, September 13–15, 2010. Springer, Berlin, Germany.
6. Yevgeniy Dodis, Kristiyan Haralambiev, Adriana López-Alt, and Daniel Wichs. Cryptography against continuous memory attacks. In *51st FOCS*, pages 511–520, Las Vegas, Nevada, USA, October 23–26, 2010. IEEE Computer Society Press.
7. Alexandre Duc, Stefan Dziembowski, and Sebastian Faust. Unifying leakage models: From probing attacks to noisy leakage. In Phong Q. Nguyen and Elisabeth Oswald, editors, *EUROCRYPT 2014*, volume 8441 of *LNCS*, pages 423–440, Copenhagen, Denmark, May 11–15, 2014. Springer, Berlin, Germany.

8. Alexandre Duc, Sebastian Faust, and François-Xavier Standaert. Making masking security proofs concrete - or how to evaluate the security of any leaking device. In Elisabeth Oswald and Marc Fischlin, editors, *EUROCRYPT 2015, Part I*, volume 9056 of *LNCS*, pages 401–429, Sofia, Bulgaria, April 26–30, 2015. Springer, Berlin, Germany.
9. Stefan Dziembowski and Sebastian Faust. Leakage-resilient circuits without computational assumptions. In Ronald Cramer, editor, *TCC 2012*, volume 7194 of *LNCS*, pages 230–247, Taormina, Sicily, Italy, March 19–21, 2012. Springer, Berlin, Germany.
10. Stefan Dziembowski, Sebastian Faust, and Maciej Skorski. Noisy leakage revisited. In Elisabeth Oswald and Marc Fischlin, editors, *EUROCRYPT 2015, Part II*, volume 9057 of *LNCS*, pages 159–188, Sofia, Bulgaria, April 26–30, 2015. Springer, Berlin, Germany.
11. Stefan Dziembowski and Krzysztof Pietrzak. Leakage-resilient cryptography. In *49th FOCS*, pages 293–302, Philadelphia, Pennsylvania, USA, October 25–28, 2008. IEEE Computer Society Press.
12. Sebastian Faust, Tal Rabin, Leonid Reyzin, Eran Tromer, and Vinod Vaikuntanathan. Protecting circuits from leakage: the computationally-bounded and noisy cases. In Henri Gilbert, editor, *EUROCRYPT 2010*, volume 6110 of *LNCS*, pages 135–156, French Riviera, May 30 – June 3, 2010. Springer, Berlin, Germany.
13. Oded Goldreich. Three xor-lemmas — an exposition. In Oded Goldreich, editor, *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, volume 6650 of *Lecture Notes in Computer Science*, pages 248–272. Springer Berlin Heidelberg, 2011.
14. Shafi Goldwasser and Guy N. Rothblum. How to compute in the presence of leakage. In *53rd FOCS*, pages 31–40, New Brunswick, NJ, USA, October 20–23, 2012. IEEE Computer Society Press.
15. Yuval Ishai, Amit Sahai, and David Wagner. Private circuits: Securing hardware against probing attacks. In Dan Boneh, editor, *CRYPTO 2003*, volume 2729 of *LNCS*, pages 463–481, Santa Barbara, CA, USA, August 17–21, 2003. Springer, Berlin, Germany.
16. Paul C. Kocher, Joshua Jaffe, and Benjamin Jun. Differential power analysis. In Michael J. Wiener, editor, *CRYPTO'99*, volume 1666 of *LNCS*, pages 388–397, Santa Barbara, CA, USA, August 15–19, 1999. Springer, Berlin, Germany.
17. Ueli Maurer, Krzysztof Pietrzak, and Renato Renner. Indistinguishability amplification. *IACR Cryptology ePrint Archive*, 2006:456, 2006.
18. Ueli M. Maurer, Krzysztof Pietrzak, and Renato Renner. Indistinguishability amplification. In Alfred Menezes, editor, *CRYPTO 2007*, volume 4622 of *LNCS*, pages 130–149, Santa Barbara, CA, USA, August 19–23, 2007. Springer, Berlin, Germany.
19. Silvio Micali and Leonid Reyzin. Physically observable cryptography (extended abstract). In Moni Naor, editor, *TCC 2004*, volume 2951 of *LNCS*, pages 278–296, Cambridge, MA, USA, February 19–21, 2004. Springer, Berlin, Germany.
20. Moni Naor and Gil Segev. Public-key cryptosystems resilient to key leakage. In Shai Halevi, editor, *CRYPTO 2009*, volume 5677 of *LNCS*, pages 18–35, Santa Barbara, CA, USA, August 16–20, 2009. Springer, Berlin, Germany.
21. J. M. Pollard. A generalisation of the theorem of cauchy and davenport. *Journal of the London Mathematical Society*, s2-8(3):460–462, 1974.
22. Emmanuel Prouff and Matthieu Rivain. Masking against side-channel attacks: A formal security proof. In Thomas Johansson and Phong Q. Nguyen, editors, *EURO-*

CRYPT 2013, volume 7881 of *LNCS*, pages 142–159, Athens, Greece, May 26–30, 2013. Springer, Berlin, Germany.
23. Anup Rao. An exposition of bourgain's 2-source extractor. *Electronic Colloquium on Computational Complexity (ECCC)*, 14(034), 2007.
24. François-Xavier Standaert, Tal Malkin, and Moti Yung. A unified framework for the analysis of side-channel key recovery attacks. In Antoine Joux, editor, *EUROCRYPT 2009*, volume 5479 of *LNCS*, pages 443–461, Cologne, Germany, April 26–30, 2009. Springer, Berlin, Germany.

A Proofs

A.1 Proof of Proposition 1

Proof. Let X be uniform over the set $\mathbb{G}$ and let ϕ be the canonical quotient homomorphism, that is $\phi(g) = g + \mathbb{H}$. For n shares $X_1, \ldots, X_n$ of X we have that $X_i|\phi(X_i)$ and X are $\left(1 - \frac{|\mathbb{H}|}{|\mathbb{G}|}\right)$-far, because $X_i|\phi(X_i) = y_i$ is uniform over a set of $|\mathbb{G}|/|\mathbb{H}|$ elements, for every choice of y_i. Similarly, $X = \sum_i X_i$ is $\left(1 - \frac{|\mathbb{H}|}{|\mathbb{G}|}\right)$-far from uniform given $\phi(X_i)$ for all i, because $\phi(X) = \sum_{i=1} \phi(X_i) = \sum_i y_i$. To see this, note that for independent uniform U we have

$$\begin{aligned}\mathsf{SD}\left(X, \phi(X_1), \ldots, \phi(X_n); U, \phi(X_1), \ldots, \phi(X_n)\right) &\geqslant \mathsf{SD}\left(X, \sum_i \phi(X_i); U, \sum_i \phi(X_i)\right) \\ &= \mathsf{SD}\left(X, \phi(X); U, \phi(X)\right) \\ &= \mathsf{SD}\left(X; U \middle| \phi(X)\right)\end{aligned}$$

where the first line follows from the fact that applying a function to two random variables only decreases the statistical distance, and the second line uses the homomorphic property of ϕ. The last expression is at least $\left(1 - \frac{|\mathbb{H}|}{|\mathbb{G}|}\right)$ as already observed for uniform X.

A.2 Proof of Proposition 2

Proof. Fix $\mathbb{G} = \mathbb{Z}_2$ and consider a uniform secret X, its shares $X_1, \ldots, X_n$, and leakage functions $f_i = f$ for $i = 1, \ldots, n$ where $f(x)$ flips the bit x with probability $\theta < \frac{1}{2}$. It is easy to see that these functions are $\left(\frac{1}{2} - \theta\right)$-noisy, that is $\mathsf{SD}\left(X_i; U \middle| f(X_i)\right) = \frac{1}{2} - \theta$ where U is an independent uniform random variable. Note that for uniform X (and any functions f_i) we have the equality of distributions

$$\left(f_1(X_1), \ldots, f_n(X_n), X\right) \stackrel{d}{=} \left(f_1(Z_1), \ldots, f_n(Z_n), Z_1 + \ldots + Z_n\right).$$

where $\{Z_i\}_{i=1}^n$ are independent and uniform on $\mathbb{G}$ (see Section 3.1.2). As a consequence we get

$$\mathsf{SD}\left(X; U \middle| f(X_1), \ldots, f(X_n)\right) = \mathsf{SD}\left(Z_1 + \ldots + Z_n; U \middle| f_1(Z_1), \ldots, f_n(Z_n)\right)$$

One can check that every X'_i has bias $\delta = \frac{1}{2} - \theta$ (it outputs a bit with probability $\frac{1}{2} \pm \delta$) conditioned on $f(X'_i) = y$ for every $y \in \{0,1\}$. Since the xor-sum $Y_1 + Y_2 + \ldots + Y_n$ of δ-biased independent bits $Y_1, Y_2, \ldots, Y_n$ has bias exactly $2^{n-1}\delta^n$, we conclude that $\mathsf{SD}\left(X'_1 + \ldots + X'_n; U \middle| f_1(X'_1), \ldots, f_n(X'_n)\right) = 2^{n-1}\delta^n$. To go below ϵ, we need $(1-2\theta)^n < 2\epsilon$ or

$$n > \ln\left((2\epsilon)^{-1}\right) / \ln\left((1-2\theta)^{-1}\right).$$

Finally, consider the case $\mathbb{G} = \mathbb{Z}_p$. We proceed as in the previous case, achieving

$$\mathsf{SD}\left(X; U \middle| f(X_1), \ldots, f(X_n)\right) = \mathsf{SD}\left(Z_1 + \ldots + Z_n; U \middle| f_1(Z_1), \ldots, f_n(Z_n)\right)$$

for arbitrary functions. We take the functions f_i so that the distribution of Z_i given $f(Z_i) = y_i$ for every i has the following form:

$$\mu_{Z_i | f(Z_i) = y_i} = \mu_{\mathbb{G}} + \delta\mu_a - \delta\mu_A$$

where a is a point and A is a set such that $a \notin A$, $|A| = \delta|\mathbb{G}|$. As it follows from the proof of Lemma 10 in Appendix A.8, we can choose A so that $|\mathbb{E}\phi(V_i)| \geqslant 1 - \theta$ for some character ϕ, where V_i is the distribution of Z_i conditioned on $f(Z_i)$. This means that the Fourier transform $\hat{V}_i$ of V_i is at least $1-\theta$ in the supremum norm, that is $\|\hat{V}_i\|_\infty \geqslant 1 - \theta$. Since the Fourier transform is multiplicative under convolution (summing independent variables) we see that we can prepare functions f_i so that $\|\hat{V}\|_\infty \geqslant (1-\theta)^n$, where $V = V_1 + \ldots + V_n$. The Parseval identity gives us $\|\hat{V}\|_2 = \|\mu_Z - \mu_U\|_2$. Since $\|\hat{Z}\|_\infty \leqslant \|\hat{V}\|_2$ and $\|\mu_V - \mu_U\|_2 \leqslant \|\mu_V - \mu_U\|_1$ we finally obtain

$$(1-\theta)^n \leqslant \|\mu_V - \mu_U\|_1 = \mathsf{SD}(Z_1 + \ldots + Z_n; U \,|\, f(Z_1) = y_1, \ldots, f(Z_n) = y_n)$$

The claim follows now by averaging over different values of $y_1, \ldots, y_n$, exactly as in the previous case.

A.3 Proof of Lemma 2.

We prove the following version, from which we conclude Lemma 2.

> Suppose that X is uniform and X_i be the encoding of X. Let g be a probabilistic function, $(G_i)_i$ be the encoding of $G = g(X)$ and let f_i be noisy leakage functions. Then we have
>
> $$\beta(X | (f_i(G_i))_i) \leqslant 3|\mathbb{G}| \cdot \beta(X | (f_i(X_i))_i) \tag{15}$$

Proof. Let V be uniform and $(V_i)_i$ be the encoding of V and let X', V' be independent copies of X, V. Note that $X, (f_i(G_i))_i$ is identically distributed as

$X, (f_i(V_i))_i | V = g(X)$. Therefore

$$\begin{aligned}\Pr[X = x | (f_i(G_i))_i = (y_i)_i] &= \Pr[X = x | (f_i(V_i))_i = (y_i)_i, V = g(X)] \\ &= \frac{\Pr[X = x, (f_i(V_i))_i = (y_i)_i, V = g(x)]}{\Pr[(f_i(V_i))_i = (y_i)_i, V = g(X)]} \\ &= \frac{\Pr[X = x] \Pr[(f_i(V_i))_i = (y_i)_i, V = g(x)]}{\sum_{x'} \Pr[X = x'] \Pr[(f_i(V_i))_i = (y_i)_i, V = g(x')]} \\ &= \frac{\Pr[V = g(x) | (f_i(V_i))_i = (y_i)_i]}{\sum_{x'} \Pr[V = g(x') | (f_i(V_i))_i = (y_i)_i]} \end{aligned} \tag{16}$$

Let $\epsilon(x) = \Pr[V = x | (f_i(V_i))_i = (y_i)_i] - \frac{1}{|G|}$. Suppose first, that g is deterministic. We have

$$\begin{aligned}\Pr[X = x | (f_i(G_i))_i = (y_i)_i] - \frac{1}{|G|} &= \frac{\frac{1}{|G|} + \epsilon(g(x))}{1 + \sum_{x'} \epsilon(g(x'))} - \frac{1}{|G|} \\ &= \frac{|G|\epsilon(g(x)) - \sum_{x'} \epsilon(x')}{|G|(1 + \sum_{x'} \epsilon(g(x')))} \end{aligned} \tag{17}$$

and

$$\sum_x \left| \Pr[X = x | (f_i(G_i))_i = (y_i)_i] - \frac{1}{|G|} \right| = \frac{\frac{1}{|G|} \sum_x \left| \epsilon(g(x)) - \frac{1}{|G|} \sum_{x'} \epsilon(g(x')) \right|}{\frac{1}{|G|} + \frac{1}{|G|} \sum_{x'} \epsilon(g(x'))} \tag{18}$$

Note that $\left| \epsilon(g(x)) - \frac{1}{|G|} \sum_{x'} \epsilon(g(x')) \right| \leqslant \sum_{x'} |\epsilon(x')|$ and $\frac{1}{|G|} \sum_{x'} \epsilon(g(x')) \leqslant \sum_{x'} \epsilon(x')$. If $\sum_{x'} \epsilon(x') \leqslant \frac{1}{\frac{3}{2}|G|}$ then we obtain

$$\sum_x \left| \Pr[X = x | (f_i(G_i))_i = (y_i)_i] - \frac{1}{|G|} \right| \leqslant \frac{\sum_{x'} |\epsilon(x')|}{\frac{1}{|G|} - \frac{1}{\frac{3}{2}|G|}} = 3|G| \sum_{x'} \epsilon(x') \tag{19}$$

otherwise

$$\sum_x \left| \Pr[X = x | (f_i(G_i))_i = (y_i)_i] - \frac{1}{|G|} \right| \leqslant 2 \leqslant 3|G| \sum_{x'} \epsilon(x') \tag{20}$$

This way, we have shown

$$\Delta(X; X' | (f_i(G_i))_i = (y_i)_i) \leqslant 3|G| \Delta(V; V' | (f_i(V_i))_i = (y_i)_i) \tag{21}$$

and by taking the average the result follows. If g is randomized, the proof is the same but $\epsilon(g(x))$ is replaced by $\mathbf{E}_g \epsilon(g(x))$ (note that we have $\beta(X | (f_i(G_i))_i) \leqslant \beta(g(X) | (f_i(G_i))_i)$.

A.4 Proof of Lemma 3.

We start with the following observation: suppose that X_i for $i = 1, \ldots, n$ are shares of the *uniform* secret X. Let X'_i for $i = 1, \ldots, n$ be all uniform and independent. Then we have the following equality of distributions

$$(X, (X_1, \ldots, X_n)) \stackrel{d}{=} \left(\sum_{i=1}^{n} X'_i,\ (X'_1, \ldots, X'_n) \right) \tag{22}$$

Therefore,

$$(X, (f_i(X_i))_{i=1}^n) \stackrel{d}{=} \left(\sum_{i=1}^{n} X'_i,\ (f_i(X'_i))_{i=1}^n \right). \tag{23}$$

As a consequence we obtain the following equality

$$\beta(X|(f_i(X_i))_{i=1}^n) = \Delta\left(\sum_{i=1}^{n} X'_i;\ U \middle| (f_i(X'_i))_{i=1}^n \right) \tag{24}$$

Thus, our problem reduces to investigate the random walk on $\mathbb{G}$ defined as $\sum_{i=1}^n X'_i | f_i(X'_i)$. We need to show that it (under some restrictions) eventually approaches the uniform distribution as n increases, and estimate the convergence speed.

A.5 Proof of Lemma 4

Proof. We can assume that $\delta + 2\gamma < 1$. We start with the following observation:

Claim. Suppose that $\delta_1, \ldots, \delta_n$ are independent random variables with expected value at most $\delta < 1$. Then with probability $1 - \exp(-2n\gamma^2)$, at least $n' = \gamma n$ of them is smaller than $\delta + 2\gamma$.

Proof (Proof of Claim). With probability $1 - \exp(-2n\gamma^2)$ we have $\frac{1}{n}\sum_i \delta_i < \delta + \gamma$. Let n' be the number of i's for which $\delta_i < \delta + 2\gamma$. Since we have $\sum_i \delta_i > (n - n')(\delta + 2\gamma)$, with probability $1 - \exp(-2n\gamma^2)$ it holds that $n(\delta + \theta) > (n - n')(\delta + 2\gamma)$ or $n' > \frac{\gamma}{\delta + 2\gamma} \cdot n > \gamma n$.

By applying the claim we see that with probability $1 - \exp(-2n\theta^2)$ over $(y_i) \leftarrow (Y_i)_i$, there always exists a set $I \subset \{1, \ldots, n\}$ such that $|I| \geqslant n'$ (possibly depending on $(y_i)_i$) such that $\mathsf{SD}\left(Z_i; U \middle| Y_i = y_i\right) \leqslant \delta + 2\theta$ for $i \in I'$. Since the distributions $(Z_i, Y_i)_i$ are independent for different i's and since $U + Z \stackrel{d}{=} U$ for any independent random variable Z, from the elementary properties of the statistical distance we obtain

$$\begin{aligned} \mathsf{SD}\left(\sum_{i=1}^{n} Z_i;\ U \middle| (Y_i)_i = (y_i)_i \right) &= \mathsf{SD}\left(\sum_{i \in I} Z_i + \sum_{i \notin I} Z_i;\ U + \sum_{i \notin I} Z_i \middle| (Y_i)_i = (y_i)_i \right) \\ &= \mathsf{SD}\left(\sum_{i \in I} Z_i;\ U \middle| (Y_i)_i = (y_i)_i \right). \end{aligned} \tag{25}$$

The lemma now easily follows, as for every I as above we have

$$\mathsf{SD}\left(\sum_{i\in I} Z_i;\ U \,\middle|\, (Y_i)_i = (y_i)_i\right) \leqslant \max_{(Z'_i)_i:\ \mathsf{SD}(Z'_i;U)\leqslant\delta'} \Delta\left(\sum_{i=1}^{n'} Z'_i; U\right). \tag{26}$$

A.6 Proof of Theorem 2

Proof. Let μ_i be a distribution of Z_i for $i = 1, 2$ and let μ_U denotes the uniform measure. Let $\Delta(\mu_i, \mu_U) = \delta_i$. Note that we can decompose $\mu_i = \mu_U + \delta_i\mu_i^+ - \delta_i\mu_i^-$. Therefore

$$\begin{aligned}\mu_1 * \mu_2 &= \left(\mu_U + \delta_1\mu_1^+ - \delta_1\mu_1^-\right) * \left(\mu_U + \delta_2\mu_2^+ - \delta_2\mu_2^-\right)\\ &= \mu_U + \delta_1\delta_2\left(\mu_1^+ * \mu_2^+ + \mu_1^- * \mu_2^- - \mu_1^+ * \mu_2^- - \mu_1^- * \mu_2^+\right)\end{aligned} \tag{27}$$

where we have made use of the fact that $\mu_U * \nu = \mu_U$ for any distribution ν. Now we have

$$\mathsf{SD}(\mu_1 * \mu_2; \mu_U) = \frac{1}{2}\left\|\mu_1^+ * \mu_2^+ + \mu_1^- * \mu_2^- - \mu_1^+ * \mu_2^- - \mu_1^- * \mu_2^+\right\|_{\ell_1(G)} \tag{28}$$

This is clearly at most 2. To identify the worst case choice of μ_i that maximizes this quantity, observe that we have to bound the last expression with respect to the constraints

$$\left\|\mu_i^-\right\|_{\ell_\infty(G)} \leqslant \frac{1}{\delta_i|G|} \quad i = 1, 2 \tag{29}$$

which come from the fact that μ_i, as decomposed, has to be positive. There is no restriction on μ_i^+. Note now that the form $\mu_1^+ * \mu_2^+ + \mu_1^- * \mu_2^- - \mu_1^+ * \mu_2^- - \mu_1^- * \mu_2^+$ is bilinear with respect to measures μ_i^+, μ_i^- and the real-valued function $\mu \to \|\mu\|_{\ell_1(G)}$ defined on *signed* measures is convex. It follows that $\left\|\mu_1^+ * \mu_2^+ + \mu_1^- * \mu_2^- - \mu_1^+ * \mu_2^- - \mu_1^- * \mu_2^+\right\|_{\ell_1(\mathbb{G})}$ attains its maximal value for measures that are extreme points of their domain. Looking at the restrictions in (29) we see that this is the case where μ_i^+ are a point mass and μ_i^- are uniform over the subset of cardinality $\delta_i|G|$[6]. Thus we can assume that $\mu_1^+ = \mu_a$, $\mu_2^+ = \mu_b$ are point mass at a, b and $\mu_1^- = \mu_A$, $\mu_2^- = \mu_B$ are uniform over A, B where $|A| = \delta_1|G|$ and $|B| = \delta_2|G|$. This way our quantity simplifies to

$$\begin{aligned}\|\mu_a * \mu_b - \mu_a * \mu_B - \mu_b * \mu_A + \mu_A * \mu_B\|_{\ell_1(G)} &= \|\mu_{a+b} - \mu_{B+a} - \mu_{b+A} + \mu_A * \mu_B\|_{\ell_1(G)}\\ &= \|\mu_0 - \mu_{B-b} - \mu_{A-a} + \mu_{A-a} * \mu_{B-b}\|_{\ell_1(G)}\end{aligned} \tag{30}$$

where we have used the fact that the norm $\ell_1(G)$ is shift invariant and that a point mass act as shifts under the convolution.

[6] Otherwise we could decompose either the positive part μ^+ into a combination of two distributions (when μ^+ is supported on more than one point) or the negative part μ^- (when the constraint Equation (29) is not binding at some point in the support).

From this we easily derive the following result

Lemma 7 (Mixing time for a sum of random variables on a group). *Let $\{Z_i\}_{i=1,\dots,n}$ be independent random variables on an abelian group $\mathbb{G}$, such that $\Delta(Z_i;U) = \delta_i$ where $\delta_i \leqslant \frac{1}{2} - \theta$ and $\theta > 0$. Then for $n \geqslant \log(1/\epsilon)/(2\theta)$ it holds that*

$$\mathsf{SD}\left(\sum_{i=1}^{n} Z_i; U\right) \leqslant \epsilon \tag{31}$$

A.7 Proof of Lemma 6

We will show that the constant given by (8) could be much better estimated when $\mathbb{G} = \mathbb{Z}_p$. The trivial estimate is 2, however this is possible only if $A + B$ is disjoint with A and B. Here we remind the following result due to Cauchy and Davenport

Theorem (Cauchy-Davenport Theorem). *For any $A, B \subset \mathbb{Z}_p$, where p is prime, we have $|A + B| \geqslant \min(|A| + |B| - 1, p)$.*

In view of this result, a better estimate is impossible if only $\delta_1+\delta_2+\max(\delta_1,\delta_2) > 1+1/p$. From this we know that the estimate (7) is not sharp for $\delta_1+\delta_2 \geqslant \frac{2}{3}+\frac{2}{3p}$. Therefore we expect to improve the estimate for *sufficiently big* values of $\delta_1 + \delta_2$ whereas for the smaller we can still use the general result. To this end, we will need a result stronger than the Cauchy-Davenport Theorem

Theorem (Pollard's Theorem [21]). *For any $A, B \subset \mathbb{Z}_p$, where p is prime, we have*

$$\sum_{x\in\mathbb{Z}_p} r_{A,B}(x)\mathbf{1}_{\{r_{A,B}(x)>t\}}(x) \leqslant |A||B| - t(|A| + |B| - t) \tag{32}$$

where $r_{A,B}(x)$ counts in how many different ways can we represent x as a sum $a + b$ with $a \in A, b \in B$.

Intuitively, Pollard's theorem says that the distribution of $r_{A,B}(x)$ cannot be too "heavy tailed".

Proof (of Lemma 6). In fact, we will show that $\mu_A * \mu_B$ *always* puts some large mass on every sufficiently big set C, essentially on A or B. Observe first that

$$\mu_A * \mu_B(x) = \frac{r_{A,B}(x)}{|A||B|} \tag{33}$$

where $r_{A,B}(x)$ counts for how many different ways can we represent x as a sum $a+b$ with $a \in A, b \in B$. By trivial estimates $r_{A,B}(x) \leqslant \min(|A|, |B|)$ we see that

$$\mu_A * \mu_B(x) \leqslant \min(\mu_A(x), \mu_B(x)), \quad x \in A \cup B \tag{34}$$

Using this we can estimate the expression in (8) as follows

$$\begin{aligned}\frac{1}{2}\left\|\mu_A+\mu_B-\mu_A*\mu_B-\mu_0\right\|_{\ell_1(G)} &= \max_{S\subset G}\left(\mu_A(S)+\mu_B(S)-\mu_A*\mu_B(S)-\mu_0(S)\right)\\ &\leqslant \max_{S\subset G}\left(\mu_A(S)+\mu_B(S)-\mu_A*\mu_B(S)\right)\\ &= \left(\mu_A(A\cup B)+\mu_B(A\cup B)-\mu_A*\mu_B(A\cup B)\right)\\ &= 2-\frac{1}{|A||B|}\sum_{x\in A\cup B} r_{A,B}(x) \qquad (35)\end{aligned}$$

From Pollard's theorem, for every set C we obtain

$$\sum_x r_{A,B}(x)\mathbf{1}_C(x) \geqslant \sum_x r_{A,B}(x)\mathbf{1}_{r_{A,B}(x)\leqslant t}(x) - t(|G|-|C|) \qquad (36)$$

$$\geqslant t(|A|+|B|-t)-t(p-|C|) = t(|A|+|B|+|C|-p-t) \qquad (37)$$

the maximum is for $t_{\max}=\frac{|A|+|B|+|C|-p}{2}$ provided that $|A|+|B|+|C|-p\geqslant 0$. We check that the required inequality $|A|+|B|-p\leqslant t_{\max}\leqslant \min(|A|,|B|)$ is true if only the set C satisfies

$$|A|+|B|-p\leqslant |C|\leqslant p-||B|-|A||. \qquad (38)$$

Note that if $t_{\max}\notin\mathbb{Z}$ then the conditions above are still sufficient provided that we replace $t_{\max}$ with $\lceil t_{\max}\rceil$ or $\lfloor t_{\max}\rfloor$. Considering the function $f(t)=t(|A|+|B|+|C|-p-t)$ by the mean-value theorem we see that

$$\begin{aligned}|f(\lceil t_{\max}\rceil)-f(\lfloor t_{\max}\rfloor)| &\leqslant \max_{\xi\in[\lfloor t_{\max}\rfloor,\lceil t_{\max}\rceil]} f'(\xi)\\ &= \max_{\xi\in[\lfloor t_{\max}\rfloor,\lceil t_{\max}\rceil]}(-2\xi+|A|+|B|+|C|-p)\\ &\leqslant -2\left(t_{\max}+\frac{1}{2}\right)+|A|+|B|+|C|-p=1. \qquad (39)\end{aligned}$$

Therefore, we obtain

$$\sum_x r_{A,B}(x)\mathbf{1}_C(x)\geqslant\left\lfloor\frac{(|A|+|B|+|C|-p)^2}{4}\right\rfloor \qquad (40)$$

Setting $C\subset A\cup B$ such that $|C|=\min\left(\max(|A|,|B|),p-||A|-|B||\right)$ we see that the condition $|C|\geqslant |A|+|B|-p$ is satisfied. Provided that $|A|+|B|+|C|-p\geqslant 0$ we obtain

$$2-\frac{1}{|A||B|}\sum_{x\in A\cup B} r_{A,B}(x)\leqslant 2-\frac{(\delta_1+\delta_2+\min(\max(\delta_1,\delta_2),1-|\delta_1-\delta_2|)-1)^2+p^{-2}}{4\delta_1\delta_2} \qquad (41)$$

and the result follows by (8).

From this result we obtain the following result, from which we conclude the part (ii) of Theorem 1 by replacing θ by $\frac{\theta}{4}$ and combining with Corollary 1 in the same way as in the the derivation of part (ii).

Lemma 8 (Mixing time for a sum of random variables on $\mathbb{Z}_p$). *Let $\{Z_i\}_{i=1,\dots,n}$ be independent random variables on $\mathbb{G} = \mathbb{Z}_p$, such that $\mathsf{SD}(Z_i; U) \leqslant \delta_i$ where $\delta_i \leqslant 1 - p^{-1} - \theta$ and $\theta > 0$. Then for $n \geqslant 3 \cdot 2^{4/\theta} \log(1/\epsilon)/\theta$ it holds that*

$$\mathsf{SD}\left(\sum_{i=1}^{n} Z_i; U\right) \leqslant \epsilon \tag{42}$$

Proof. First, using Corollary 3, we show that every sufficiently long sum has distance at most $\frac{1}{3}$. Once we have that, it is enough to split the entire sum into sufficiently many blocks and then apply Theorem 2. Consider $n_0 = 2^m$. By applying Lemma 6 several times we see that

$$\mathsf{SD}\left(\sum_{i=1}^{n_0} Z_i;\ U\right) \leqslant B_m \tag{43}$$

where B_i are numbers defined by the following recursion

$$B_0 = 1 - p^{-1} - \theta, \quad B_i = h(B_{i-1}, B_{i-1}) \text{ for } i \geqslant 1 \tag{44}$$

We will prove that $1-p^{-1}$ is the *repelling point*: if we start from any B_0 satisfying $\frac{1}{3} \leqslant B_0 < 1 - p^{-1}$ then B_i decreases below $\frac{1}{3}$. Let $C_i = 1 - B_i$. If $B_{i-1} \geqslant \frac{1}{3}$, then by Corollary 3 we get

$$\begin{aligned} C_i =& 1 - B_i = 1 - h(B_{i-1}, B_{i-1}) \\ &= 2B_{i-1}^2 - \frac{(3B_{i-1} - 1)^2}{4} - \frac{1}{4p^2} \\ &= C_{i-1} + \frac{C_{i-1}^2}{4} - \frac{1}{4p^2} \\ &= C_{i-1}\left(1 + \frac{C_{i-1}}{4}\left(1 - \frac{1}{C_{i-1}^2 p^2}\right)\right) \end{aligned} \tag{45}$$

From this we conclude that if $\frac{1}{3} \leqslant B_{i-1} < 1 - p^{-1}$ then $C_{i-1} > p^{-1}$ and hence $C_i > C_{i-1}$ or equivalently $B_i < B_{i-1}$. Moreover, if $C_{i-1} \geqslant p^{-1} + \theta$, we get

$$\begin{aligned} C_i &\geqslant C_{i-1}\left(1 + \theta \cdot \frac{2 + p\theta}{4 + 4p\theta}\right) \\ &\geqslant C_{i-1}\left(1 + \frac{\theta}{4}\right) \end{aligned} \tag{46}$$

Since $C_i \leqslant 1$ and $C_0 \geqslant p^{-1} + \theta > \theta$, for some $j \leqslant \frac{4}{\theta} \log\left(\frac{1}{\theta}\right)$ we must have $B_j < \frac{1}{3}$. Thus for $m = \lceil \frac{4}{\theta} \log\left(\frac{1}{\theta}\right) \rceil$ we have

$$\mathsf{SD}\left(\sum_{i=1}^{2^m} Z_i; U\right) \leqslant \frac{1}{3} \tag{47}$$

Consider $\ell = \log(1/\epsilon)$ blocks of random variables $\{(Z_{2^m j+1}, \ldots, (Z_{2^m j+2^m})\}_j$ for $j = 0, \ldots, N-1$. For every such a 2^m-element block from the last observation it follows that

$$\mathsf{SD}\left(\sum_{i=1}^{2^m} Z_{2^m j+i};\ U\right) \leqslant \frac{1}{3} \tag{48}$$

Applying ℓ times Lemma 2 yields the estimate

$$\mathsf{SD}\left(\sum_{i=1}^{\ell 2^m} Z_i; U\right) \leqslant \left(\frac{2}{3}\right)^{\ell} \tag{49}$$

which finishes the proof.

A.8 Harmonic Analysis

We need the following lemma, being a generalization of Vazirani's XOR lemma.

Lemma 9 (XOR lemma for abelian groups, [23]). *Let Z be a distribution over a finite abelian group $\mathbb{G}$, such that $|\mathbb{E}\phi(Z)| \leqslant \epsilon$ for every non-trivial character ϕ on $\mathbb{G}$. Then X is $\epsilon\sqrt{|\mathbb{G}|}$-close to uniform.*

Lemma 10 (Mixing times of random sums over $\mathbb{Z}_p$). *Let $\{Z_i\}_{i=1,\ldots,n}$ be independent random variables on $\mathbb{G} = \mathbb{Z}_p$, such that $\mathsf{SD}(Z_i; U) \leqslant 1 - p^{-1} - \theta$ and $\theta > 0$. Then for $n \geqslant 8 \cdot \log(|\mathbb{G}|/\epsilon)/\theta^3$ it holds that*

$$\mathsf{SD}\left(\sum_{i=1}^{n} Z_i; U\right) \leqslant \epsilon \tag{50}$$

Proof. We apply some facts from harmonic analysis. Let Z_i be the worst-case distributions that maximize $\mathsf{SD}(\sum_{i=1}^{n} X_i, U)$ under the constraints $\mathsf{SD}(Z_i, U) \leqslant 1 - p^{-1} - \theta$. By Equation (6) that

$$\mu_{Z_i} = \left(1 - \frac{|A|}{p}\right) \cdot \mu_0 + \frac{|A|}{p} \cdot \frac{1}{|A|}\mu_A, \quad |A| = p\theta \tag{51}$$

Consider a non-trivial character $\phi(x) = \exp(2k\pi i/p)$ on $\mathbb{Z}_p$. Since $A \neq \emptyset$ we have $\theta \geqslant \frac{1}{p}$. We will show an upper bound on $\mathbf{E}\phi(X_i)$. First, observe that

$$|\mathbf{E}\phi(X_i)| = \left|1 - \frac{|A|}{p} + \frac{|A|}{p} \cdot \frac{1}{|A|} \sum_{x \in A} \exp\left(\frac{2k\pi i x}{p}\right)\right| \tag{52}$$

is maximized exactly when $kA = \left\{-\frac{|A|-1}{2}, \ldots, 0, \ldots, \frac{|A|-1}{2}\right\}$. Indeed, we have

Claim. For any subset A of $\mathbb{Z}_p$ and any non-trivial character ϕ over $\mathbb{Z}_p$ we have the following estimate

$$|\mathbf{E}\phi(X_i)| = \left|1 - \frac{|A|}{p} + \frac{|A|}{p} \cdot \frac{1}{|A|} \sum_{x \in A} \exp(2k\pi i x/p)\right| \leqslant 1 - \theta + \frac{\sin \pi\theta}{p \sin \frac{\pi}{p}}$$

Proof. Note that every non-trivial character is of the form $\phi(x) = \exp(2k\pi i x/p)$ where $k \in \{1, 2, \ldots, p-1\}$. Next, we can assume that $k = 1$, by replacing A with $A' = k \cdot A$, which doesn't change the set size. Now, by the triangle inequality we have

$$\left|1 - \frac{|A|}{p} + \frac{1}{p} \sum_{x \in A} \phi(x)\right| \leqslant 1 - \frac{|A|}{p} + \frac{|A|}{p} \cdot \left|\frac{1}{|A|} \sum_{x \in A} \phi(x)\right|$$

It remains to estimate $|m_A|$ where

$$m_A = \frac{1}{|A|} \sum_{x \in A} \phi(x)$$

is the mass center of the set $\phi(A) = \{\phi(x) : \ x \in A\}$. Note that $\phi(A)$ may be any arbitrary $|A|$-element subset of the set of all p-th roots of unity (because ϕ is a bijection), see Figure 2 for an illustration. Our task is therefore to maximize

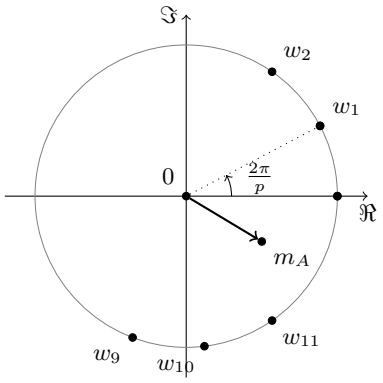

Fig. 2: The mass center of the set $\phi(A)$ should be as close to the circle as possible.

the length of m_A which happens when A is the set of subsequent unity roots. In particular

$$|\mathbf{E}\phi(X_i)| \leqslant \left|1 - \frac{|A|}{p}\right| + \frac{1}{p}\left|\sum_{|x| \leqslant \frac{|A|-1}{2}} \exp(2\pi i x)\right| = 1 - \theta + \left|\frac{\sin \pi\theta}{p \sin \frac{\pi}{p}}\right|, \qquad (53)$$

where the last equality follows by known trigonometric identities. Since $\theta < 1$ we can omit the absolute value here, and this finishes the proof.

We will prove the following inequality

Claim. For any $\theta < 1$ and any $c \leqslant \frac{4}{3} - \frac{\pi^2}{18}$, we have $1 - \theta + \frac{\sin \pi\theta}{p \sin \frac{\pi}{p}} \leqslant 1 - c\theta^3$

Proof. We want to prove that $f(\theta) = c\theta^3 - \theta + \frac{\sin \pi\theta}{p \sin \frac{\pi}{p}} \leqslant 0$. We have $f(0) = 0$ and $\frac{\partial f(\theta)}{\partial \theta} = -1 + 3c\theta^2 + \pi \frac{\cos \pi\theta}{p \sin \frac{\pi}{p}}$ Since for $t \in \left[0, \frac{\pi}{2}\right]$ it holds that $\cos t \leqslant 1 - \frac{4t^2}{\pi^2}$ and $\sin t \geqslant t - \frac{t^3}{6}$, we obtain

$$\frac{\partial f(\theta)}{\partial \theta} \leqslant -1 + 3c\theta^2 + \frac{1 - 4\theta^2}{1 - \frac{\pi^2}{6p^2}} = \frac{\frac{\pi^2}{6p^2} - \theta^2 \left(4 - 3c + \frac{3c\pi^2}{6p^2}\right)}{1 - \frac{\pi^2}{6p^2}} \tag{54}$$

and since $\theta \geqslant \frac{1}{p}$, the result follows.

From the last claim it follows that we can put $c = \frac{1}{2}$ and thus

$$\begin{aligned} |\mathbf{E}\phi(X)| &= \left|\mathbf{E}\phi\left(\sum_{i=1}^{n} X_i\right)\right| \\ &= \prod_{i=1}^{n} |\mathbf{E}\phi(X_i)| \\ &\leqslant \left(1 - \theta^3/2\right)^n . \end{aligned} \tag{55}$$

Now the result follows by Lemma 9.