On Perfect and Adaptive Security in Exposure-Resilient Cryptography

Yevgeniy Dodis[1], Amit Sahai[2], and Adam Smith[3]

[1] Department of Computer Science, New York University,
251 Mercer St, New York, NY 10012, USA. dodis@cs.nyu.edu
[2] Department of Computer Science, Princeton University,
35 Olden St, Princeton, NJ 08540, USA. sahai@cs.princeton.edu
[3] Laboratory for Computer Science, Massachusetts Institute of Technology,
545 Main St, Cambridge, MA 02139, USA. asmith@theory.lcs.mit.edu

Abstract. We consider the question of *adaptive* security for two related cryptographic primitives: all-or-nothing transforms and exposure-resilient functions. Both are concerned with retaining security when an intruder learns some bits of a string which is supposed to be secret: *all-or-nothing transforms* (AONT) protect their input even given partial knowledge of the output; *exposure-resilient functions* (ERF) hide their output even given partial exposure of their input. Both of these primitives can be defined in the perfect, statistical and computational settings and have a variety of applications in cryptography. In this paper, we study how these notions fare against adaptive adversaries, who may choose which positions of a secret string to observe on the fly.

In the *perfect* setting, we prove a new, strong lower bound on the constructibility of (perfect) AONT. This applies to both standard and adaptively secure AONT. In particular, to hide an input as short as $\log n$ bits, the adversary must see *no more than half* of the n-bit output. This bound also provides a new impossibility result on the existence of (ramp) secret-sharing schemes [6] and relates to a combinatorial problem of independent interest: finding "balanced" colorings of the hypercube.

In the statistical setting, we show that adaptivity adds strictly more power to the adversary. We relate and reduce the construction of adaptive ERF's to that of *almost-perfect resilient functions* [19], for which the adversary can actually *set* some of the input positions and still learn nothing about the output. We give a probabilistic construction of these functions which is essentially optimal and substantially improves on previous constructions of [19, 5]. As a result, we get nearly optimal adaptively secure ERF's and AONT's. Finally, extending the statistical construction we obtain optimal *computational* adaptive ERF's, "public-value" AONT's and resilient functions.

1 Introduction

Recently, there has been an explosion of work [23,9,10,20,18,7,1,26,14] surrounding an intriguing notion introduced by Rivest called the *All-Or-Nothing Transform* (AONT) [23]. Roughly speaking, an AONT is a randomized mapping which

B. Pfitzmann (Ed.): EUROCRYPT 2001, LNCS 2045, pp. 301–324, 2001.

can be efficiently inverted if given the output in *full*, but which leaks *no* information about its input to an adversary even if the adversary obtains *almost all* the bits of the output. The AONT has been shown to have important cryptographic applications ranging from increasing the efficiency of block ciphers [20,18,7] to protecting against almost complete exposure of secret keys [10]. The first formalization and constructions for the AONT were given by Boyko [9] in the Random-Oracle model. However, recently Canetti et al. [10] were able to formalize and exhibit efficient constructions for the AONT in the standard computational model. They accomplished this goal by reducing the task of constructing AONT's to constructing a related primitive which they called an *Exposure-Resilient Function* (ERF) [10]. An ERF is a deterministic function whose output looks random to an adversary even if the adversary obtains *almost all* the bits of the input. A salient feature of the work of [10] is the fact that they were able to achieve good results for the computational (and most cryptographically applicable) versions of these notions by first focusing on the perfect and statistical forms of AONT's and ERF's.

1.1 Background

We first recall informally the definitions of the two main notions we examine in this paper. An ℓ-AONT [23,9,10] is an efficiently computable and *invertible* randomized transformation T, which transforms any string x into a pair of strings (y_s, y_p), respectively called the *secret* and the *public* part of T. While the invertability of T allows to reconstruct x from the *entire* $T(x) = (y_s, y_p)$, we require that any adversary learning all of y_p and all but ℓ bits of y_s obtains "no information" about x.

On the other hand, an ℓ-ERF [10] is an efficiently computable deterministic function f on strings such that even if an adversary learns all but ℓ bits of a *randomly chosen* input r, it still cannot distinguish the output $f(r)$ from a random string. As usual, we can define *perfect, statistical,* and *computational* versions of these notions. It is easy to see that in the perfect or statistical settings, the length of the output of an ℓ-ERF can be at most ℓ; whereas for perfect or statistical ℓ-AONT's, the length of the input is at most ℓ. To beat these trivial bounds, one must examine the computational forms of ERF's and AONT's. Indeed, if we are given a pseudorandom generator, it is easy to see that by applying the generator to the output of a perfect or statistical ERF, we can obtain ERF's with arbitrary (polynomial) output size.

Canetti et al. [10] showed that the following simple construction suffices to construct AONT's from ERF's. Given an ℓ-ERF f mapping $\{0,1\}^n$ to $\{0,1\}^k$, we construct an ℓ-AONT T transforming k bits to n bits of secret output and k bits of public output: $T(x) = \langle r,\ f(r) \oplus x \rangle$. Intuitively, if at least ℓ bits of r are missed, then $f(r)$ "looks" random. Hence $f(r) \oplus x$ also looks random, thus hiding all information about the input x.

Applications. The All-Or-Nothing Transform and its variants have been applied to a variety of problems. In the perfect setting, it is a special case of a *ramp* *scheme* [6], useful for sharing secrets efficiently. Its statistical variant can be used to provide secure communication over the "wire-tap channel II", a partly public channel where the adversary can observe almost all the bits communicated (but the sender and the receiver do not know which) [22,3]. In the computational setting, it also has many uses. Rivest [23], and later Desai [14], use it to enhance the security of block ciphers against brute-force key search. Matyas et al. [20] propose to use AONT to increase the efficiency of block ciphers: rather than encrypt all blocks of the message, apply an AONT to the message and encrypt only one or very few blocks. The same idea is used in various forms by Jackobson et al. [18] and Blaze [7] to speed up remotely-keyed encryption. Similarly, it can be combined with authentication to yield a novel encryption technique [24,1]. Several other applications have been suggested by [9,26].

Another class of applications for (computational) AONT's was suggested by Canetti et al. [10]. They considered a situation where one of our most basic cryptographic assumptions breaks down — the secrecy of a key can become partially compromised (a problem called *partial key exposure*). [10] point out that most standard cryptographic definitions do not guarantee (and often violate) security once even a small portion of the key has been exposed. The AONT offers a solution to this problem. Namely, rather than store a secret key x, one stores $y = T(x)$ instead. Now the adversary gets no information about the secret key even if he manages to get all but ℓ bits of y. The problem of *gradual* key exposure is also raised by [10], where information about a (random) private key is slowly but steadily leaked to an adversary. In this situation, the private key can be "renewed" using an ERF to protect it against discovery by the adversary, while additionally providing forward security when the "current" key is totally compromised.

1.2 Adaptive Security

In many of the applications above, the question of adaptive security arises naturally. For example, in the problem of partial key exposure, it is natural to consider an adversary that is able to first gain access to some fraction of the bits of the secret, and then decides which bits to obtain next as a function of the bits the adversary has already seen.

Perfect AONT's and Adaptive Security. In the definition of a *perfect* ℓ-AONT, we demand that any subset of all but ℓ bits of the output must be *completely independent* of the input x.[1] In this case, it is trivial to observe that there is no difference between adaptive and non-adaptive security. Hence, if we could construct good perfect AONT's, this would also solve the problem of constructing adaptively secure AONT's.

Consider ℓ-AONT's that transform k bits to n bits. [10] show how to construct perfect ℓ-AONT's where $\ell = n(\frac{1}{2} + \varepsilon)$ for any $\varepsilon > 0$ (at the expense of smaller $k = \Omega(n)$), but were unable to construct perfect AONT's with $\ell < n/2$ (i.e. perfect AONT's where the adversary could learn more than half of the output).

[1] In the perfect setting, public output is not needed (e.g., can be fixed a-priori).

Perfect AONT's — Our Contribution. In our work, we show that unfortunately this limitation is inherent. More precisely, whenever $n \le 2^k$, the adversary *must* miss at least half of the output in order not to learn anything about the input. We prove this bound by translating the question of constructing perfect ℓ-AONT's to the question of finding *"ℓ-balanced" weighted colorings of the hypercube*, which is of independent combinatorial interest. Namely, we want to color and weight the nodes of the n-dimensional hypercube $\mathcal{H} = \{0,1\}^n$ using $c = 2^k$ colors, such that every ℓ-dimensional subcube of $\mathcal{H}$ is "equi-colored" (i.e. has the same total weight for each of the c colors). We prove our result by non-trivially extending the beautiful lower bound argument of Friedman [15] (which only worked for unweighted colorings) to our setting. Our bound also gives a new bound on *ramp secret sharing schemes* [6]. In such schemes one divides the secret of size k into n schares such that there are two thresholds t and $(t-\ell)$ such that any t shares suffice to reconstruct the secret but no $(t-\ell)$ shares yield any information. To our knowledge, the best known bound for ramp schemes [8,17,21] was $\ell \ge k$. Our results imply a much stronger bound of $\ell \ge t/2$ (when each share is a bit; over larger alphabets of size q we get $\ell > t/q$).

Therefore, we show that despite their very attractive perfect security, perfect AONT's are of limited use in most situations, and do not offer a compelling way to achieve adaptive security.

Statistical ERF's and Adaptive Security. The definition of a *perfect ℓ-*ERF (mapping n bits to k bits) states that the output, when considered jointly with any subset of $(n-\ell)$ bits of the input, must be truly uniform. In this case, clearly once again adaptive and non-adaptive security collapse into one notion. The definition of a (non-adaptive) *statistical* ℓ-ERF, however, allows for the the joint distribution above to be merely *close* to uniform. In this case, the non-adaptive statistical definition does *not* imply adaptive security, and in particular the construction given in [10] of statistical ERF's fails to achieve adaptive security.[2] Intuitively, it could be that a small subset of the input bits S_1 determines some non-trivial boolean relation of another small subset of the input bits S_2 with the output of the function (e.g., for a fixed value of the bits in S_1, one output bit might depend only on bits in S_2). In the adaptive setting, reading S_1 and then S_2 would break an ERF. In the non-adaptive setting, however, any *fixed* subset of the input bits is very unlikely to contain $S_1 \cup S_2$. (A similar discussion applies to AONT's.) In other words, statistical constructions of [10] were able to produce statistical ℓ-ERF's (and ℓ-AONT's) with nearly optimal $\ell = k + o(k)$, but failed to achieve adaptive security, while perfect ERF's achieve adaptive security, but are limitted to $\ell > n/2$ [15].

Statistical ERF's — Our Contribution. Thus, we seek to identify notions lying somewhere in between perfect and statistical (non-adaptive) ERF's that would allow us to construct adaptively secure ERF's (and AONT's), and yet achieve better parameters than those achievable by perfect ERF's (and AONT's). In this task, we make use of *resilient functions* (RF's). These were first defined

[2] For more details, see Section 2.2.

in the perfect setting by Vazirani [28] and first studied by Chor et al. [12] and independently by Bennett et al. [3]. An ℓ-RF is identical to an ℓ-ERF except that the adversary, instead of merely *observing* certain bits of the input, gets to *set* all but ℓ bits of the input.[3] Note that the notions of ERF and RF are the same when considered in the perfect setting. A statistical variant of resilient functions (no longer equivalent to ERF's) was first considered by Kurosawa et al. [19], who also gave explicit constructions of such functions (improved by [5]).

We show that the strong notion of statistical RF's introduced by Kurosawa et al. [19] suffices to construct adaptively secure ERF's (and AONT's). While the construction of Kurosawa et al. [19] already slightly beats the lower bound for perfect ERF's, it is very far from the trivial lower bound of $\ell > k$ (in fact, it is still limited to $\ell > n/2$). We present an efficient probabilistic construction of such "almost-perfect" RF's achieving optimal $\ell = k + o(k)$. While not fully deterministic, our construction has to be run only *once and for all*, after which the resulting efficient function is "good" with probability exponentially close to 1, and can be *deterministically* used in all the subsequent applications. As a result of this construction and its relation to adaptive ERF's and AONT's, we achieve essentially optimal security parameters for adaptive security by focusing on a stronger notion of almost-perfect RF's.

We also take the opportunity to study several variants of statistical RF's and (static/adaptive) ERF's, and give a complete classification of these notions, which may be of additional, independent interest.

Computational Setting. As we pointed out, [10] used their statistical (non-adaptive) constructions to get ERF's and AONT's in the computational setting. We show that the same techniques work with our adaptive definitions. Coupled with our statistical constructions, we get nearly optimal computational constructions as well.

Larger alphabets. To simplify the presentation and the discussion of the results in this paper, as well as to relate them more closely with the previous work, we restrict ourselves to discussing exposure-resilient primitives over the alphabet $\{0, 1\}$. However, all our notions and results can be easily generalized to larger alphabets.

1.3 Organization

In Section 2, we define the central objects of study in our paper, and review some of the relevant previous work of [10]. In Section 3 we study perfect AONT's, relate them to hypecube colorings and prove the strong lower bound on ℓ (showing the limitations of perfect AONT's). Finally, in Section 4 we study variants of statistical ERF's will allow us to achieve adaptive security. We show that "almost-rerfect" RF's of [19] achieve this goal, and exhibit a simple and almost optimal (probabilistic) construction of such functions. In particular, we show

[3] In much of the literature about resilient functions, such a function would be called an $(n - \ell)$-resilient function. We adopt our notation for consistency.

the existence of adaptively secure AONT's and ERF's with essentially optimal parameters.

2 Preliminaries

Let $\{{n \atop \ell}\}$ denote the set of size-ℓ subsets of $[n] = \{1 \dots n\}$. For $L \in \{{n \atop \ell}\}$, $y \in \{0,1\}^n$, let $[y]_{\bar{L}}$ denote y restricted to its $(n-\ell)$ bits *not* in L. We say a function $\epsilon(n)$ is negligible (denoted by $\epsilon = negl(n)$) if for every constant c, $\epsilon(n) = O\left(\frac{1}{n^c}\right)$. We denote an algorithm $\mathcal{A}$ which has oracle access to some string y (i.e., can query individual bits of y) by $\mathcal{A}^y$.

2.1 Definitions for Non-adaptive Adversaries

For static adversaries, the definitions of AONT and ERF can be stated quite efficiently in terms of perfect, statistical or computational indistinguishability (see [16]). For consistency we have also provided a definition of RF (where adaptivity does not make sense, and hence the adversary can be seen as "static").

Note that for full generality, we follow the suggestion of [10] and allow the all-or-nothing transform to have two outputs: a *public* part which we assume the adversary always sees; and a *secret* part, of which the adversary misses ℓ bits.

Definition 1. *A polynomial-time randomized transformation* $T : \{0,1\}^k \to \{0,1\}^s \times \{0,1\}^p$ *is an* ℓ-AONT (all-or-nothing transform) *if*

1. *T is polynomial-time invertible, i.e. there exists efficient I such that for any $x \in \{0,1\}^k$ and any $y = (y_1, y_2) \in T(x)$, we have $I(y) = x$. We call y_1 is the* secret part *and y_2, the* public part *of T.*
2. *For any $L \in \{{s \atop \ell}\}, x_0, x_1 \in \{0,1\}^k$: $\langle x_0, x_1, [T(x_0)]_{\bar{L}}\rangle \approx \langle x_0, x_1, [T(x_1)]_{\bar{L}}\rangle$*[4] *Here $\approx$ can refer to perfect, statistical or computational indistinguishability.*

If $p = 0$, the resulting AONT *is called* secret-only.

Definition 2. *A polynomial time function $f : \{0,1\}^n \to \{0,1\}^k$ is an* ℓ-ERF (exposure-resilient function) *if for any $L \in \{{n \atop \ell}\}$ and for a randomly chosen $r \in \{0,1\}^n$, $R \in \{0,1\}^k$, we have: $\langle [r]_{\bar{L}}, f(r)\rangle \approx \langle [r]_{\bar{L}}, R\rangle$.*
Here $\approx$ can refer to perfect, statistical or computational indistinguishability.

Definition 3. *A polynomial time function $f : \{0,1\}^n \to \{0,1\}^k$ is* ℓ-RF (resilient function) *if for any $L \in \{{n \atop \ell}\}$, for any assignment $w \in \{0,1\}^{n-\ell}$ to the positions* not *in L, for a randomly chosen $r \in \{0,1\}^n$ subject to $[r]_{\bar{L}} = w$ and random $R \in \{0,1\}^k$, we have: $\langle f(r) \mid [r]_{\bar{L}} = w\rangle \approx \langle R\rangle$.*
Here $\approx$ can refer to perfect, statistical or computational indistinguishability.

[4] Notice, for $L \in \{{s \atop \ell}\}$ we have notationally that $[(y_1, y_2)]_{\bar{L}} = ([y_1]_{\bar{L}}, y_2)$.

As an obvious note, a ℓ-RF is also a static ℓ-ERF (as we shall see, this will *no longer hold* for adaptive ERF; see Lemma 5).

Perfect primitives. It is clear that perfect ERF are the same as perfect RF. Additionally, perfect AONT's are easy to construct from perfect ERF's. In particular one could use the simple one-time pad construction of [10]: $T(x) = \langle r, f(r) \oplus x \rangle$, where r is the secret part of the AONT. However, we observe that (ignoring the issue of efficiency) there is no need for the public part in the perfect AONT (i.e., we can fix it to any valid setting y_2 and consider the restriction of the AONT where the public part is always y_2). Setting $y_2 = \mathbf{0}$ in the one-time pad construction implies an AONT where we output a random r subject to $f(r) = x$. Thus, in the perfect setting the "inverse" of an ℓ-ERF *is* an ℓ-AONT, and we get:

Lemma 1. *(Ignoring issues of efficiency) A perfect ℓ-ERF $f : \{0,1\}^n \rightarrow \{0,1\}^k$ implies the existence of a perfect (secret-only) ℓ-AONT $T : \{0,1\}^k \rightarrow \{0,1\}^n$.*

While the reduction above does *not* work with statistical ERF (to produce statistical AONT), we will show that it works with a stronger notion of *almost-perfect* RF (to produce statistical AONT). See Lemma 7.

2.2 Definitions for Adaptive Adversaries

Adaptively Secure AONT. In the ordinary AONT's the adversary has to "decide in advance" which $(s - \ell)$ bits of the (secret part of) the output it is going to observe. This is captured by requiring the security for all *fixed* sets L of cardinality ℓ. While interesting and non-trivial to achieve, in many applications (e.g. partial key exposure, secret sharing, protecting against exhaustive key search, etc.) the adversary potentially has the power to choose which bits to observe *adaptively*. For example, at the very least it is natural to assume that the adversary could decide which bits of the secret part to observe after it learns the public part. Unfortunately, the constructions of [10] do not even achieve this minimal adaptive security, invalidating their claim that "public part requires no protection and can be given away for free". More generally, the choice of which bit(s) to observe next may partially depend on which bits the adversary has already seen. Taken to the most extreme, we can allow the adaptive adversary to read the bits of the secret part "one-bit-at-a-time", as long as he misses at least ℓ of them.

Definition 4. *A polynomial time randomized transformation $T : \{0,1\}^k \rightarrow \{0,1\}^s \times \{0,1\}^p$ is a (perfect, statistical or computational)* adaptive ℓ-AONT (adaptive all-or-nothing transform) *if*

1. *T is efficiently invertible, i.e. there is a polynomial time machine I such that for any $x \in \{0,1\}^k$ and any $y = (y_1, y_2) \in T(x)$, we have $I(y) = x$.*
2. *For any adversary $\mathcal{A}$ who has oracle access to string $y = (y_s, y_p)$ and is required not to read at least ℓ bits of y_s, and for any $x_0, x_1 \in \{0,1\}^k$, we have: $\left|\Pr(\mathcal{A}^{T(x_0)}(x_0, x_1) = 1) - \Pr(\mathcal{A}^{T(x_1)}(x_0, x_1) = 1)\right| \le \epsilon$, where*

- *In the perfect setting $\epsilon = 0$.*
- *In the statistical setting $\epsilon = negl(s + p)$.*
- *In the computational setting $\epsilon = negl(s + p)$ for any PPT $\mathcal{A}$.*

We stress that the adversary can base its queries on x_0, x_1, the public part of the output, as well as those parts of the secret output that it has seen so far. We also remark that in the perfect setting this definition is *equivalent* to that of an ordinary perfect ℓ-AONT. Thus, adaptivity does not help the adversary in the perfect setting (because the definition of a perfect AONT is by itself very strong!). In particular, good perfect AONT's are good adaptive AONT's. Unfortunately, we will later show that very good perfect AONT's do not exist.

Adaptively Secure ERF. In the original definition of ERF [10], the adversary has to "decide in advance" which $(n - \ell)$ input bits it is going to observe. This is captured by requiring the security for all *fixed* sets L of cardinality ℓ. However, in many situations (e.g., the problem of gradual key exposure [10]), the adversary has more power. Namely, it can decide which $(n - \ell)$ bits of the secret to learn *adaptively* based on the information that it has learned so far. In the most extreme case, the adversary would decide which bits to observe "one-bit-at-a-time". Unfortunately, the definition and the construction of [10] do not satisfy this notion.

There is one more particularity of adaptive security for ERF's. Namely, in some applications (like the construction of AONT's using ERF's [10]) the adversary might observe some partial information about the secret output of the ERF, $f(r)$, *before it starts to compromise the input* r. Is it acceptable in this case that the adversary can learn more partial information about $f(r)$ than he already has? For example, assume we use $f(r)$ as a stream cipher and the adversary learns the first few bits of $f(r)$ before it chooses which $(n - \ell)$ bits of r to read. Ideally, we will not want the adversary to be able to learn some information about the remaining bits of $f(r)$ — the ones that would be used in the stream cipher in the future. Taken to the extreme, even if the adversary sees either the *entire* $f(r)$ (i.e., has complete information on $f(r)$), or a random R, and *only then* decides which $(n - \ell)$ bits of r to read, it cannot distinguish the above two cases.

As we argued, we believe that a good notion of adaptive ERF should satisfy both of the properties above, which leads us to the following notion.

Definition 5. *A polynomial time function $f : \{0,1\}^n \to \{0,1\}^k$ is a (perfect, statistical or computational)* adaptive ℓ-ERF (adaptive exposure-resilient function) *if for any adversary $\mathcal{A}$ who has access to a string r and is required not to read at least ℓ bits of r, when r is chosen at random from $\{0,1\}^n$ and R is chosen at random from $\{0,1\}^k$, we have:* $|\Pr(\mathcal{A}^r(f(r)) = 1) - \Pr(\mathcal{A}^r(R) = 1)| \le \epsilon$*, where*

- *In the perfect setting $\epsilon = 0$.*
- *In the statistical setting $\epsilon = negl(n)$.*
- *In the computational setting $\epsilon = negl(n)$ for any PPT $\mathcal{A}$.*

Notice that in the perfect setting this definition is equivalent to that of an ordinary (static) perfect ℓ-ERF, since for any L, the values $[r]_{\bar{L}}$ and $f(r)$ are uniform and independent. In the statistical setting, the notions are no longer equivalent: indeed, the original constructions of [10] fail dramatically under an adaptive attack. We briefly mention the reason. They used so-called randomness extractors in their construction of statistical ERF's (see [10] for the definitions). Such extractors use a small number of truly random bits d to extract all the randomness from any "reasonable" distribution X. However, it is crucial that this randomness d is chosen independently from and *after* the distribution X is specified. In their construction d was part of the input r, and reading upto $(n-\ell)$ of the remaining bits of r defined the distribution X that they extracted randomness from. Unfortunately, an adaptive adversary can first read d, and only then determine which other bits of r to read. This alters X depending on d, and the notion of an extractor does not work in such a scenario. In fact, tracing the particular extractors that they use, learning d first indeed allows an adaptive adversary to break the resulting static ERF.

Also notice that once we have good adaptive statistical ERF's, adaptive computational ERF's will be easy to construct in same same way as with regular ERF [10]: simply apply a good pseudorandom generator to the output of an adaptive statistical ERF. Finally, we notice that the generic one-time pad construction of [10] of AONT's from ERF's extends to the adaptive setting, as long as we use the strong adaptive definition of ERF given above. Namely, the challenge has to be given first, since the adversary for the AONT may choose which bits of the secret part r to read when having already read the entire public part — either $f(r) \oplus x_0$ or $f(r) \oplus x_1$ (for known x_0 and x_1!). Thus, we get

Lemma 2. *If $f : \{0,1\}^n \rightarrow \{0,1\}^k$ is an adaptive ℓ-ERF, then $T(x) = \langle r, x \oplus f(r) \rangle$ is an adaptive ℓ-AONT with secret part r and public part $x \oplus f(r)$.*

3 Lower Bound on Perfect AONT

In this section we study perfect AONT's. We show that there exists a strong limitation in constructing perfect AONT's: the adversary must miss at least half of the n-bit output, even if the input size k is as small as $\log n$. Recall that perfect AONT's are more general than perfect ERF's (Lemma 1), and thus our bound non-trivially generalizes the lower bound of Friedman [15] (see also another proof by [4]) on perfect ERF. As we will see, the proof will follow from the impossibility of certain weighted "balanced" colorings of an n-dimensional hypercube, which is of independent interest.

Theorem 1. *If $T : \{0,1\}^k \rightarrow \{0,1\}^n$ is a perfect (secret-only) ℓ-AONT, then*

$$\ell \geq 1 + n \cdot \frac{2^{k-1}-1}{2^k-1} = \frac{n}{2} + \left(1 - \frac{n}{2(2^k-1)}\right) \tag{1}$$

In particular, for $n \leq 2^k$ we get $\ell > \frac{n}{2}$, so at least half of the output of T has to remain secret even if T exponentially expands its input! Moreover, the equality can be achieved only by AONT's constructed from ERF's via Lemma 1.

3.1 Balanced Colorings of the Hypercube

A *coloring* of the n-dimensional hypercube $\mathcal{H} = \{0,1\}^n$ with c colors is any map which associates a color from $\{1, \ldots, c\}$ to each node in the graph. In a *weighted* coloring, each node y is also assigned a non-negative real weight $\chi(y)$. We will often call the nodes of weight 0 *uncolored*, despite them having an assigned nominal color. For each color i, we define the weight vector χ_i of this color by assigning $\chi_i(y) = \chi(y)$ if y has color i, and 0 otherwise. We notice that for any given $y \in \mathcal{H}$, $\chi_i(y) > 0$ for at most one color i, and also $\sum \chi_i = \chi$. A coloring where all the nodes are uncolored is called *empty.* Since we will never talk about such colorings, we will assume that $\sum_{y \in \mathcal{H}} \chi(y) = 1$. A *uniform* coloring has all the weights equal: $\chi(y) = 2^{-n}$ for all y.

An *ℓ-dimensional subcube $\mathcal{H}_{L,a}$* of the hypercube is given by a set of ℓ "free" positions $L \in \{ {n \atop \ell} \}$ and an assignment $a \in \{0,1\}^{n-\ell}$ to the remaining positions, and contains the resulting 2^ℓ nodes of the hypercube consistent with a.

Definition 6. *We say a weighted coloring of the hypercube is ℓ-*balanced *if, within every subcube of dimension ℓ, each color has the same weight. That is, for each L and a, $\sum_{y \in \mathcal{H}_{L,a}} \chi_i(y)$ is the same for all colors i.*

Notice, ℓ-balanced coloring is also ℓ'-balanced for any $\ell' > \ell$, since an ℓ' dimensional subcube is the disjoint union of ℓ-dimensional ones. We study balanced colorings since they exactly capture the combinatorial properties of ℓ-AONT's and ℓ-ERF's. We get the following equivalences.

Lemma 3. *Ignoring efficiency, the following equivalences hold in the perfect setting:*

1. *ℓ-AONT's from k to n bits* $\Longleftrightarrow$ weighted *ℓ-balanced colorings of n-dimensional hypercube with 2^k colors.*
2. *ℓ-ERF's from n to k bits* $\Longleftrightarrow$ uniform *ℓ-balanced colorings of n-dimensional hypercube with 2^k colors.*

Proof Sketch. For the first equivalence, the color of node $y \in \mathcal{H}$ corresponds to the value if the inverse map $I(y)$, and its weight corresponds to $\Pr_{x,T}(T(x) = y)$. For the second equivalence, the color of node $y \in \mathcal{H}$ is simply $f(y)$. □

Notice, the lemma above also gives more insight into why perfect AONT's are more general than perfect ERF's (and an alternative proof of Lemma 1). We now restate our lower bound on perfect AONT's in Theorem 1 in terms of weighted ℓ-balanced colorings of $\mathcal{H}$ with $c = 2^k$ colors (proving it for general c).

Theorem 2. *Any (non-empty) ℓ-balanced weighted coloring of the n-dimensional hypercube using c colors must have $\ell \geq \frac{n}{2} + \left(1 - \frac{n}{2(c-1)}\right)$. Moreover, equality can hold only if the coloring is uniform and no two adjacent nodes of positive weight have the same color.*

We believe that the theorem above is interesting in its own right. It says that once the number of colors is at least 3, it is impossible to find a c-coloring (even weighted!) of the hypercube such that all ℓ-dimensional subcubes are "equicolored", unless ℓ is very large (linear in n).

3.2 Proof of the Lower Bound (Theorem 2)

In our proof of Theorem 2, we will consider the 2^n-dimensional vector space V consisting of real-valued (not boolean!) vectors with positions indexed by the strings in $\mathcal{H}$, and we will use facts about the Fourier decomposition of the hypercube.

FOURIER DECOMPOSITION OF THE HYPERCUBE. Like the original proof of Friedman [15] for the case of uniform colorings, we use the adjacency matrix A of the hypercube. A is a $2^n \times 2^n$ dimensional 0-1 matrix, where the entry $A_{x,y} = 1$ iff x and y (both in $\{0,1\}^n$) differ in exactly one coordinate. Recall that a non-zero vector $\mathbf{v}$ is an *eigenvector* of the matrix A corresponding to an *eigenvalue* λ, if $A\mathbf{v} = \lambda\mathbf{v}$. Since A is symmetric, there is an orthonormal basis of $\mathbb{R}^{2^n}$ in which all 2^n vectors are eigenvectors of A. For two strings in x, z in $\{0,1\}^n$, let $x \cdot z$ denote their inner product modulo 2 and let $weight(z)$ be the number of positions of z which are equal to 1. Then:

Fact 1 *A has an orthonormal basis of eigenvectors* $\{\mathbf{v}_z : z \in \{0,1\}^n\}$*, where the eigenvalue of* $\mathbf{v}_z$ *is* $\lambda_z = n - 2 \cdot weight(z)$*, and the value of* $\mathbf{v}_z$ *at position* y *is* $\mathbf{v}_z(y) = \frac{1}{\sqrt{2^n}} \cdot (-1)^{z \cdot y}$.

We will use the notation $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^\top \mathbf{v} = \sum_i u_i v_i$ to denote the inner product of $\mathbf{u}$ and $\mathbf{v}$, and let $\|\mathbf{u}\|^2 = \langle \mathbf{u}, \mathbf{u} \rangle = \sum_i u_i^2$ denote the square of the Euclidean norm of $\mathbf{u}$. We then get the following useful fact, which follows as an easy exercise from Fact 1 (it is also a consequence of the Courant-Fischer inequality).

Fact 2 *Assume* $\{\mathbf{v}_z : z \in \{0,1\}^n\}$ *are the eigenvectors of A as above, and let* $\mathbf{u}$ *be a vector orthogonal to all the* $\mathbf{v}_z$*'s corresponding to z with* $weight(z) < t$: $\langle \mathbf{u}, \mathbf{v}_z \rangle = 0$*. Then we have:* $\mathbf{u}^\top A \mathbf{u} \le (n - 2t) \cdot \|\mathbf{u}\|^2$*. In particular, for* any $\mathbf{u}$ *we have:* $\mathbf{u}^\top A \mathbf{u} \le n \cdot \|\mathbf{u}\|^2$.

EXPLOITING BALANCEDNESS. Consider a non-empty ℓ-balanced weighted coloring χ of the hypercube using c colors. Let χ_i be the characteristic weight vector corresponding to color i (i.e. $\chi_i(y)$ is the weight of y when y has color i and 0 otherwise). As we will show, the χ_i's have some nice properties which capture the balancedness of the coloring χ. In particular, we know that for any colors i and j and for any ℓ-dimensional subcube of $\mathcal{H}$, the sum of the components of χ_i and of χ_j are the same in this subcube. Hence, if we consider the difference $(\chi_i - \chi_j)$, we get that the sum of its coordinates over any ℓ-dimensional subcube is 0.

To exploit the latter property analytically, we consider the quantity $(\chi_i - \chi_j)^\top A(\chi_i - \chi_j)$, where A is the adjacency matrix of the n-dimensional hypercube. As suggested by Fact 2, we can bound this quantity by calculating the Fourier coefficients of $(\chi_i - \chi_j)$ corresponding to large eigenvalues. We get:

Lemma 4. *For any* $i \neq j$*, we have:* $(\chi_i - \chi_j)^\top A(\chi_i - \chi_j) \le (2\ell - n - 2) \cdot \|\chi_i - \chi_j\|^2$.

We postpone the proof of this crucial lemma until the the end of this section, and now just use it to prove our theorem. First, note that the lemma above only gives us information on two colors. To simultaneously use the information from all pairs, we consider the *sum* over all pairs i, j, that is

$$\varDelta \stackrel{\text{def}}{=} \sum_{i,j} (\chi_i - \chi_j)^\top A (\chi_i - \chi_j) \tag{2}$$

We will give upper and lower bounds for this quantity (Equation (3) and Equation (4), respectively), and use these bounds to prove our theorem. We first give the upper bound, based on Lemma 4.

Claim.

$$\varDelta \leq 2\ (2\ell - n - 2)\ (c-1) \cdot \sum_i \|\chi_i\|^2 \tag{3}$$

Proof. We can ignore the terms of $\varDelta$ when $i = j$ since then $(\chi_i - \chi_j)$ is the 0 vector. Using Lemma 4 we get an upper bound:

$$\sum_{i,j} (\chi_i - \chi_j)^\top A (\chi_i - \chi_j) \leq (2\ell - n - 2) \cdot \sum_{i \neq j} \|\chi_i - \chi_j\|^2$$

Now the vectors χ_i have disjoint supports (since each $y \in \mathcal{H}$ is assigned only one color), so we have $\|\chi_i - \chi_j\|^2 = \|\chi_i\|^2 + \|\chi_j\|^2$. Substituting into the equation above, we see that each $\|\chi_i\|^2$ appears $2(c-1)$ times (recall that c is the number of colors), which immediately gives the desired bound in Equation (3). □

Second, we can expand the definition of $\varDelta$ to directly obtain a lower bound.

Claim.

$$\varDelta \geq -2n \cdot \sum_i \|\chi_i\|^2 \tag{4}$$

Proof. Since A is symmetric we have $\chi_i^\top A \chi_j = \chi_j^\top A \chi_i$. Then:

$$\begin{aligned} \sum_{i,j} (\chi_i - \chi_j)^\top A (\chi_i - \chi_j) &= \sum_{i,j} \left(\chi_i^\top A \chi_i + \chi_j^\top A \chi_j - 2\chi_i^\top A \chi_j \right) \\ &= 2c \cdot \sum_i \chi_i^\top A \chi_i \ - 2 \cdot \sum_{i,j} \chi_i^\top A \chi_j \end{aligned}$$

Let us try to bound this last expression. On the one hand, we know that $\chi_i^\top A \chi_i \geq 0$ since it is a product of matrices and vectors with non-negative entries. On the other hand, we can rewrite the last term as a product:

$$\sum_{i,j} \chi_i^\top A \chi_j = \left(\sum_i \chi_i \right)^\top A \left(\sum_i \chi_i \right)$$

This quantity, however, we can bound using the fact that the maximum eigenvalue of A is n (see Fact 2). We get:

$$\left(\sum_i \chi_i\right)^\top A \left(\sum_i \chi_i\right) \leq n \cdot \left\|\sum_i \chi_i\right\|^2$$

Since the vectors χ_i have disjoint support (again, each node y is assigned a unique color), they are orthogonal and so $\|\sum_i \chi_i\|^2 = \sum_i \|\chi_i\|^2$. Combining these results, we get the desired lower bound:

$$\sum_{i,j}(\chi_i - \chi_j)^\top A(\chi_i - \chi_j) \geq 0 - 2n \cdot \sum_i \|\chi_i\|^2 = -2n \cdot \sum_i \|\chi_i\|^2 \qquad \square$$

Combining the lower and the upper bounds of Equation (3) and Equation (4), we notice that $\sum_i \|\chi_i\|^2 > 0$ and can be cancelled out (since the coloring χ is non-empty). This gives us $2(2\ell - n - 2)(c - 1) \geq -2n$, which exactly implies the needed bound on ℓ.

Proof of Lemma 4. It remains to prove Lemma 4, i.e. $(\chi_i - \chi_j)^\top A(\chi_i - \chi_j) \leq (2\ell - n - 2) \cdot \|\chi_i - \chi_j\|^2$. By Fact 2, it is sufficient show that all the Fourier coefficients of $(\chi_i - \chi_j)$ which correspond to eigenvalues $\lambda_z \geq 2\ell - n = n - 2(n - \ell)$ are 0. In other words, that $(\chi_i - \chi_j)$ is orthogonal to all the eigenvectors $\mathbf{v}_z$ whose eigenvalues are at least $(n - 2(n - \ell))$, i.e. $weight(z) \leq n - \ell$. But recall that by the definition of balancedness, on any subcube of dimension at least ℓ, the components of $(\chi_i - \chi_j)$ sum to 0! On the other hand, the eigenvectors $\mathbf{v}_z$ are constants on very large-dimensional subcubes of $\mathcal{H}$ when λ_z is large (see Fact 1). These two facts turn out to be exactly what we need to in order to show that $\langle \mathbf{v}_z, \chi_i - \chi_j \rangle = 0$ whenever $\lambda_z \geq 2\ell - n$, and thus to prove Lemma 4.

Claim. For any $z \in \{0,1\}^n$ with $weight(z) \leq n - \ell$ (i.e. $\lambda_z \geq 2\ell - n$), we have: $\langle \mathbf{v}_z, \chi_i - \chi_j \rangle = 0$.

Proof. Pick any vector $z = (z_1, \ldots, z_n) \in \{0,1\}^n$ with $weight(z) \leq n - \ell$, and let S be the support of z, i.e. $S = \{j : z_j = 1\}$. Note that $|S| \leq n - \ell$. Also, recall that $\mathbf{v}_z(y) = \frac{1}{\sqrt{2^n}} \cdot (-1)^{z \cdot y}$ (see Fact 1). Now consider any assignment a to the variables of S. By letting the remaining variables take on all possible values, we get some subcube of the hypercube, call it $\mathcal{H}_a$.

One the one hand, note that $\mathbf{v}_z$ is constant (either $1/\sqrt{2^n}$ or $-1/\sqrt{2^n}$) on that subcube, since if y and y' differ only on positions *not* in S, we will have $z \cdot y = z \cdot y'$. Call this value C_a. On the other hand, since the coloring is ℓ-balanced and since $|S| \leq n - \ell$, the subcube $\mathcal{H}_a$ has dimension at least ℓ and so we know that both colors i and j have equal weight on $\mathcal{H}_a$. Thus summing the values of $(\chi_i - \chi_j)$ over this subcube gives 0.

Using the above two observations, we show that $\langle \chi_i - \chi_j, \mathbf{v}_z \rangle = 0$ by rewriting the inner product as a sum over all assignments to the variables in S:

$$\langle \chi_i - \chi_j, \mathbf{v}_z \rangle = \sum_{y \in \mathcal{H}} \mathbf{v}_z(y)[\chi_i(y) - \chi_j(y)] = \sum_{a \in \{0,1\}^{|S|}} \left(\sum_{y \in \mathcal{H}_a} \mathbf{v}_z(y)[\chi_i(y) - \chi_j(y)] \right)$$

$$= \sum_a C_a \cdot \left(\sum_{y \in \mathcal{H}_a} \chi_i(y) - \sum_{y \in \mathcal{H}_a} \chi_j(y) \right) = \sum_a C_a \cdot 0 = 0 \qquad \square$$

Equality Conditions. We now determine the conditions on the colorings so that we can achieve equality in Theorem 2 (and also Theorem 1). Interestingly, such colorings are very structured, as we can see by tracing through our proof. Namely, consider the lower bound proved in Equation (4), i.e. that $\sum_{i,j}(\chi_i - \chi_j)^\top A(\chi_i - \chi_j) \leq -2n \sum_i \|\chi_i\|^2$. Going over the proof, we see that equality can occur only if two conditions occur.

On the one hand, we must have $\chi_i^\top A \chi_i = 0$ for all colors i. An easy calculation shows that $\chi_i^\top A \chi_i$ is 0 only when there is no edge of non-zero weight connecting two nodes of color i. Thus, this condition implies that the coloring is in fact a c-coloring in the traditional sense of complexity theory: no two adjacent nodes will have the same color. On the other hand, the inequality $(\sum_i \chi_i)^\top A(\sum_i \chi_i) \leq n \cdot \| \sum_i \chi_i\|^2$ must be tight. This can only hold if the vector $\chi = \sum_i \chi_i$ is parallel to $(1, 1, \ldots, 1)$ since that is the only eigenvector with the largest eigenvalue n. But this means that all the weights $\chi(y)$ are the same, i.e. that the coloring must be *uniform*.

We also remark that Chor et al. [12] showed (using the Hadamard code) that our bound is tight for $k \leq \log n$.

3.3 Extension to Larger Alphabets

Although the problem of constructing AONT's is usually stated in terms of bits, it is natural in many applications (e.g., secret-sharing) to consider larger alphabets, namely to consider $T : \{0, \ldots, q-1\} \to \{0, \ldots, q-1\}^n$. All the notions from the "binary" case naturally extend to general alphabets as well, and so does our lower bound. However, the lower bound we obtain is mostly interesting when the alphabet size q is relatively small compared to n. In particular, the threshold $n/2$, which is so crucial in the binary case (when we are trying to encode more than $\log n$ bits), becomes n/q (recall, q is the size of the alphabet). Significantly, this threshold becomes meaningless when $q > n$. This isn't surprising, since in this case we can use Shamir's secret sharing [25] (provided q is a prime power) and achieve $\ell = k$. We also remark that our bound is tight if $q^k \leq n$ and can be achieved similarly to the binary case by using the q-ary analog of the Hadamard code.

Theorem 3. *For any integer $q \geq 2$, let $T : \{0, \ldots, q-1\}^k \to \{0, \ldots, q-1\}^n$ be a perfect ℓ-AONT. Then*

$$\ell \geq \frac{n}{q} + \left(1 - \frac{q-1}{q} \cdot \frac{n}{q^k - 1}\right)$$

In particular, $\ell > n/q$ when $q^k > n$.

Similarly to the binary case, there is also a natural connection between ℓ-AONT's and weighted ℓ-balanced colorings of the "multi-grid" $\{0,\ldots,q-1\}^n$ with $c = q^k$ colors. And again, the bound of Theorem 2 extends here as well and becomes $\ell \geq \frac{n}{q} + \left(1 - \frac{q-1}{q} \cdot \frac{n}{c-1}\right)$.

The proof techniques are essentially identical to those for the binary case. We now work with the graph $\{0,\ldots,q-1\}^n$, which has an edge going between every pair of words that differ in a single position. We think of vertices in this graph as vectors in $\mathbb{Z}_q^n$. If ω is a primitive q-th root of unity in $\mathbb{C}$, then a orthonormal basis of eigenvectors of the adjacency matrix is given by the q^n-dimensional complex vectors $\mathbf{v}_z$ for $z \in \{0,\ldots,q-1\}^n$, where $\mathbf{v}_z(y) = \frac{1}{\sqrt{q^n}} \cdot \omega^{z \cdot y}$ (here, $z \cdot y$ is the standard dot product modulo q). Constructing upper and lower bounds as above, we eventually get $(q\ell - n - q)(c-1)\sum_i \|\chi_i\|^2 \geq -n(q-1)\sum_i \|\chi_i\|^2$ which implies the desired inequality. Equality conditions are the same.

4 Adaptive Security in the Statistical Setting

We now address the question of adaptive security in the *statistical* setting. Indeed, we saw that both perfect ERF's and perfect AONT's have strong limitations. We also observed in Lemma 2 that we only need to concentrate on ERF's — we can use them to construct AONT's. Finally, we know that applying a regular pseudorandom generator to a good adaptively secure statistical ERF will result in a good adaptively secure *computational* ERF. This leaves with the need to construct adaptive statistical ERF's (recall that unfortunately, the construction of [10] for the static case is not adaptively secure). Hence, in this section we discuss only the statistical setting, and mainly resilient functions (except for Section 4.3; see below).

More specifically, in Section 4.1 we discuss several flavors of statistical resilient functions, and the relation among them, which should be of independent interest. In particular, we argue that the notion of almost-perfect resilient functions (APRF) [19] is the strongest one (in particular, stronger than adaptive ERF). In Section 4.2 we show how to construct APRF's. While seemingly only slightly weaker than perfect RF's, we show that we can achieve much smaller, optimal resilience for such functions: $\ell \approx k$ (compare with $\ell \geq n/2$ for perfect RF's). In particular, this will imply the existence of nearly optimal statistical RF's and adaptive statistical ERF's with the same parameters. Finally, in Section 4.3 we will show that APRF's can also be used to show the existence of optimal *secret-only* adaptive statistical AONT's (which improves the one-time pad construction from Lemma 2 and was not known even in the non-adaptive setting of [10]).

4.1 Adaptive ERF and Other Flavors of Resilient Functions

The definition presented in section 2 for adaptive security of an ERF is only one of several possible notions of adaptive security. Although it seems right for most applications involving resilience to *exposure*, one can imagine stronger attacks in which the security of resilient functions (RF), which tolerate even partly fixed inputs, would be desired. In this section we relate these various definitions, and reduce them to the stronger notion of an *almost-resilient* function [19], which are of independent combinatorial interest.

There are several parameters which one naturally wants to vary when considering "adaptive" security of an ERF, which is in its essence an extractor for producing good random bits from a partially compromised input.

1. Does the adversary get to see the challenge (output vs. a random string) before deciding how to "compromise" the input?
2. Does the adversary get to decide on input positions to "compromise" one at a time or all at once?
3. Does the adversary get to fix (rather than learn) some of the positions?

Flavors of Resilient Functions. To address the above questions, we lay out the following definitions. Unless stated otherwise, f denotes an efficient function $f : \{0,1\}^n \to \{0,1\}^k$, $L \in \{{n \atop \ell}\}$, r is chosen uniformly from $\{0,1\}^n$, R is chosen uniformly from $\{0,1\}^k$. Finally, the adversary $\mathcal{A}$ is computationally unbounded, and has to obtain a non-negligible advantage in the corresponding experiment.

1. **(Weakly) Static ERF:** (This is the original notion of [10].)
   $r \in \{0,1\}^n$ is chosen at random. The adversary $\mathcal{A}$ specifies L and learns $w = [r]_{\bar{L}}$. $\mathcal{A}$ is then given the challenge Z which is either $f(r)$ or R. $\mathcal{A}$ must distinguish between these two cases.
2. **Strongly Static ERF:** (In this notion, the challenge is given first).
   $r \in \{0,1\}^n$ is chosen at random. The adversary $\mathcal{A}$ is then given the challenge Z which is either $f(r)$ or R. Based on Z, $\mathcal{A}$ specifies L, then learns $w = [r]_{\bar{L}}$, and has to distinguish between $Z = f(r)$ and $Z = R$.
3. **Weakly Adaptive ERF:** (This is a natural notion of adaptivity for ERF.)
   $r \in \{0,1\}^n$ is chosen at random. The adversary $\mathcal{A}$ learns up to $(n-\ell)$ bits of r, one at a time, basing each of his choices on what he has seen so far. $\mathcal{A}$ is then given the challenge Z which is either $f(r)$ or R, and has to distinguish between these two cases.
4. **(Strongly) Adaptive ERF:** (This is the notion defined in Section 2.)
   $r \in \{0,1\}^n$ is chosen at random. The adversary $\mathcal{A}$ is then given the challenge Z which is either $f(r)$ or R. Based on Z, $\mathcal{A}$ learns up to $(n-\ell)$ bits of r, one at a time, and has to distinguish between $Z = f(r)$ and $Z = R$.
5. **Statistical RF:** (This is the extension of resilient functions [12,3] to the statistical model, also defined in Section 2.)
   $\mathcal{A}$ chooses any set $L \in \{{n \atop \ell}\}$ and any $w \in \{0,1\}^{n-\ell}$. $\mathcal{A}$ requests that $[r]_{\bar{L}}$ is set to w. The remaining ℓ bits of r in L are set at random. $\mathcal{A}$ is then given a challenge Z which is either $f(r)$ or R, and has to distinguish between these two cases. (Put another way, $\mathcal{A}$ loses if for any $L \in \{{n \atop \ell}\}$ and any $w \in \{0,1\}^{n-\ell}$, the distribution induced by $f(r)$ when $[r]_{\bar{L}} = w$ and the other ℓ bits of r chosen at random, is statistically close to the uniform on $\{0,1\}^k$.)

6. **Almost-Perfect** RF **(**APRF**):** (This is the notion of [19].)
$\mathcal{A}$ chooses any set $L \in \{{n \atop \ell}\}$ and any $w \in \{0,1\}^{n-\ell}$. $\mathcal{A}$ requests that $[r]_{\bar{L}}$ is set to w. The remaining ℓ bits of r in L are set at random and $Z = f(r)$ is evaluated. $\mathcal{A}$ wins if there exists $y \in \{0,1\}^k$ such that $\Pr(Z = y)$ in this experiment does not lie within $2^{-k}(1 \pm \epsilon)$, where ϵ is negligible.[5]

Note that for each of the first five notions above, we can define the "error parameter" ϵ as the advantage of the adversary in the given experiment (for the sixth notion, ϵ is already explicit).

Let us begin by discussing the notion we started with — adaptive ERF. First, it might seem initially like the notion of weakly adaptive ERF is all that we need. Unfortunately, we have seen that to construct adaptive AONT's from ERF's via Lemma 2, we need strong adaptive ERF's. Second, the "algorithmic" adaptive behavior of the adversary is difficult to deal with, so it seems easier to deal with a more combinatorial notion. For example, one might hope that a statistical RF is by itself an adaptive ERF (notice, such RF is clearly a *static* ERF), and then concentrate on constructing statistical RF's. Unfortunately, this hope is false, as stated in the following lemma.

Lemma 5. *There are functions which are statistical* RF *but not statistical adaptive (or even strongly static!)* ERF.

Proof Sketch. Let n be the input size. Let f' be an statistical RF from $n' = \frac{n}{2}$ bits to $k' = \frac{n}{6}$ bits such that $\ell' = \frac{n}{4}$. Such functions exist, as we prove in Section 4.2.

Define f as follows: on an n-bit input string r, break r into two parts r_1 and r_2 both of length $\frac{n}{2}$. Apply f' to r_1 to get a string s of length $\frac{n}{6}$. Now divide s into $\frac{n}{6(\log n-1)}$ blocks of size $\log \frac{n}{2}$, which can be interpreted as a random subset S from $\{1,\ldots,\frac{n}{2}\}$ with $\frac{n}{6(\log n-1)}$ elements. Let $\bigoplus S$ be the parity of the bits in $[r_2]_S$. The output of f is the pair $\langle s, \bigoplus S\rangle$. Thus $k \approx \frac{n}{6}$.

Now let $\ell = n - \frac{n}{6(\log n-1)}$. Clearly, an adversary who sees the challenge first, can (non-adaptively) read the bits $[r_2]_S$ and check the parity (giving him advantage at least 1/2 over the random string). Thus, f is not an adaptively secure ERF. On the other hand, an adversary who can fix only $(n-\ell) \approx n/6\log(n)$ input bits can still not learn anything about the output of f' and thus is unlikely to know the value of *all* the bits in S. Such an adversary will always have negligible advantage. Hence f is a statistical RF. □

Since the opposite direction (from adaptive ERF's to statistical RF's) is obviously false as well, we ask if some notion actually can *simultaneously* achieve both adaptive security for ERF, and statistical security for RF. Fortunately, it turns that by satisfying the stronger condition of an almost-perfect resilient function (APRF) [19], one obtains an adaptive ERF. Since APRF's will play such a crucial role in our study, we give a separate, more formal definition.

[5] Note that in [19] the error parameter was measured slightly differently: they define ϵ as the maximum absolute deviation. Our convention makes sense in the cryptographic setting since then the adversary's advantage at distinguishing $f(r)$ from random in any of the above experiments is comparable ϵ, as opposed to $\epsilon 2^k$.

Definition 7. *A polynomial time function $f : \{0,1\}^n \to \{0,1\}^k$ is ℓ-*APRF (almost-perfect resilient function) *if for any $L \in \{{n \atop \ell}\}$, for any assignment $w \in \{0,1\}^{n-\ell}$ to the positions* not *in L, for a randomly chosen $r \in \{0,1\}^n$ and for some negligible $\epsilon = negl(n)$, we have:*

$$\Pr(f(r) = y \;\Big|\; [r]_{\bar{L}} = w) = (1 \pm \epsilon)2^{-k} \tag{5}$$

While it is obvious that any APRF is a statistical RF (by summing over 2^k values of y), the fact that it is also an adaptive ERF is less clear (especially considering Lemma 5), and is shown below.

Theorem 4. *If f is an* APRF*, then f is a statistical adaptive* ERF*.*

Proof. By assumption, f is an ℓ-APRF with error ϵ: for every set $L \in \{{n \atop \ell}\}$ and every assignment w to the variables not in L, Equation (5) above holds when r is chosen at random. Now suppose that we have an adaptive adversary $\mathcal{A}$ who, given either $Z = f(r)$ or $Z = R$ and (limited) access to r, can distinguish between the two cases with advantage ϵ'. We will show that $\epsilon' \leq \epsilon$.

At first glance, this may appear trivial: It is tempting to attempt to prove it by conditioning on the adversary's view at the end of the experiment, and concluding that there must be *some* subset L and appropriate fixing w which always leads to a good chance of distinguishing. However, this argument fails since the adversary $\mathcal{A}$ may base his choice of L on the particular challenge he receives, and on the bits he considers.

So we use a more sophisticated argument, although based on a similar intuition. First, we can assume w.l.o.g. that the adversary $\mathcal{A}$ is deterministic, because there is some setting of his random coins conditioned on which he will distinguish with advantage at least ϵ', and so we may as well assume that he always uses those coins.

Following the intuition above, we consider the adversary's view at the end of the experiment, just before he outputs his answer. This view consists of two components: the input challenge Z and the $(n-\ell)$ observed bits $w = w_1, \ldots, w_{n-\ell}$ (which equal $[r]_{\bar{L}}$ for some set L of size at least ℓ). Significantly, L need not be explicitly part of the view: since $\mathcal{A}$ is deterministic, L is a function of Z and w.

Denote by $\mathsf{View}_{\mathcal{A}}^{(Z)}$ the view of $\mathcal{A}$ on challenge Z. When $Z = R$, it is easy to evaluate the probability that $\mathcal{A}$ will get a given view. Since the values $r \in \{0,1\}^n$ and $R \in \{0,1\}^k$ are independent, we have

$$\Pr\left[\mathsf{View}_{\mathcal{A}}^{(R)} = (y, w)\right] = 2^{-(n-\ell+k)}$$

On the other hand, when $Z = f(r)$, we have to be careful. If L is the subset corresponding to $\mathcal{A}$'s choices on view (y, w), then we do indeed have:

$$\Pr\left[\mathsf{View}_{\mathcal{A}}^{(f(r))} = (y, w)\right] = \Pr\left[f(r) = y \wedge [r]_{\bar{L}} = w\right]$$

This last equality holds even though the choice of L may depend on y. Indeed, $\mathcal{A}$ is deterministic and so he will always choose the subset L when $[r]_{\bar{L}} = w$, regardless of the other values in r. *Thus, we can in some sense remove the adversary from the discussion entirely.* Now this last probability can be approximated by conditioning and using Equation (5):

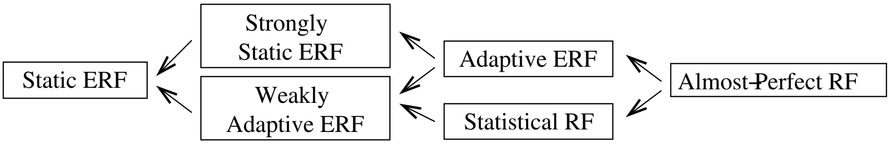
$$\begin{aligned}\Pr\Big[f(r) = y \wedge [r]_{\bar{L}} = w\Big] &= \Pr\Big[f(r) = y \;\Big|\; w = [r]_{\bar{L}}\Big] \Pr\Big[w = [r]_{\bar{L}}\Big] \\ &= (1 \pm \epsilon)2^{-k} \cdot 2^{-(n-\ell)} \\ &= (1 \pm \epsilon)2^{-(n-\ell+k)}\end{aligned}$$

We can now explicitly compute the adversary's probability of success in each of the two experiments we are comparing. Let $A(y, w) = 1$ if $\mathcal{A}$ accepts on view (y, w) and 0 otherwise. Then:

$$\begin{aligned}\epsilon' &= \left|\Pr\Big[\mathcal{A}^r(f(r)) = 1\Big] - \Pr\Big[\mathcal{A}^r(R) = 1\Big]\right| \\ &= \left|\sum_{y,w}\left(\Pr\Big[\mathsf{View}_{\mathcal{A}}^{(f(r))} = (y, w)\Big] - \Pr\Big[\mathsf{View}_{\mathcal{A}}^{(R)} = (y, w)\Big]\right) \cdot A(y, w)\right| \\ &\le \sum_{y,w}\left|(1 \pm \epsilon)2^{-(n-\ell+k)} - 2^{-(n-\ell+k)}\right| \le \epsilon\end{aligned}$$

Thus $\epsilon' \le \epsilon$, and so f is a statistical adaptive ERF. □

Classification of Resilient Functions. In fact, we can completely relate all the six notions of resilient functions that we introduced:



This diagram is complete: if there is no path from notion A to notion B, then there is a function which satisfies A but not B. We notice that except for the two proofs above, only one non-trivial proof is needed in order to complete the diagram: the separation between weakly adaptive ERF's and static ERF's (other implications and separations are easy exercises). However, this separation follows from the static construction of Canetti et al. [10], which, as we mentioned, need not yield a weakly adaptive ERF.

We also remark that while the diagram above is useful from a structural point of view, in the next section we show how to build APRF's — the strongest of the above notions — achieving $l \approx k$, which is nearly optimal even for *static* ERF's — the weakest of the above notions. Thus, all the above notions are almost the "same" in terms of the optimal parameters they achieve (which are also substantially better than those possible in the perfect setting).

4.2 Obtaining Nearly Optimal Almost-Resilient Functions

Given the discussion of the previous section, it is natural to try to construct good APRF's. These were first defined and studied by Kurosawa et al. [19]. Using techniques from coding theory, they construct[6] ℓ-APRF such that $\ell \geq \frac{n+k}{2} + 2\log\left(\frac{1}{\epsilon}\right)$. Although this beats the lower bound on perfect ERF of [15,4], it is very far from the trivial lower bound $\ell \geq k$, especially when $k = o(n)$. Thus, it is natural to ask whether this is a fundamental limitation on APRF's, or whether indeed one can approach this simplistic lower bound.

As a first step, we can show that if f is picked at random from all the functions from $\{0,1\}^n$ to $\{0,1\}^k$, it is very likely to be a good APRF (we omit the proof since we subsume it later). However, this result is of little practical value: storing such a function requires $k \cdot 2^n$ bits. Instead, we replace the random function with a function from a t-wise independent hash family [11] for t roughly on the order of n. Functions in some such families (e.g., the set of all degree $t-1$ polynomials over the field $GF(2^n)$) require as little as tn bits of storage, and are easy to evaluate.

Using tail-bounds for t-wise independent random variables, one can show that with very high probability we will obtain a good APRF:

Theorem 5. *Fix any n, ℓ and ϵ. Let $\mathcal{F}$ be a family of t-wise independent functions from n bits to k bits, where $t = n/\log n$ and*

$$k = \ell - 2\log\left(\frac{1}{\epsilon}\right) - O(\log n)$$

Then with probability at least $(1 - 2^{-n})$ a random function f sampled from $\mathcal{F}$ will be an ℓ-APRF (and hence adaptive ℓ-ERF and statistical ℓ-RF) with error ϵ.

Corollary 1. *For any $\ell = \omega(\log n)$, there exists an efficient statistical adaptive ℓ-ERF $f : \{0,1\}^n \to \{0,1\}^k$ with $k = \ell - o(\ell)$.*

The proof of Theorem 5 uses the following lemma, which is used (implicitly) in the constructions of deterministic extractors of [27]. Recall that a distribution X over $\{0,1\}^n$ has *min-entropy m*, if for all x, $\Pr(X = x) \leq 2^{-m}$.

Lemma 6. *Let $\mathcal{F}$ be a family of t-wise independent functions (for even $t \geq 8$) from n to k bits, let X be a distribution over $\{0,1\}^n$ of min-entropy m, and let $y \in \{0,1\}^k$. Assume for some $\alpha > 0$*

$$k \leq m - \left(2\log\frac{1}{\epsilon} + \log t + 2\alpha\right). \qquad (6)$$

Let f be chosen at random from $\mathcal{F}$ and x be chosen according to X. Then

$$\Pr_{f \in \mathcal{F}}\left(\left|\Pr_x(f(x) = y) - \frac{1}{2^k}\right| \geq \epsilon \cdot \frac{1}{2^k}\right) \leq 2^{-\alpha t} \qquad (7)$$

[6] This result *looks* (but is not) different from the one stated in [19] since we measure ϵ differently.

In other words, for any $y \in \{0,1\}^k$, if f is chosen from $\mathcal{F}$ then with overwhelming probability we have that the probability that $f(X) = y$ is $\frac{1}{2^k}(1 \pm \epsilon)$.

Theorem 5 follows trivially from this lemma. Indeed, set $\alpha = 3\log n$, $t = n/\log n$. Notice that for any $L \in \{{n \atop \ell}\}$ and any setting w of bits not in L, the random variable $X = \langle r \mid [r]_{\bar{L}} = w\rangle$ has min-entropy $m = \ell$. Then k given in Theorem 5 indeed satisfies Equation (6). Now we apply Lemma 6 and take the union bound in Equation (7) over all possible fixings of some $(n-\ell)$ input bits, and over all $y \in \{0,1\}^n$. Overall, there are at most $\binom{n}{\ell}2^{n-\ell}2^k \leq 2^{2n}$ terms in the union bound, and each is less than $2^{-\alpha t} = 2^{-3n}$, finishing the proof of Theorem 5.

For completeness, we give a simple proof of Lemma 6. We will make use of the following "tail inequality" for sums of t-wise independent random variables proven by Bellare and Rompel [2]. There they estimate $\Pr[|Y - \mathsf{Exp}[Y]| > A]$, where Y is a sum of t-wise independent variables. We will only be interested in $A = \epsilon \cdot \mathsf{Exp}[Y]$, where $\epsilon \leq 1$. In this case, tracing the proof of Lemma 2.3 (and Lemma A.5 that is used to prove it) of [2], we get the following:

Theorem 6 ([2]). *Let t be an even integer, and assume $Y_1, \ldots, Y_N$ are t-wise independent random variables in the interval $[0,1]$. Let $Y = Y_1 + \ldots + Y_N$, $\mu = \mathsf{Exp}[Y]$ and $\epsilon < 1$. Then*

$$\Pr(|Y - \mu| \geq \epsilon\mu) \leq C_t \cdot \left(\frac{t}{\epsilon^2\mu}\right)^{t/2} \tag{8}$$

where the constant $C_t < 3$ and in fact $C_t < 1$ for $t \geq 8$.

Now we can prove Lemma 6:

Proof. Let p_x denote the probability that $X = x$, and let q denote the random variable (only over the choice of f) which equals to the probability (over the choice of x *given* f) that $f(x) = y$, i.e.

$$q = \sum_{x \in \{0,1\}^n} p_x \cdot I_{\{f(x)=y\}}$$

where $I_{\{f(x)=y\}}$ is an indicator variable which is 1 if $f(x) = y$ and 0 otherwise. Since for any x the value of $f(x)$ is uniform over $\{0,1\}^k$, we get that $\mathsf{Exp}_f[I_{\{f(x)=y\}}] = 2^{-k}$, and thus $\mathsf{Exp}_f[q] = 2^{-k}$. Notice also that the variables $I_{\{f(x)=y\}}$ are t-wise independent, since f is chosen at random from a family of t-wise independent functions. And finally notice that since X has min-entropy m, we have that all $p_x \leq 2^{-m}$.

Thus, if we let $Q_x = 2^m \cdot p_x \cdot I_{\{f(x)=y\}}$, and $Q = \sum_{x \in \{0,1\}^n} Q_x = 2^m q$, we get that the variables Q_x are t-wise independent, all reside in the interval $[0,1]$, and $\mathsf{Exp}[Q] = 2^m \mathsf{Exp}[q] = 2^{m-k}$. Now we can apply the tail inequality given in Theorem 6 and obtain:

$$\begin{aligned}\Pr_f\left[\left|q-\frac{1}{2^k}\right|\geq\epsilon\cdot\frac{1}{2^k}\right]&=\Pr_f\left[\left|Q-2^{m-k}\right|\geq\epsilon\cdot 2^{m-k}\right]\\&\leq\left(\frac{t}{\epsilon^2\cdot 2^{m-k}}\right)^{t/2}=\left(\frac{1}{2^{m-k-2\log\frac{1}{\epsilon}-\log t}}\right)^{t/2}\\&\leq 2^{-\alpha t}\end{aligned}$$

where the last inequality follows from Equation (6). □

4.3 Adaptively Secure AONT

We already remarked that that the construction of optimal adaptive statistical ERF's implies the construction of adaptive computational ERF's. Combined with Lemma 2, we get optimal constructions of AONT's as well. We notice also that the public part of these AONT construction is k. In the statistical setting, where we achieved optimal $\ell = k+o(k)$, we could then combine the public and the secret part of the AONT to obtain a *secret-only* adaptive AONT with $\ell = 2k+o(k)$. One may wonder if there exist statistical secret-only AONT's with $\ell = k+o(k)$, which would be optimal as well. Using our construction of almost-perfect resilient functions, we give an affirmative answer to this question. Our construction is not efficient, but the existential result is interesting because it was not known even in the static setting.

Lemma 7. *Ignoring the issue of efficiency, there exist adaptive statistical secret-only ℓ-AONT $T:\{0,1\}^k\to\{0,1\}^n$ with $\ell = k+o(k)$.*

Proof. Recall, Lemma 1 used an inverse of a perfect RF (or ERF, which is the same) to construct perfect secret-only AONT. We now show that the same construction can be made to work in the statistical setting provided we use APRF rather than weaker statistical RF. In particular, let $f:\{0,1\}^n\to\{0,1\}^k$ be an ℓ-APRF. We know that we can achieve $\ell = k+o(k)$. We define $T(x)$ to be a random $r\in\{0,1\}^n$ such that $f(r)=x$. (This is well-defined since APRF's are surjective.)

Now take any distingusher $\mathcal{A}$, any $x\in\{0,1\}^k$ and any possible view of $\mathcal{A}$ having oracle access to $T(x)=r$. Since we can assume that $\mathcal{A}$ is deterministic, this view can be specified by the $(n-\ell)$ values w that $\mathcal{A}$ read from r (in particular, the subset L is also determined from w). Now, we use Bayes law to estimate $\Pr(\mathsf{View}_{\mathcal{A}}^{(T(x))}=w)$. Notice, since $r=T(x)$ is a random preimage of x, we could assume that r was chosen at random from $\{0,1\}^n$, and use conditioning on $f(r)=x$. This gives us:

$$\begin{aligned}\Pr(\mathsf{View}_{\mathcal{A}}^{(T(x))}=w)&=\Pr(\mathsf{View}_{\mathcal{A}}^{(r)}=w\ \Big|\ f(r)=x)\ =\ \Pr([r]_{\bar{L}}=w\ \Big|\ f(r)=x)\\&=\frac{\Pr(f(r)=x\ \Big|\ [r]_{\bar{L}}=w)\cdot\Pr([r]_{\bar{L}}=w)}{\Pr(f(r)=x)}\\&=\frac{(1\pm\epsilon)\cdot 2^{-k}\cdot 2^{\ell-n}}{(1\pm\epsilon)\cdot 2^{-k}}\ =\ (1\pm 2\epsilon)\cdot 2^{\ell-n}\end{aligned}$$

Notice that this bound is independent on $\mathcal{A}$, x and w. Hence, for any x_0, x_1 and any adversary $\mathcal{A}$, $\mathsf{View}_{\mathcal{A}}^{(T(x_0))}$ and $\mathsf{View}_{\mathcal{A}}^{(T(x_1))}$ are within statistical distance 4ϵ from each other, implying that T is an adaptive statistical AONT. □

Acknowledgments

It is our pleasure to thank Dan Spielman for help in simplifying the proof of Theorem 2, and Salil Vadhan for extremely helpful discussions and for showing us an early version of [27].

References

1. M. Bellare and A. Boldyreva. The Security of Chaffing and Winnowing. In *Proc. of Asiacrypt*, 2000.
2. M. Bellare, J. Rompel. Randomness-Efficient Oblivious Sampling. In *Proc. of 35th FOCS*, pp. 276–287, 1994.
3. C. Benett, G. Brassard, J. Robert. Privacy Amplification by public discussion. In *SIAM J. on Computing*, pp. 17(2):210–229, 1988.
4. J. Bierbrauer, K. Gopalakrishnan, D. Stinson. Orthogonal Arrays, resilient functions, error-correcting codes and linear programming bounds. In *SIAM J. of Discrete Math*, 9:424–452, 1996.
5. J. Bierbrauer, H. Schellwat. Almost Independent and Weakly Biased Arrys: Efficient Constructions and Cryptologic Applications. In *Proc. of CRYPTO*, pp. 531–543, 2000.
6. G. R. Blakley and C. Meadows. Security of Ramp Schemes. In *Proc. of CRYPTO*, pp. 242–268, 1984.
7. M. Blaze. High Bandwidth Encryption with low-bandwidth smartcards. In *Fast Software Encryption*, pp. 33–40, 1996.
8. C. Blundo, A. De Santis, U. Vaccaro. Efficient Sharing of Many Secrets. In *Proc. of STACS*, LNCS 665, pp. 692-703, 1993.
9. V. Boyko. On the Security Properties of the OAEP as an All-or-Nothing Transform. In *Proc. of Crypto*, pp. 503–518, 1999.
10. R. Canetti, Y. Dodis, S. Halevi, E. Kushilevitz and A. Sahai. Exposure-Resilient Functions and All-Or-Nothing Transforms. In *Proc. of EuroCrypt*, 2000.
11. L. Carter and M. Wegman. Universal classes of hash functions. *JCSS*, vol. 18, pp. 143–154, 1979.
12. B. Chor, J. Friedman, O. Goldreich, J. Håstad, S. Rudich, R. Smolensky. The Bit Extraction Problem or t-resilient Functions. In *Proc. of FOCS*, pp. 396–407, 1985.
13. Y. Dodis. Exposure-Resilient Cryptography. Ph.D. Thesis., MIT, 2000.
14. A. Desai. The security of All-or-Nothing Encryption: Protecting Against Exhaustive Key Search In *Proc. of CRYPTO*, pp. 359–375, 2000.
15. J. Friedman. On the Bit Extraction Problem In *Proc. of FOCS*, pp. 314–319, 1992.
16. O. Goldreich. Foundations of Cryptography (Fragments of a Book). URL: `http://www.wisdom.weizmann.ac.il/home/oded/public_html/frag.html`
17. W.-A. Jackson and K. Martin. A Combinatorial Interpretation of Ramp Schemes *Australasian Journal of Combinatorics*, **14** (1996), 51–60.

18. M. Jakobsson, J. Stern, M. Yung. Scramble All, Encrypt Small. In *Proc. of Fast Software Encryption*, pp. 95–111, 1999.
19. K. Kurosawa, T. Johansson and D. Stinson. Almost k-wise independent sample spaces and their cryptologic applications. Submitted to *J. of Cryptology*, preliminary version appeared in *Proc. of EuroCrypt*, pp. 409–421, 1997.
20. S. Matyas, M. Peyravian and A. Roginsky. Encryption of Long Blocks Using a Short-Block Encryption Procedure. Available at `http://grouper.ieee.org/groups/1363/P1363a/LongBlock.html`.
21. W. Ogata and K. Kurosawa. Some Basic Properties of General Nonperfect Secret Sharing Schemes. *Journal of Universal Computer Science*, Vol.**4**, No. 8 (1998), pp. 690–704.
22. L. H. Ozarow and A. D. Wyner. Wire-Tap Channel II In *Proc. of EUROCRYPT*, pp. 33–50, 1984.
23. R. Rivest. All-or-Nothing Encryption and the Package Transform. In *Fast Software Encryption, LNCS*, 1267:210–218, 1997.
24. R. Rivest. Chaffing and Winnowing: Confidentiality without Encryption. *CryptoBytes (RSA Laboratories)*, 4(1):12–17, 1998.
25. A. Shamir. How to share a secret. In *Communic. of the ACM*, 22:612-613, 1979.
26. S. U. Shin, K. H. Rhee. Hash functions and the MAC using all-or-nothing property. In *Proc. of Public Key Cryptography, LNCS*, 1560:263–275, 1999.
27. L. Trevisan and S. Vadhan. Extracting randomness from samplable distributions In *Proc. of FOCS*, 2000.
28. U. Vazirani. Towards a Strong Communication Complexity Theory or Generating Quasi-Random Sequences from Two Communicating Semi-Random Sources. In *Combinatorica*, 7(4):375–392, 1987.