WBE signal set with a corresponding number of oversized users $|\mathcal{O}|$. The values we used for numerical simulations are summarized in the following table.

E_k	$\lvert\mathcal{O}\rvert$	E_k	$\lvert\mathcal{O}\rvert$
k^2	198	$\sqrt{k}$	196
k	198	$1+0.1k$	194

Fig. 4 illustrates that even mild energy disparities yield very poor performance. For instance, the JER floors at $1/2$ for the first two power distributions. Here, the 198 more powerful users are oversized and hence experience a single-user Gaussian channel; by contrast, the two remaining users share the only remaining direction in the signal space, and hence, interfere in such a way that the AEE of the weakest user is equal to zero. Finally, as the energy disparities decrease, there are fewer oversized users, and the floor decreases and is reached for higher E_b/N_0. The JER is still unacceptably high for all practical purposes though. Consider for instance that when $E_k = 1+0.1k$, the JER floors at 0.06 for $E_b/N_0 \geq 35$ dB.

V. Conclusion

This correspondence has characterized exactly the severe limitation of generalized WBE signals under linear MMSE detection. Specifically, we have shown that such signals do not satisfy even the basic requirement that the error rate of every user decreases exponentially as noise vanishes. Our results hold for arbitrary overload, modulation, and received energies. Moreover, when the received powers are equal, the error rate of every user floors. Therefore, it appears that the full benefit of generalized WBE signals can only be leveraged by nonlinear detection. It remains an open problem as to whether there exists a signal set with nonzero symmetric energy under linear MMSE detection (and hence not the generalized WBE set) for overloaded CDMA systems.

References

[1] P. Viswanath and V. Anantharam, "Optimal sequences and sum capacity of synchronous CDMA systems," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1984–1991, Sep. 1999.

[2] M. Rupf and J. L. Massey, "Optimum sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1261–1266, Jul. 1994.

[3] T. Guess, "Optimal sequences for CDMA with decision-feedback receivers," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 886–900, Apr. 2003.

[4] P. Viswanath, V. Anantharam, and D. N. C. Tse, "Optimal sequences, power control, and user capacity of synchronous CDMA systems with linear MMSE receivers," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1968–1983, Sep. 1999.

[5] L. Gao and T. F. Wong, "Power control and spreading sequence allocation in a CDMA forward link," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 105–124, Jan. 2004.

[6] S. Ulukus and R. D. Yates, "Iterative construction of optimum signature sequence sets in synchronous CDMA systems," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1989–1998, Jul. 2001.

[7] P. Anigstein and V. Anantharam, "Ensuring convergence of the MMSE iteration for interference avoidance to the global optimum," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 873–885, Apr. 2003.

[8] C. Rose, S. Ulukus, and R. D. Yates, "Wireless systems and interference avoidance," *IEEE Trans. Wireless Commun.*, vol. 1, no. 3, pp. 415–428, Jul. 2002.

[9] T. Strohmer, R. W. Heath Jr., and A. J. Paulraj, "On the design of optimal spreading sequences for CDMA systems," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2002, pp. 1434–1438.

[10] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[11] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.

[12] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth, and user capacity," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 641–657, Mar. 1999.

[13] A. Kapur and M. K. Varanasi, "Multiuser detection for overloaded CDMA systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1728–1742, Jul. 2003.

[14] M. K. Varanasi, "Decision feedback multiuser detection: A systematic approach," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 219–240, Jan. 1999.

[15] J. L. Massey and T. Mittelholzer, "Welch's bound and sequence sets for code-division multiple-access systems," in *Sequences II, Methods in Communication, Security, and Computer Science*. New York: Springer-Verlag, 1993.

Extracting Randomness from Multiple Independent Sources

Chia-Jung Lee, Chi-Jen Lu, Shi-Chun Tsai, and Wen-Guey Tzeng

***Abstract*—We study the problem of deterministically extracting almost perfect random bits from multiple weakly random sources that are mutually independent. With two independent sources, we have an explicit extractor which can extract a number of random bits that matches the best construction currently known, via the generalized leftover hash lemma. We also extend our construction to extract randomness from more independent sources. One nice feature is that the extractor still works even with all but one source exposed. Finally, we apply our extractor for a cryptographic task in which a group of parties wants to agree on a secret key for group communication over an insecure channel, without using ideal local randomness.**

***Index Terms*—Deterministic extractor, two-sources extractor, multi-sources-extractor, leftover hash lemma.**

I. Introduction

Randomness has become a useful resource in computer science. For several important computational problems, randomized algorithms are simpler, run faster, or use smaller space than the known deterministic ones. In cryptography, randomness is essential for protocols to generate or hide the secret. Hence, how to obtain and manipulate randomness has become an important topic in computer science. However, random sources we have access to are usually imperfect. We say that a source has *min-entropy* k if every string occurs with probability at most 2^{-k}. From such a weakly random source, one would like to extract almost

Manuscript received August 16, 2004; revised January 38, 2005. The work of C.-J. Lee and S.-C. Tsai was supported in part by the National Science Council of Taiwan under Contract NSC-92-2213-E-009-024. The work of C.-J. Lu was supported in part by the National Science Council of Taiwan under Contract NSC–93-2213-E-001-004. The work of W.-G. Tzeng was supported in part by the National Science Council of Taiwan under Contract NSC-93-2213-E-009-009, and in part under a MediaTek Grant, 2003.

C.-J. Lee and S.-C. Tsai are with the Department of Computer Science and Info Engineering, National Chiao-Tung University, Hsinchu 30050, Taiwan, R.O.C. (e-mail: leecj@csie.nctu.edu.tw; sctsai@csie.nctu.edu.tw).

C.-J. Lu is with the Institute of Information Science, Academia Sinica, Taipei, Taiwan, R.O.C. (e-mail: cjlu@iis.sinica.edu.tw).

W.-G. Tzeng is with the Department of Computer and Info Science, National Chiao-Tung University, Hsinchu 30050, Taiwan, R.O.C. (e-mail: tzeng@cis.nctu.edu.tw).

Communicated by T. Johansson, Associate Editor for Complexity and Cryptography.

Digital Object Identifier 10.1109/TIT.2005.847746

perfect randomness, using a procedure called *extractor* [10]. Chor and Goldreich [2] show that one cannot deterministically extract even one bit from a source of length n with min-entropy $b < n$. One way around this is to add an additional short truly random seed to catalyze the extracting process. The goal is to extract as much randomness as possible using a seed as short as possible. This line of research has received much attention during the past decade (see [11] for a nice survey), and an explicit construction has been given recently which is optimal up to constant factors [9].

When the sources have better structures, it becomes possible to have deterministic (seedless) extractors. One example is the bit-fixing source, in which some bits are fixed while others are perfectly random. Kamp and Zuckerman [7] gave a deterministic extractor that can extract $\Omega(n^{2\gamma})$ bits from bit-fixing sources of length n, in which all but $n^{1/2} + \gamma$ bits are fixed in an "oblivious" way. König and Maurer [8] proposed a deterministic extractor that can extract $\log q$ bits from generalized symbol-fixing sources with n independent symbols on $\mathbb{Z}_p$. Another example is when multiple sources are available which are *mutually independent*. With two sources of length n and of min-entropy b_1, b_2, respectively, Graham, and Spencer [5] show implicitly how to extract one random bit with $b_1 \geq (n/2) + \text{poly}\log(n)$ and $b_2 \geq \log n$. For extraction of many bits, the best construction, by Dodis *et al.* [3], is able to extract $b_1 + b_2 + 2 - n - 2\log(1/\epsilon)$ bits which are ϵ-close to random, even with one of the two sources exposed. On the other hand, Barak *et al.* [1] recently showed that the min-entropy rate can be lowered if more, but still a constant number of, independent sources can be used. In particular, for any constant $\delta \in (0, 1)$, they can extract n bits from a constant (depending on δ) number of independent and identical sources of length n with min-entropy δn.

In this correspondence, we also work on deterministic extraction from multiple independent sources. Our first result is a simple extractor for two sources. The number of bits we extract matches the current best result of Dodis *et al.* [3], but both our construction and analysis are considerably simpler. For example, suppose that we want to extract m bits from two weak sources of length n where $m \mid n$. Then their time complexity is $O(mn)$, and ours is $O(n)$. One of our main technical contribution is a generalization of the well-known *leftover hash lemma* [6]. The leftover hash lemma says that if we sample a function h uniformly from a family H of pairwise independent functions and apply it on an input x sampled from a source with enough min-entropy, the output $h(x)$ will look almost like random. This is usually applied in the setting of seeded extractors, in which the perfect random seed is used to sample uniformly from H. We generalize the leftover hash lemma to allow sampling from H according to any distribution with high enough min-entropy. This provides us a way to extract from two independent weakly random sources: one source to sample the input x while the other to sample the function h. We also extend our construction for the case when there are $t \geq 3$ independent sources available.

Our deterministic extractor can extract $k_1 + k_2 + 2 - n - 2\log(1/\epsilon)$ bits, where k_1 and k_2 are the two largest min-entropies of the t sources. It has the following nice features. First, our extractor works as long as two sources have enough min-entropy; it can work even when only two sources contain randomness (thus, with a very low average min-entropy rate). Second, as is in [3], [4], our extractor can still work even with all but one source exposed. Finally, to construct our extractor, we do not need to know beforehand the specific min-entropy of each source.

Finally, we introduce one possible application with strong multisource extractors. We consider the following cryptographic task which generalizes the two-party case in [4]. Suppose a group of parties $P_1, \ldots, P_t$ are together initially and later go far away from each other, and then they want to establish a secret key for group communication over an insecure channel. Can this task be achieved without using ideal local randomness? We give one solution. Initially, these parties share some $\mathcal{X}$ sampled from a weak source when they are together. After departing from each other, each party P_i samples $\mathcal{X}_i$ from his/her own local weak source, and sends it to the others. Once receiving all $\mathcal{X}_i$'s, each party computes the secret key $\text{EXT}(\mathcal{X}, \mathcal{X}_1, \ldots, \mathcal{X}_t)$ using our extractor EXT, which is secure even against an adversary who knows $\mathcal{X}_2, \ldots, \mathcal{X}_t$. This can be augmented with an authentication process to prevent an adversary from impersonating a legitimate party.

II. Preliminaries

Throughout this correspondence, we will use the terms *random variable* and *distribution* interchangeably. All logarithms will have base two. For $n \in \mathbb{N}$, let $\mathcal{U}_n$ denote the uniform distribution over $\{0,1\}^n$. For two random variables $\mathcal{X}, \mathcal{Y}$ over a finite set S, their *statistical distance* is

$$\|\mathcal{X} - \mathcal{Y}\| \equiv (1/2) \sum_{s \in S} |\Pr[\mathcal{X} = s] - \Pr[\mathcal{Y} = s]|$$

and the min-entropy of $\mathcal{X}$ is

$$\mathrm{H}_\infty(\mathcal{X}) \equiv \min_{s \in S} \log(1/\Pr[\mathcal{X} = s]).$$

For a sequence of values $v_1, \ldots, v_t$, let $v_{[i,j]}$, for $i \leq j$, denote the subsequence $v_i, \ldots, v_j$. In this correspondence, we study deterministic extractors for multiple independent sources.

Definition 1: For $t \in \mathbb{N}$, a function $\text{EXT}\colon (\{0,1\}^n)^t \to \{0,1\}^m$ is called a $(b_1, b_2, \ldots, b_t, \epsilon)$-extractor if for any t *independent* random variables $\mathcal{X}_1, \ldots, \mathcal{X}_t$, with each $\mathcal{X}_i$ distributed over $\{0,1\}^n$ and $\mathrm{H}_\infty(\mathcal{X}_i) \geq b_i$, we have

$$\|\text{EXT}(\mathcal{X}_1, \ldots, \mathcal{X}_t) - \mathcal{U}_m\| \leq \epsilon.$$

EXT is called a $(b_1, \ldots, b_t, \epsilon)$-strong-two-source-extractor if it satisfies the stronger property that

$$\left\|\text{EXT}(\mathcal{X}_1, \ldots, \mathcal{X}_t) \circ \mathcal{X}_{[2,t]} - \mathcal{U}_m \circ \mathcal{X}_{[2,t]}\right\| \leq \epsilon.$$

A distribution is called *flat* if it is a uniform distribution over some set S. It is well known that any distribution of min-entropy k is a convex combination of flat distributions of min-entropy k, so to analyze extractors it suffices to work for flat distributions.

III. Generalized Leftover Hash Lemma

Definition 2: We call a family H of functions from $\{0,1\}^n$ to $\{0,1\}^m$ *pairwise independent* if

$$\forall x_1 \neq x_2: \quad \Pr_{h \in H}[h(x_1) = h(x_2)] = \frac{1}{2^m}.$$

The well-known leftover hash lemma [6] says that if h is sampled uniformly from such a family H and x is sampled from a distribution with enough min-entropy, the distribution of $h(x) \circ h$ is close to uniform. We extend it to the case that h is sampled from a large enough subset G of H. Note that the original leftover hash lemma is a special case of our lemma with $G = H$.

Lemma 1 (Generalized Leftover Hash Lemma): Let H be any family of pairwise independent functions from $\{0,1\}^n$ to $\{0,1\}^m$. Let $\mathcal{G}$ denote the uniform distribution over a set $G \subseteq H$ and let $\mathcal{X}$ denote the uniform distribution over a set $X \subseteq \{0,1\}^n$. Then

$$\|\mathcal{G}(\mathcal{X}) \circ \mathcal{G} - \mathcal{U}_m \circ \mathcal{G}\| \leq \frac{1}{2}\sqrt{\frac{2^m |H|}{|X||G|}}.$$

Proof: $4\|\mathcal{G}(\mathcal{X}) \circ \mathcal{G} - \mathcal{U}_m \circ \mathcal{G}\|^2$ is equal to

$$\begin{aligned}
&\left(\sum_{h \in G} \sum_{z \in \{0,1\}^m} \frac{1}{|G|} \left| \Pr_{x \in X}[h(x) = z] - \frac{1}{2^m} \right| \right)^2 \\
&\le \frac{2^m}{|G|} \cdot \left[\sum_{h \in H} \sum_{z \in \{0,1\}^m} \left(\Pr_{x \in X}[h(x) = z] - \frac{1}{2^m} \right)^2 \right] \qquad (1) \\
&= \frac{2^m}{|G|} \left[\left(\sum_{h \in H} \sum_{z \in \{0,1\}^m} \Pr_{x, x' \in X}[h(x) = h(x') = z] \right) - \frac{|H|}{2^m} \right] \\
&= \frac{2^m |H|}{|G|} \left(\Pr_{h \in H; x, x' \in X}[h(x) = h(x')] - \frac{1}{2^m} \right) \\
&\le \frac{2^m |H|}{|G|} (\Pr[x = x'] + \Pr[h(x) = h(x') \,|\, x \ne x'] \\
&\quad - 1/2^m) \\
&\le \frac{2^m |H|}{|G|} \left(\frac{1}{|X|} + \frac{1}{2^m} - \frac{1}{2^m} \right) \\
&= \frac{2^m |H|}{|X| |G|} \qquad (2)
\end{aligned}$$

where (1) is due to the Jensen's inequality, and (2) holds because H is pairwise independent. □

IV. Extracting From Two Independent Sources

We apply the generalized leftover hash lemma to extract randomness from two independent sources $\mathcal{X}$ and $\mathcal{Y}$ over $\{0,1\}^n$ with $\mathrm{H}_\infty(\mathcal{X}) \ge b_1$ and $\mathrm{H}_\infty(\mathcal{Y}) \ge b_2$. For any $n, m \in \mathbb{N}$ with $m \,|\, n$, let $\ell = (n/m)$, and we treat any $v \in \{0,1\}^n$ as an ℓ-dimensional vector $v = (v_1, v_2, \ldots, v_\ell)$ with each $v_i \in \mathrm{GF}(2^m)$. Now define our extractor $\mathrm{Ext}^2 : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ as

$$\mathrm{EXT}^2(x, y) = \langle x, y \rangle_m = \sum_{i=1}^{\ell} x_i \cdot y_i \in \mathrm{GF}(2^m)$$

which is the inner product of x and y over $\mathrm{GF}(2^m)$.

Theorem 1: The function Ext^2 is a (b_1, b_2, ϵ)-strong-two-source-extractor with $\epsilon = 2^{-(b_1+b_2+2-n-m)/2}$.

Proof: Let $H = \{h_y \,|\, y \in \{0,1\}^n\}$, where $h_y(x) = \langle x, y \rangle_m$ for $x, y \in \{0,1\}^n$. It is easy to check that the family H is pairwise independent. Then the theorem follows immediately from Lemma 1. □

V. Extracting From t Independent Sources

Next, we show how to extract randomness from t independent sources $\mathcal{X}_1, \ldots, \mathcal{X}_t$. Define the extractor $\mathrm{Ext}^t : (\{0,1\}^n)^t \to \{0,1\}^m$ as

$$\mathrm{EXT}^t(x_1, \ldots, x_t) = \sum_{1 \le i < j \le t} \langle x_i, x_j \rangle_m.$$

Theorem 2: The function Ext^t is a $(b_1, \ldots, b_t, \epsilon)$-strong-multisource-extractor with $\epsilon = 2^{-(b_1+k+2-n-m)/2}$, where $k = \max(b_2, \ldots, b_t)$.

Proof: Assume, without loss of generality, that $\mathcal{X}_2$ is the source with the largest min-entropy among $\mathcal{X}_2, \ldots, \mathcal{X}_t$. Fix any values $x_3, \ldots, x_t$, let $s = x_3 + \cdots + x_t$, and let $\alpha = \sum_{3 \le i < j \le t} \langle x_i, x_j \rangle_m$. Then

$$\mathrm{EXT}^t(\mathcal{X}_1, \mathcal{X}_2, x_3, \ldots, x_t) = \langle \mathcal{X}_1, \mathcal{X}_2 \rangle_m + \langle \mathcal{X}_1, s \rangle_m + \langle \mathcal{X}_2, s \rangle_m + \alpha.$$

Consider the family of functions $H = \{h_y \,|\, y \in \{0,1\}^n\}$ where $h_y(x) = \langle x, y \rangle_m + \langle x, s \rangle_m + \langle y, s \rangle_m + \alpha$. It is pairwise independent because for any $x \ne x'$

$$\begin{aligned}
&\Pr_y[h_y(x) = h_y(x')] \\
&\quad = \Pr_y[\langle x, y \rangle_m + \langle x, s \rangle_m = \langle x', y \rangle_m + \langle x', s \rangle_m] \\
&\quad = \Pr_y[\langle x - x', y \rangle_m = \langle x' - x, s \rangle_m] \\
&\quad = \frac{1}{2^m}.
\end{aligned}$$

Therefore, Theorem 1 implies

$$\|\mathrm{EXT}^t(\mathcal{X}_1, \mathcal{X}_2, x_3, \ldots, x_t) \circ \mathcal{X}_2 - \mathcal{U}_m \circ \mathcal{X}_2\| \le 2^{-(b_1+b_2+2-n-m)/2}$$

for any $x_3, \ldots, x_t$. Thus,

$$\begin{aligned}
&\left\|\mathrm{EXT}^t(\mathcal{X}_1, \ldots, \mathcal{X}_t) \circ \mathcal{X}_{[2,t]} - \mathcal{U}_m \circ \mathcal{X}_{[2,t]}\right\| \\
&\quad \le \sum_{x_{[3,t]}} \big(\Pr[X_{[3,t]} = x_{[3,t]}] \\
&\qquad \cdot \|\mathrm{EXT}^t(\mathcal{X}_1, \mathcal{X}_2, x_3, \ldots, x_t) \circ \mathcal{X}_2 - \mathcal{U}_m \circ \mathcal{X}_2\|\big) \\
&\quad \le 2^{-(b_1+b_2+2-n-m)/2}.
\end{aligned}$$

□

If the sources are not exposed, we can have a slightly better result. Note that

$$\begin{aligned}
&\|\mathrm{EXT}^t(\mathcal{X}_1, \ldots, \mathcal{X}_t) - \mathcal{U}_m\| \\
&\qquad \le \left\|\mathrm{EXT}^t(\mathcal{X}_1, \ldots, \mathcal{X}_t) \circ \mathcal{X}_{[2,t]} - \mathcal{U}_m \circ \mathcal{X}_{[2,t]}\right\|
\end{aligned}$$

so by taking $\mathcal{X}_1$ in Theorem 2 to be the source with the highest min-entropy, we have the following.

Corollary 1: The function Ext^t is a $(b_1, \ldots, b_t, \epsilon)$-extractor with $\epsilon = 2^{-(k_1+k_2+2-n-m)/2}$, where k_1 and k_2 are the two largest values among $b_1, \ldots, b_t$.

Note that in the construction of our extractor Ext^t, we do not need to know beforehand the specific min-entropy of each source. It works as long as the sum of the two largest min-entropies is large enough.

VI. Application

Consider the following cryptographic setting in which a group of parties $P_1, P_2, \ldots, P_u$ want to establish a secret key for group communication. Suppose initially these parties are together and can sample $\mathcal{A}_1, \mathcal{B}_1, \ldots, \mathcal{A}_u, \mathcal{B}_u, \mathcal{X}$ from some blockwise source [2], where each block ($\mathcal{A}_i, \mathcal{B}_i$, or $\mathcal{X}$) is n-bit long and has min-entropy at least b even given all the previous blocks. After that, all parties go far away from each other but are connected by an insecure network. If they want to communicate securely later on, they can execute the following protocol.

1) In the order of i from 1 to u, party P_i samples $\mathcal{X}_i$ from his/her own local source, computes $\mathcal{Y}_i = \mathcal{A}_i \mathcal{X}_i + \mathcal{B}_i$, and sends $(\mathcal{X}_i, \mathcal{Y}_i)$ to all other parties.
2) When receiving $(\tilde{\mathcal{X}}_j, \tilde{\mathcal{Y}}_j)$ from an alleged party P_j, each P_i verifies whether $\tilde{\mathcal{Y}}_j = \mathcal{A}_j \tilde{\mathcal{X}}_j + \mathcal{B}_j$. Let $T = \{P_{i_1}, P_{i_2}, \ldots, P_{i_{t-1}}\}$ be the set of parties who pass this authentication test.
3) Each party in T computes the secret key
$$\mathcal{K} = \mathrm{EXT}^t(\mathcal{X}, \mathcal{X}_{i_1}, \mathcal{X}_{i_2}, \ldots, \mathcal{X}_{i_{t-1}})$$
which can be used, for example, as the secret key of the one-time pad encryption.

We discuss two security issues. For authentication, we know that after seeing $(\mathcal{X}_i, \mathcal{Y}_i)$ for every $i < j$, $(\mathcal{A}_j, \mathcal{B}_j)$ still has min-entropy $2b$, so from [4], an adversary can only impersonate a party P_j with probability $2^{-(2b-n)}$.

For the security of $\mathcal{K}$, note that $\mathcal{X}$ is assumed to have min-entropy b even given $\mathcal{A}_1, \mathcal{B}_1, \ldots, \mathcal{A}_u, \mathcal{B}_u$, and we can assume that $\mathcal{X}, \mathcal{X}_{i_1}, \mathcal{X}_{i_2}, \ldots, \mathcal{X}_{i_{t-1}}$ are mutually independent as they are generated in distant places. Thus, Theorem 2 implies

$$\|\langle \mathcal{X}_{i_1}, \ldots, \mathcal{X}_{i_{t-1}}, \mathcal{K}\rangle - \langle \mathcal{X}_{i_1}, \ldots, \mathcal{X}_{i_{t-1}}, \mathcal{U}\rangle\| \le 2^{-(b+k+2-n-m)/2}$$

where $k = \max(\mathrm{H}_\infty(\mathcal{X}_{i_1}), \ldots, \mathrm{H}_\infty(\mathcal{X}_{i_{t-1}}))$. That is, $\mathcal{K}$ is secure enough when $b + k \gg n + m$. Note that any strong extractor will also work.

Acknowledgment

The authors would like to thank the anonymous referees for very useful comments.

References

[1] B. Barak, R. Impagliazzo, and A. Wigdersom, "Extracting randomness from few independent sources," in *Proc. IEEE 45th Annu. IEEE Symp. Foundations of Computer Science (FOCS'04)*, Rome, Italy, Oct. 2004, pp. 384–393.

[2] B. Chor and O. Goldreich, "Unbiased bits from sources of weak randomness and probabilistic communication complexity," *SIAM J. Comput.*, vol. 17, no. 2, pp. 230–261, Apr. 1988.

[3] Y. Dodis, A. Elbaz, R. Oliveira, and R. Raz, "Improved randomness extraction from two independent sources," in *Proc. APPROX-RANDOM*, Cambridge, MA, Aug. 2004, pp. 334–344.

[4] Y. Dodis and R. Oliveira, "On extracting private randomness over a public channel," in *Proc. APPROX-RANDOM*, Princeton, NJ, Aug. 2003, pp. 252–263.

[5] R. Graham and J. Spencer, "A constructive solution to a tournament problem," *Canad. Math. Bull.*, vol. 14, pp. 45–48, Jan. 1971.

[6] R. Impagliazzo, L. A. Levin, and M. Luby, "Pseudo-random generation from one-way functions," in *Proc. 21st Annu. ACM Symp. Theory of Computing (STOC'89)*, Seattle, WA, May 1997, pp. 290–294.

[7] J. Kamp and D. Zuckerman, "Deterministic extractors for bit-fixing sources and exposure-resilient cryptography," in *Proc. IEEE 44th Annu. IEEE Symp. Foundations of Computer Science (FOCS'03)*, Cambridge, MA, Oct. 2003, pp. 92–101.

[8] R. Konig and U. Maurer, "Extracting randomness from generalized symbol-fixing and Markov sources," in *Proc. IEEE Int. Symp. Information Theory (ISIT'04)*, Chicago, IL, Jun./Jul. 2004, p. 232.

[9] C. Lu, O. Reingold, S. Vadhan, and A. Wigderson, "Extractors: Optimal up to constant factors," in *Proc. 35st Annu. ACM Symp. Theory of Computing (STOC'03)*, San Diego, CA, Jun. 2003, pp. 602–611.

[10] N. Nisan and D. Zuckerman, "Randomness is linear in space," *J. Comp. Syst. Sci.*, vol. 52, no. 1, pp. 43–52, Feb. 1996.

[11] R. Shaltiel, "Recent developments in explicit constructions of extractors," in *Bull. EATCS*, vol. 77, 2002, pp. 67–95.

Bounds on the Performance of Vector-Quantizers Under Channel Errors

Gal Ben-David, *Senior Member, IEEE,* and David Malah, *Fellow, IEEE*

Abstract*—Vector quantization (VQ) is an effective and widely known method for low-bit-rate communication of speech and image signals. A common assumption in the design of VQ-based communication systems is that the compressed digital information is transmitted through a perfect channel. Under this assumption, quantization distortion is the only factor in output signal fidelity. Moreover, the assignment of channel symbols to the VQ reconstruction vectors is of no importance. However, under physical channels, errors may be present, causing degradation in overall system performance. In such a case, the effect of channel errors on the coding system performance depends on the index assignment of the reconstruction vectors. The index assignment problem is a special case of the Quadratic Assignment Problem (QAP) and is known to be NP-complete. For a VQ with N reconstruction vectors there are N! possible assignments, meaning that an exhaustive search over all possible assignments is practically impossible. To help the VQ designer, we present in this correspondence lower and upper bounds on the performance of VQ systems under channel errors, over all possible assignments. The bounds coincide with a general bound for the QAP. Nevertheless, the proposed derivation allows us to compare the bounds with published results on VQ index assignment. A related expression for the average performance is also given and discussed. Special cases and numerical examples are given in which the bounds and average performance are compared with index assignments obtained by known algorithms.

***Index Terms*—Channel coding, index assignment (IA), performance bounds, vector quantization (VQ).**

I. Introduction

Vector quantization (VQ) is a method for mapping signals into digital sequences. A typical VQ-based communication system is shown in Fig. 1.

A discrete-time *source* emits signal samples over an infinite (or densely finite) alphabet. These samples should be sent to the *destination* with the highest possible fidelity. The *VQ encoder* translates source output vectors into *channel* digital sequences. The *VQ decoder*'s goal is to reconstruct source samples from this digital information. Since the analog information cannot be perfectly represented by the digital information some *quantization distortion* must be tolerated.

In each channel transmission, the VQ encodes a K-dimensional vector of source samples $\underline{x}(t)$ into a *reconstruction vector index* $y(t)$, where the discrete variable t represents the time instant or a channel-use counter. The index is taken from a finite alphabet $y(t) \in \{0, 1, \ldots, N-1\}$, where N is the number of reconstruction vectors (hence the number of possible channel symbols).

The *index assignment* (IA) is represented in Fig. 1 by a permutation operator

$$\Pi : y(t) \in \{0, 1, \ldots, N-1\} \rightarrow z(t) \in \{0, 1, \ldots, N-1\}. \quad (1)$$

Manuscript received December 9, 2000; revised July 8, 2004. This work was carried out at the Electrical Engineering Department, Technion–Israel Institute of Technology, Haifa, Israel. The material in this correspondence was presented in part at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP01), Salt Lake City, UT, May 2001.

G. Ben-David is a Consultant in Digital Signal Processing, Mitzpe Adi, Israel (e-mail: galb@netvision.net.il).

D. Malah is with the Electrical Engineering Department, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: malah@ee.technion.ac.il).

Communicated by P. A. Chou, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2005.847750