Explicit Exclusive Set Systems with Applications to Broadcast Encryption

Craig Gentry
Stanford University
cgentry@cs.stanford.edu

Zulfikar Ramzan
Symantec, Inc.
zulfikar_ramzan@symantec.com

David P. Woodruff *
MIT and Tsinghua University
dpwood@mit.edu

Abstract

A family of subsets $\mathcal{C}$ of $[n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$ is (r,t)-exclusive if for every $S \subset [n]$ of size at least $n - r$, there exist $S_1, \ldots, S_t \in \mathcal{C}$ with $S = S_1 \cup S_2 \cup \cdots \cup S_t$. These families, also known as complement-cover families, have cryptographic applications, and form the basis of information-theoretic broadcast encryption and multi-certificate revocation. We give the first explicit construction of such families with size $\mathrm{poly}(r,t)n^{r/t}$, essentially matching a basic lower bound. Our techniques are algebraic in nature.

When $r = O(t)$, as is natural for many applications, we can improve our bound to $\mathrm{poly}(r,t)\binom{n}{r}^{1/t}$. Further, when r, t are small, our construction is tight up to a factor of r. We also provide a $\mathrm{poly}(r,t,\log n)$ algorithm for finding $S_1, \ldots, S_t$, which is crucial for efficient use in applications. Previous constructions either had much larger size, were randomized and took super-polynomial time to find $S_1, \ldots, S_t$, or did not work for arbitrary n, r, and t. Finally, we improve the known lower bound on the number of sets containing each $i \in [n]$. Our bound shows that our derived broadcast encryption schemes have essentially optimal total number of keys and keys per user for n users, transmission size t, and revoked set size r.

1. Introduction

In [1], Aiello, Lodha, and Ostrovsky put forth the notion of a *complement-cover family*, which, informally speaking, is a family of sets for which every large subset of the universe can be written as the union of some small collection of subsets from the family. This notion was rediscovered by Kumar and Russell [10], who referred to such families as *exclusive set systems*. More formally,

Definition 1 *A family of subsets $\mathcal{C} = \{S_1, \ldots, S_k\}$ over $[n]$ is (n,k,r,t)-exclusive if for any subset $R \subset [n]$ with $|R| \leq r$, we can write $[n] \setminus R = \cup_{j=1}^{t} S_{i_j}$ for some $1 \leq i_j \leq k$.*

Determining the exact tradeoff between $n, k, r,$ and t is a fundamental combinatorial problem with significant applications in cryptography. One of the most notable of these is *information-theoretic broadcast encryption* [4, 11]. In such a scheme, there is a server sending *broadcasts* to n users. Each broadcast consists of t encryptions of a session key, with the property that any coalition of members from a revoked set $R \subset [n]$ with $|R| \leq r$ learns no information from the broadcast. Each encryption in the broadcast is done using a different key, held by both the server and some specific set of users. Thus, k is the number of keys in the scheme, which is proportional to the storage complexity, and S_i denotes the users who hold the server's ith key. If a broadcast consists of encryptions under keys $S_{i_1}, \ldots, S_{i_t}$, it should hold that $\cup_{j=1}^{t} S_{i_j} = [n] \setminus R$. Then even if all the users in R collude, they collectively lack each key in the broadcast.

Since each subset of t keys can correspond to at most one set $[n] \setminus R$, we need $\sum_{i=0}^{t}\binom{k}{i} \geq \sum_{i=0}^{r}\binom{n}{i} \geq \binom{n}{r}$. After some algebra, this gives the lower bound $k = \Omega(t\binom{n}{r}^{1/t})$.

Kumar and Russell [10] use the probabilistic method to show that for sufficiently large n and any $r \leq t$, there exist exclusive set systems with size $O(t^3(nt)^{r/t}\ln n)$. This, however, is only an existence result and has several drawbacks. First, it is not known how to efficiently verify that the randomly chosen sets do indeed form an exclusive set system. Second, the sets have a large description size since they are chosen at random. Third, the number of sets grows as $n^{r/t}$ rather than $\binom{n}{r}^{1/t}$, which is important for large r.

Finally, and perhaps most importantly, [10] does not provide an efficient algorithm for generating $S_{i_1}, \ldots, S_{i_t}$ with $[n] \setminus R = S_{i_1} \cup \cdots \cup S_{i_t}$, which is equivalent to solving Set-Cover on a certain input distribution. As far as we are aware, this problem has not been extensively studied, with the only results appearing in [15, 16]. The strongest result is in [15], which shows how to obtain an additive, slightly sublogarithmic approximation factor in $\mathrm{poly}(n)$-time. Oftentimes though, even time polynomial in n is considered too large, as r, t are usually much smaller. Hence, it is desirable to have algorithms running in time[1] $\mathrm{poly}(r,t,\log n)$.

*Some of this research was done while the authors were all at DoCoMo Labs.

[1] This is the time to determine the members of a collection of sets that form the desired union. However, it may not be enough time to output the

In this paper, we give an *explicit* construction of an (n, k, r, t)-exclusive set system with $k = \text{poly}(r, t)n^{r/t}$ sets. Unlike previous constructions, our construction works for *any* values of r, t, and n. Further, assuming that $r = O(t)$, this can be improved to $\text{poly}(r, t)\binom{n}{r}^{1/t}$, which is optimal up to the $\text{poly}(r, t)$ factor. In applications such as broadcast encryption, usually the communication is at least r since the server needs to describe the set R to the users. Thus, it is likely that $r \leq t$, and we can apply our improved construction. For the case when r and t are slow-growing functions of n, as is the case for broadcast encryption, we can optimize our storage complexity to $k = O(rt\binom{n}{r}^{1/t})$. This improves the previous best known complexity of [10], and is tight up to a factor of r. Moreover, we provide a deterministic $\text{poly}(r, t, \log n)$-time algorithm, which given R, finds $S_{i_1}, \ldots, S_{i_t}$ with $[n] \setminus R = S_{i_1} \cup \cdots \cup S_{i_t}$. Thus, broadcasting is extremely efficient.

In [13], Luby and Staddon derive lower bounds on the number of sets in an exclusive set system which contain a given element $i \in [n]$. They show that there exists an $i \in [n]$ occurring in $\binom{n}{r}^{1/t}/(rt)$ sets. In this paper, we improve their bound to $\binom{n}{r}^{1/t}/r$ using a variation of the sunflower lemma. These bounds show that for broadcast encryption, the *total number* of keys in our scheme is about the same as the number of keys required *per user* in *any scheme*. Thus, our number of keys per user is also essentially optimal.

There have been many other constructions of exclusive set systems. Aiello, Lodha, and Ostrovsky [1] construct $(n, n2^c/(c-1), r, r\log_c n/r)$-exclusive set systems for any constant $c \geq 2$. They also show how to use any (n, k, r, t)-exclusive set system for efficient *certificate revocation*. Here n refers to the number of users, k the number of certificates held by the certificate authority, r the number of revoked users, and t the communication complexity of an update phase.

In the context of broadcast encryption, Gafni, Staddon, and Yin [5] provide an $(n, (r\log n/\log r)^2, r, (r\log n/\log r)^2)$-exclusive set system. In the same context, Lotspiech, Naor, and Naor [11] give $(n, 2n, r, r\log n/r)$ and $(n, n\log n, r, 2r)$-exclusive set systems based on binary trees. Using algebraic-geometric codes, Kumar, Rajagopalan, and Sahai [9] construct $(n, r^3\log n/\log r, r, r^3\log n/\log r)$-exclusive set systems.

Although these schemes are equipped with efficient algorithms for generating $S_{i_1}, \ldots, S_{i_t}$ with $[n] \setminus R = S_{i_1} \cup \cdots \cup S_{i_t}$, a serious disadvantage of all of these schemes is that once n and r are chosen, both the broadcast size t and the number of keys (certificates) k are determined. However, as pointed out in [10], it is clear that given n, r and t, for sufficiently large k there exists an (n, k, r, t)-exclusive set system. In contrast with previous schemes, our schemes can support arbitrary n, r, and t. In fact, below we will see that our schemes may even improve the parameters of these specific schemes. Thus, our constructions significantly generalize the previous ones, and are almost tight.

Improved Cryptographic Applications: In the table below, we have listed the previous results, as well as two settings of parameters of our scheme. We stress that our results are listed for general n and r, and for small values of r the degree of our polynomial factors is very small. Below we also discuss the tradeoffs on the number of keys per user.

Paper	Communication complexity	Number of Keys
[9]	$O(r^3\log n/\log r)$	$O(r^3\log n/\log r)$
[5]	$O(r^2\log^2 n/\log^2 r)$	$O(r^2\log^2 n/\log^2 r)$
[1, 11]	$r\log n/r$	$2n$
[11](SD)	$2r$	$\Theta(n\log n)$
this paper	$r\log n/r$	$\text{poly}(r, \log n)$
this paper	$2r$	$\text{poly}(r)n^{1/2}$

The first setting of parameters in this paper outperforms [9] and [5] in terms of communication, while paying an extra $\text{poly}(r, \log n)$ factor in the key complexity. This is useful when communication is the bottleneck. This setting also provides an exponential improvement to the key complexity of [1] and [11] for small r. Our improvement comes from the fact that the schemes in [1] and [11] are not sensitive to r, whereas we parameterize our complexity in terms of r, which is likely to be small in practice. We note that the number of keys per user in our scheme is only $\text{poly}(r, \log n)$, which is comparable to that of previous schemes.

The second setting of parameters in our scheme is useful for a comparison to another scheme proposed in [11], known as the subset-difference (SD) scheme. For the same n, r, and t, we achieve $O(r^2n^{1/2})$ keys, roughly the square-root of the SD scheme for small r. One may argue that the SD scheme focuses instead on the number of keys per user. At first glance, it appears that their scheme achieves only $\Theta(\log^2 n)$ keys per user, contradicting our lower bound. A more careful inspection shows that their scheme only provides *computational security*, and thus is incomparable with ours, which is information-theoretic. Making their scheme information-theoretically secure requires $\Omega(n)$ keys per user, while ours only requires $\text{poly}(r)n^{1/2}$. We achieve similar improvements over the LSD broadcast encryption scheme of Halevy and Shamir [6].

Other Applications: Although the most immediate applications of our results are to broadcast encryption and certificate revocation, our results may also apply to data structures and group testing. We note that key distribution patterns, a generalization of broadcast encryption, have been studied in connection to group testing before [14].

Techniques: The idea behind our construction is to first construct exclusive set systems for the case when r and t are much smaller than n. We then create an exclusive system for general n, r, and t with a divide-and-conquer approach:

members of each set, though for our applications this is not needed.

roughly speaking, we carefully partition the universe $[n]$ into blocks and use our smaller set systems independently on each block.

The construction for small r, t is algebraic in nature. Namely, we associate $[n]$ with points in affine space. Sets then correspond to functions f on this space. More precisely, a set corresponds to the points on which f does not vanish. Then a point u belongs to the set union $S_1 \cup \cdots \cup S_t$ provided it does not vanish on all of the corresponding functions $f_1, \ldots, f_t$. Algebraically, this means that u is not in the variety of $f_1, \ldots, f_t$. The main problem is to find a small explicit collection of functions for which every set of at most r points is the variety of some t functions in the collection. In this way, we have reduced the problem to a specific algebraic question. We find an explicit family using multivariate polynomials together with certain expanders and MDS codes.

Our improved lower bound for the number of sets containing each $i \in [n]$ is based on the sunflower lemma with relaxed disjointness.

Organization: In section 2, we develop our polynomial-based system for small r and t. In section 2.1 we use expanders to improve the first construction. In section 2.2 we use a small amount of randomness for further improvements, while preserving deterministic, efficient broadcast. In section 2.3 we balance the different types of sets that we use, giving further improvements. In section 3, we construct an exclusive set system for general n, r, and t. For readability, we give our lower bound on the keys per user in Appendix A. We also give more intuition for the scheme in section 2 in Appendix B.

2. The Polynomial System

Recall that n is the universe size, $n - r$ is the size of the sets $[n] \setminus R$ we wish to cover, and t the number of sets we use to cover each set $[n] \setminus R$. We start by describing a simplified scheme under the assumption that

$$r^\alpha t^2 \leq n^{1/t}$$

for a constant $\alpha > 2$ to be specified. For now the reader should just think of r and t as being much smaller than n. Let $p \geq n^{1/t}$ be prime, and let $\mathbb{F} = \mathbb{F}_p$. For $x \in [n]$, we identify x with a point $(x_0, \ldots, x_{t-1}) \in \mathbb{F}^t$.

Our scheme works by choosing a small collection $\mathcal{C}$ of polynomials in the ring $\mathbb{F}[X_0, \ldots, X_{t-1}]$, where $X_0, \ldots, X_{t-1}$ are formal variables. For each $f \in \mathcal{C}$, we create a set S_f consisting of all the points u in $\mathbb{F}^t$ for which $f(u) \neq 0$. Given a set $R \subset [n]$ with $|R| \leq r$, we will find t functions $f_0, \ldots, f_{t-1} \in \mathcal{C}$ for which $\mathbf{Var}(f_0, \ldots, f_{t-1}) = R$, where $\mathbf{Var}(f_0, \ldots, f_{t-1})$ denotes the common zeros of $f_0, \ldots, f_{t-1}$, that is, the variety of these functions. By construction, any $u \in [n] \setminus R$ occurs in some set, while any $u \in R$ does not.

The problem is therefore to find an explicit polynomial collection $\mathcal{C}$ with these properties. We consider the following collection

$$\begin{aligned}\mathcal{C} &= \{\prod_{j=1}^{r'}(X_0 - i_j) \mid r' \leq r, \text{ distinct } i_1, \ldots, i_{r'} \in \mathbb{F}\} \\ &\cup \{f(X_i) - X_{i+1} \mid 0 \leq i \leq t-2, \deg(f) \leq r-1\}\end{aligned}$$

The number of polynomials of the form $\prod_{j=1}^{r'}(X_0 - i_j)$ is $\sum_{i=0}^{r}\binom{p}{i} \leq \sum_{i=0}^{r} p^i \leq 2p^r$, and the number of univariate polynomials f of degree at most $r-1$ is at most p^r, so $|\mathcal{C}| = O(tp^r)$.

Intuition: The idea we use is that polynomials of the form $f(X_i) - X_{i+1}$ implement an AND operation between adjacent coordinates. Since the polynomials have degree $r-1$, we can only use a given polynomial to implement r constraints. By chaining t of the polynomials together, we can exclude exactly those points in R, coordinate by coordinate. Finally, we need polynomials of the form $\prod_{j=1}^{r'}(X_0 - i_j)$ for the base case, that is, to begin the chaining. One important observation is that by using polynomials to implement these local constraints, we greatly reduce the total number of sets k. The reason is that the mapping from sets of r constraints to polynomials is many-to-one.

We start with the following lemma, which formalizes this intuition.

Lemma 2 *Suppose that for each i in $[t]$, no two points in R have the same ith coordinate. Then we can find $f_0, \ldots, f_{t-1} \in \mathcal{C}$ for which $\mathbf{Var}(f_0, \ldots, f_{t-1}) = R$.*

Proof: Since the coordinates have distinct values and $|R| \leq r$, we can choose $f_0 = \prod_{u \in R}(X_0 - u_0)$. For $i > 1$, we find a univariate polynomial g_i by interpolating from $g_i(u_i) = u_{i+1}$ for each $u \in R$, and then setting the multivariate polynomial $f_i = g_i(X_i) - X_{i+1}$. For any point $x \notin R$, if $x \in \mathbf{Var}(f_0, \ldots, f_{t-1})$, then $f_0(x_0) = 0$, so that $x_0 = u_0$ for some $u \in R$. It inductively follows that $u_{i+1} = g_i(u_i) = g_i(x_i) = x_{i+1}$, where the first equality follows from the definition of g_i, the second equality follows from the inductive hypothesis, and the last equality from the fact that $f_i(x) = g_i(x_i) - x_{i+1} = 0$. Thus $x = u$, a contradiction. On the other hand, if $x \in R$, then it is easy to see that $x \in \mathbf{Var}(f_0, \ldots, f_{t-1})$. ∎

The remaining problem is how to handle the case when points in R share coordinates. One idea is to carefully choose a small set of invertible linear transformations $L_1, \ldots, L_m$ on the space $X_0, \ldots, X_{t-1}$ so that for any set R, there is some index B for which each row of $L_B R$ consists of distinct entries. We then proceed as before in this

new coordinate system. In this case we say that L_B is *good* for R. Here, L_B is interpreted as a $t \times t$ matrix and R as a $t \times r$ matrix. We then define $\mathcal{C}$ to be $\cup_{B=1}^{m} \mathcal{C}_B$, where $\mathcal{C}_B$ is given by

$$\mathcal{C}_B = \{\prod_{j=1}^{r'}(L_B X_0 - i_j) \mid r' \leq r, \text{ distinct } i_1, \ldots, i_{r'} \in \mathbb{F}\}$$

$$\cup \{f(L_B X_i) - L_B X_{i+1} \mid 0 \leq i \leq t-2, \deg(f) \leq r-1\}.$$

The size of $\mathcal{C}$ is $O(mtp^r)$. For a given R with $|R| \leq r$, we find a B for which L_B is good for R, and then apply the previous scheme using the sets in $\mathcal{C}_B$. To complete the specification, we use the following lemma.

Lemma 3 *There is an explicit set of $m = r^2 t$ linear transformations $L_1, \ldots, L_m$ such that for all $R \subset [n]$ of size at most r, there is some L_B that is good for R.*

Proof: Divide $\mathbb{F}$ into $m = r^2 t$ disjoint blocks $B = \{b_1, \ldots, b_t\}$ each containing t distinct elements. This is possible since $r^2 t^2 \leq n^{1/t} \leq p$. Define the linear transformations

$$L_B = \begin{bmatrix} 1 & b_1 & b_1^2 & \ldots & b_1^{t-1} \\ 1 & b_2 & b_2^2 & \ldots & b_2^{t-1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & b_t & b_t^2 & \ldots & b_t^{t-1} \end{bmatrix}.$$

The L_B are invertible, and $L_B(x) = p_x(b_1), p_x(b_2), \ldots, p_x(b_t)$, where $p_x(Y) = \sum_{i=0}^{t-1} x_i Y^i$. As two distinct degree $t-1$ polynomials can agree on at most $t-1$ points, it follows that for any given R, at most $\binom{r}{2}(t-1) < r^2 t$ elements v of $\mathbb{F}$ can be such that $p_x(v) = p_y(v)$ for distinct $x, y \in R$. Therefore, one of the L_B is good for R. ■

We summarize our findings thus far. We use the term *broadcast* to mean, given a set R of cardinality at most r, find $i_1, \ldots, i_t$ for which $[n] \setminus R = S_{i_1} \cup \cdots \cup S_{i_t}$. We refer to R as the set of *revoked* points. This terminology coincides with that for broadcast encryption, where points are users.

Theorem 4 *Let $\alpha > \frac{1}{1-.525}$ be any constant, and assume $\max(r^\alpha, r^2 t^2) = O(n^{1/t})$. For sufficiently large n, there is an explicit $(n, O((rt)^2 n^{r/t}), r, t)$-exclusive set system. Further, broadcasting can be done in $\mathrm{poly}(r, t, \log n)$ time.*

Proof: By lemma 3 we can set $m = r^2 t$ in the discussion above. Thus $k = O((rt)^2 p^r)$. It remains to find a small prime p with $p \geq n^{1/t}$. Using a result of [3], we can find a prime p with $n^{1/t} \leq p < n^{1/t} + n^{\beta/t}$ for any constant $\beta > .525$ and sufficiently large $n^{1/t}$. Since $t \leq n^{1/t}$, we have $t = O\left(\frac{\log n}{\log \log n}\right)$, and thus $n^{1/t} = \Omega\left(\frac{\log n}{\log \log n}\right)$ so that $n^{1/t} \to \infty$ as $n \to \infty$. Therefore we can find such a prime for sufficiently large n. The number of sets is bounded by

$$\begin{aligned} O((rt)^2 p^r) &= O((rt)^2 (n^{1/t} + n^{\beta/t})^r) \\ &= O((rt)^2 n^{r/t} (1 + n^{(\beta-1)/t})^r) \\ &= O((rt)^2 n^{r/t} e^{r/n^{(1-\beta)/t}}) \\ &= O((rt)^2 n^{r/t}), \end{aligned}$$

where we have used the bound on α to conclude that $r \leq n^{(1-\beta)/t}$. The time for broadcasting is dominated by the search for a good L_B and the $t-1$ degree-$(r-1)$ polynomial interpolations, each of which can be done in $\mathrm{poly}(r, t, \log n)$ time. ■

2.1. Using expanders

We can do a bit better with a slightly different way of handling points in R that share coordinates.

Intuition: The previous scheme had $r^2 t$ coordinate systems, each of which was good for a different collection of $R \subset [n]$. In each system we interpreted a point $u \in \mathbb{F}^t$ as a polynomial, and evaluated it on t elements of $\mathbb{F}$. The system was good for R if for each of the t elements, each of the at most r polynomials in R had different evaluations. The disadvantage is that even if only one pair of polynomials collided on one element in a coordinate system, the system could not be used for R. In the worst case this happens $r^2 t$ times, so we need $r^2 t$ systems.

But only $r^2 t$ elements can have collisions, and so if we had $r^2 t + t$ elements, for any R we could find t elements to use for a coordinate system. However, if we allow any two elements to occur together in a system, the number of sets in our overall system would be too large. Interpreting the elements as nodes of a graph and pairs of elements that can occur together as edges, the property we want is that the graph is well-connected and has low degree. This is exactly the property of an expander graph. We will find a connected component of size t amongst collision-free elements and use this as a coordinate system.

Choose $m = \gamma r^2 t$ elements $1, \ldots, m \subseteq \mathbb{F}$ for some constant $\gamma > 2$ to be determined, and say an element v is *good* for a set R if, using the notation of lemma 3, for distinct $x, y \in R$, $p_x(v) \neq p_y(v)$. From the proof of lemma 3, for any R we can find $(\gamma - 1) r^2 t$ good elements for R.

The idea is to consider graphs G with constant degree d, vertex set $[m]$, and the property that any induced subgraph on a large constant fraction of vertices has a connected component of size at least $m/2 \geq t$. This property holds for certain expander graphs. Recall that a graph G is an (m, d, c)-*expander* if it has m-vertices, each vertex has degree d, and for every set of vertices $W \subset V$ with $|W| \leq m/2$, there are at least $c|W|$ vertices in $V \setminus W$ adjacent to some vertex in W.

Lemma 5 *Let G be an (m, d, c)-expander. Then any induced subgraph on more than $\frac{dm}{c+d}$ vertices of G has a connected component of size at least $m/2$.*

Proof: Let H be an arbitrary subgraph on more than $\frac{dm}{c+d}$ vertices, and let $C_1, \ldots, C_k$ be its connected components. If any of the C_i contain $m/2$ vertices, we are done. Otherwise, since G is an expander and $|C_i| < m/2$, C_i is incident to at least $c|C_i|$ distinct vertices in $G \setminus C_i$, and thus in $G \setminus H$. The multiset of vertices in $G \setminus H$ connected to H has cardinality more than $c\frac{dm}{c+d}$, which is impossible since each of the (at most) $\frac{cm}{c+d}$ vertices in $G \setminus H$ can occur at most d times. ■

For an explicit family of expanders, we use the following.

Fact 6 *[2, 12] There is an explicit family of $(m_i, 6, \frac{1}{2} - \frac{\sqrt{5}}{6})$ expanders with $m_i < m_{i+1} = O(m_i)$.*

Theorem 7 *Let $\alpha > \frac{1}{1-.525}$ be any constant, and assume $\max(r^\alpha, r^2 t) = O(n^{1/t})$. For sufficiently large n, there is an explicit $(n, O(r^2 t n^{r/t}), r, t)$-exclusive set system. Further, broadcasting can be done in $\mathrm{poly}(r, t, \log n)$ time.*

Proof: In order to apply lemma 5, we choose our constant γ and use fact 6 to construct an $(m = \gamma r^2 t,\ 6,\ \frac{1}{2} - \frac{\sqrt{5}}{6})$-expander, subject to

$$\frac{\gamma - 1}{\gamma} > \frac{6}{6 + \frac{1}{2} - \frac{\sqrt{5}}{6}}.$$

Identify $G = (V, E)$'s vertices V with $[m] \subset \mathbb{F}$, and define

$$\mathcal{C} = \{\prod_{j=1}^{r'} (\sum_{i=0}^{t-1} b^i X_i - i_j) \mid r' \le r,\ \text{distinct } i_j \in \mathbb{F},\ b \in V\}$$

$$\cup \{f(\sum_{i=0}^{t-1} b^i X_i) - \sum_{i=0}^{t-1} c^i X_i \mid \deg(f) \le r-1,\ (b, c) \in E\}.$$

The size of $\mathcal{C}$, and thus k, is $O(mdp^r)$. As in the proof of theorem 4, we can choose p so that this quantity is $O(r^2 t n^{r/t})$.

To broadcast with a revoked set R, find $(\gamma - 1)r^2 t$ vertices in G which are good for R. Then, using lemma 5 and the bound on γ, find a connected component C of at least $m/2 \ge t$ vertices good for R. This step can be done efficiently using a breadth-first search. Let v be the root of the BFS tree containing the first t vertices visited in C. For a vertex w in the tree, let $par(w)$ be its parent. Set

$$f_0 = \prod_{u \in R} (\sum_{i=0}^{t-1} v^i X_i - p_u(v)).$$

Choose the remaining $t-1$ polynomials as follows: for each $w \neq v$, find g_i by interpolating from

$$g_i(\sum_{i=0}^{t-1} w^i u_i) = \sum_{i=0}^{t-1} par(w)^i u_i$$

for $u \in R$, and set

$$f_i = g_i(\sum_{i=0}^{t-1} w^i X_i) - \sum_{i=0}^{t-1} par(w)^i X_i.$$

Every $u \in R$ vanishes on these t functions. To see that no other point x vanishes, observe that if $f_0(\sum_{i=0}^{t-1} v^i x_i) = 0$, then $\sum_{i=0}^{t-1} v^i x_i = p_u(v)$ for some $u \in R$ since f_0 has only $|R|$ zeros. By induction on the height of the tree, $\sum_{i=0}^{t-1} w^i x_i = \sum_{i=0}^{t-1} w^i u_i$ for all vertices w. As there are t vertices and p_x, p_u are degree-$(t-1)$ polynomials, $p_x = p_u$, so $x = u$, a contradiction.

The time complexity is dominated by the search for good vertices for R, the breadth-first search, and the polynomial interpolations, all of which can be done in $\mathrm{poly}(r, t, \log n)$ time. ■

2.2. Using randomness

An unfortunate drawback of the construction in [10] is that there is no efficient algorithm given to find $S_{i_1}, \ldots, S_{i_t}$ whose union is $[n] \setminus R$. We removed this problem with our explicit construction above. Further, our explicit construction achieved size $O(r^2 t n^{r/t})$ versus the $O(t^3 n^{r/t} \log n)$ complexity of the randomized construction in [10], which held only for $r \le t$.

In this section we improve our complexity further to $O(rtn^{r/t})$ via a randomized construction. Although the construction is randomized, it does not suffer from the efficiency problems of [10]. Rather, broadcasting can still be done in $\mathrm{poly}(r, t, \log n)$ time, and $\mathcal{C}$ has a short description.

Intuition: The idea is to choose the set of m points in section 2.1 randomly from $\mathbb{F}$. For a given R it then becomes unlikely that we will choose many points with collisions on R. We show this allows us to choose $O(rt)$ points rather than $O(r^2 t)$.

Lemma 8 *Let $\epsilon > 0$ and $\gamma > 1$ be any constants. Assume*

$$r^2 t < n^{(1-\epsilon)/t},$$

and choose a set of $m = 2\gamma rt/\epsilon$ elements S uniformly at random from $\mathbb{F}$. With probability $1 - n^{-\Theta(r)}$, for all R, the set S contains $2(\gamma - 1)rt/\epsilon$ good elements for R.

Proof: Fix a revoked set $R \subseteq [n]$. For $s \in S$, let v be the probability that s is not good for R, that is, there exist

distinct $x, y \in R$ for which $p_x(s) = p_y(s)$. For fixed $x \neq y$, we have

$$\Pr_s[p_x(s) = p_y(s)] \leq (t-1)/p,$$

and thus

$$v \leq \binom{r}{2}(t-1)/p \leq r^2 t/n^{1/t} < n^{-\epsilon/t}$$

by the assumption of the lemma. The probability that more than $2rt/\epsilon$ elements of S are not good for R is bounded by

$$\begin{aligned}\sum_{i=2rt/\epsilon}^{m} \binom{m}{i} v^i (1-v)^{m-i} &\leq m 2^m v^{2rt/\epsilon} \\ &< 2^{2m} n^{-2r} \\ &= n^{-2r+2m/\log n}.\end{aligned}$$

For any $n^{-\Theta(r)} \leq \delta < 1$, this is less than δn^{-r} if

$$-2r + 2m/\log n \leq -r + \log \delta/\log n,$$

or equivalently, $r \log n \geq 2m + \log 1/\delta$. By assumption, this holds for sufficiently large n because $m = O(rt)$ and $\delta > n^{-\Theta(r)}$, while $t = O(\log n/\log\log n)$ since $t < n^{1/t}$. Then the probability there exists an R for which more than $2rt/\epsilon$ elements of S are not good for R is less than $\sum_{i=0}^{r}\binom{n}{i}\delta n^{-r} \leq n^{-\Theta(r)}$. ■

Using the set S as the vertex set of an expander as in section 2.1, we conclude,

Theorem 9 *Let $\alpha > \frac{1}{1-.525}$ and $\epsilon > 0$ be any constants, and assume*

$$\max(r^\alpha, r^2 t) < n^{(1-\epsilon)/t}.$$

There is an efficient algorithm that with probability $1 - n^{-\Theta(r)}$, generates an $(n, O(rtn^{r/t}), r, t)$-exclusive set system. Broadcasting can be done in $\mathrm{poly}(r, t, \log n)$ time.

Remark 10 We do not know how to derandomize the choice of S, and consider it an interesting research direction. For our construction, the derandomization comes down to the following: find a set S of $O(rt)$ points of $\mathbb{F}$ such that any polynomial of the form $\prod_{i<j}(q_i - q_j)$ does not vanish on a constant fraction of S, where $q_1, \ldots, q_r$ are arbitrary degree-$(t-1)$ polynomials in $\mathbb{F}[X]$.

2.3. Balancing the key complexity

We have shown how to achieve complexity $k = O(rtn^{r/t})$. In this section we achieve $k = O(rt\binom{n}{r}^{1/t})$. To illustrate the technique, we first apply it to the scheme of theorem 4. There are two types of sets, those of the form $\prod_{j=1}^{r'}(L_B X_0 - i_j)$ for $r' \leq r$ and distinct $i_1, \ldots, i_r \in \mathbb{F}$, and those of the form $f(L_B X_i) - L_B X_{i+1}$, where f is a polynomial of degree at most $r-1$. If m is the number of linear combinations L_B, then the number of sets of the first type is $m\sum_{i=0}^{r}\binom{p}{i}$. To apply theorem 4, we assume $r^2t^2 = O(n^{1/t})$, so that $r = O(p^{1/2})$. It follows that[2] $m\sum_{i=0}^{r}\binom{p}{i} = \Theta(m\binom{p}{r})$. On the other hand, the number of sets of the second type is $m(t-1)p^r$.

Intuition: The complexity is dominated from sets of the second type. We will reduce the alphabet size p to some prime q, while including more alphabet symbols (other than just the first) in sets of the first type. This balances the contribution to the complexity from the two types.

Using [3], for large enough n we can choose a prime q in the interval

$$\left[\binom{n}{r}^{1/(rt)}, \binom{n}{r}^{1/(rt)} + \binom{n}{r}^{\beta/(rt)}\right]$$

for any constant $\beta > .525$. This follows if we assume $\max(r^{1+\epsilon}, t) \leq n^{1/t}$ for some constant $\epsilon > 0$. Indeed, this implies $n/r = n^{\Omega(1)}$ and $t = O(\log n/\log\log n)$, so $\binom{n}{r}^{1/(rt)} \geq (n/r)^{1/t}$, and the latter tends to ∞. We will show $k = O(mtq^r)$. Note that

$$O(mtq^r) = O\left(mt\binom{n}{r}^{1/t}\right)$$

for $r \leq \binom{n}{r}^{(1-\beta)/t}$.

Since $(n/r)^r \leq \binom{n}{r} \leq (ne/r)^r$, there is a constant $1 \leq c \leq e$, with $\binom{n}{r} = (nc/r)^r$. We represent $[n]$ by points in

$$[\lceil r/c \rceil] \times \mathbb{F}_q^t.$$

This allows elements to have distinct representations. For the moment, assume our revoked set R is such that no two members of R share their ith cooordinate for any $i > 1$. Sets of the first type contain those points x whose first two coordinates do not agree with those of any element of R. By the distinctness assumption, the number of such sets is

$$\sum_{i=0}^{r}(\lceil r/c \rceil)^r \binom{q}{i} = \Theta((\lceil r/c \rceil)^r \binom{q}{r}),$$

since the fact that $r \leq \binom{n}{r}^{(1-\beta)/t}$ implies that $r = O(\sqrt{q})$, so that the binomial sum is dominated by the last term. Sets of the second type have the form $f(X_i) - X_{i+1}$, where f has degree less than r and $2 \leq i \leq t$. Since $i \geq 2$, these polynomials do not involve the first coordinate. The number of sets of this type is $(t-1)q^r$. To show that $k = O(tq^r)$,

[2] To see this, for any constant c, $\binom{p}{c\sqrt{p}}/\binom{p}{c\sqrt{p}-1} = \Theta(\sqrt{p})$, so that $\sum_{i=0}^{c\sqrt{p}}\binom{p}{i} = \binom{p}{c\sqrt{p}} + \sum_{i=0}^{c\sqrt{p}-1}\binom{p}{i} = \Theta(\binom{p}{c\sqrt{p}})$.

we bound the sets of the first type. Up to a constant factor, this number is,

$$\begin{aligned}\left\lceil\frac{r}{c}\right\rceil^r\binom{q}{r} &\leq \left\lceil\frac{r}{c}\right\rceil^r\left(\frac{qc}{r}\right)^r \\ &\leq \left(\left(\frac{r}{c}+1\right)\frac{qc}{r}\right)^r \\ &\leq \left(1+\frac{c}{r}\right)^r q^r \leq e^c q^r = O(q^r),\end{aligned}$$

where we used the constraints $\binom{n}{r} = (nc/r)^r$ and $q \leq n$ to deduce that $\binom{q}{r} \leq (qc/r)^r$.

Theorem 11 *Let $\alpha > \frac{1}{1-\beta} + \frac{1}{t}$ and $\beta > .525$ be any constants, and assume*

$$\max(r^\alpha, r^{2+1/t}t^2) = O(n^{1/t}).$$

Then for sufficiently large n, there is an explicit $(n, O((rt)^2\binom{n}{r}^{1/t}), r, t)$-exclusive set system. Further, broadcasting can be done in $\mathrm{poly}(r, t, \log n)$ time.

Proof: If the revoked set R is such that no two members share their ith coordinate for any $i > 1$, then if a point x doesn't appear in the broadcast its first two coordinates must agree with those of some $u \in R$. It follows from distinctness and our construction that $x = u$.

If points in R share their ith coordinate for some $i > 1$, we simply proceed as in lemma 3, ignoring the first coordinate. However, now we need the stronger assumption that $r^2t^2 \leq q$.

All that is left to show is that our three assumptions in the discussion above

1. $\max(r^{1+\epsilon}, t) \leq n^{1/t}$,
2. $r \leq \binom{n}{r}^{(1-\beta)/t}$,
3. $r^2t^2 \leq q$

follow from $\max(r^\alpha, r^{2+1/t}t^2) = O(n^{1/t})$. The derivations are straightforward and are omitted. ■

To apply the technique to the construction of theorem 7, we proceed as before, ignoring the first coordinate. The only assumption in the proof of theorem 11 that changes is the third one, which is now $r^2t = O(q)$. One can now show,

Theorem 12 *Let $\alpha > \frac{1}{1-\beta} + \frac{1}{t}$ and $\beta > .525$ be any constants, and assume $\max(r^\alpha, r^{2+1/t}t) = O(n^{1/t})$. For sufficiently large n, there is an explicit $(n, O(r^2t\binom{n}{r}^{1/t}, r, t)$-exclusive set system. Further, broadcasting can be done in $\mathrm{poly}(r, t, \log n)$ time.*

To adapt theorem 9, we just need to change the third assumption to $r^2t = O(q^{1-\epsilon})$ for some $\epsilon > 0$. Indeed, as in the proof of lemma 8, it is not hard to show that the probability v that some $s \in S$ is not good for R can be bounded above by $q^{-\epsilon}$. By our assumption that $r^{1+\epsilon} \leq n^{1/t}$, we have $n/r = n^{\Omega(1)}$ so that $q^{-\epsilon} = n^{-\Omega(1/t)}$, and the proof of lemma 8 goes through (with larger constants).

Theorem 13 *Let $\alpha > \frac{1}{1-\beta} + \frac{1}{t}$, $\beta > .525$, and $\epsilon > 0$ be any constants, and assume that we have*

$$\max(r^\alpha, r^{2+(1-\epsilon)/t}t) < n^{(1-\epsilon)/t}.$$

Then there is an efficient algorithm that with probability $1-n^{-\Theta(r)}$, generates an $(n, O(rt\binom{n}{r}^{1/t}), r, t)$-exclusive set system. Broadcasting takes time $\mathrm{poly}(r, t, \log n)$.

3. The General System

We reduce the case of arbitrary n, r, t to the schemes of section 2. We construct many small exclusive set systems on different subsets of $[n]$ and take their union to obtain the final explicit exclusive set system. Each of the small systems will be constructed with parameters n_i, r_i, t_i satisfying the requirements of the schemes in section 2.

The size of our final system will be $\mathrm{poly}(r, t)n^{r/t}$, matching the lower bound up to the $\mathrm{poly}(r, t)$ factor and the optimizations in section 2.3. At the end of the section we show how to replace the $n^{r/t}$ term with $\binom{n}{r}^{1/t}$ when $r = O(t)$, and sketch how to improve the $\mathrm{poly}(r, t)$ factor.

We may assume that $|R| = r$ because for each $0 \leq i \leq r$ we can construct an exclusive set system for those R with $|R| = i$, and then take their union. The complexity is largest when $|R| = r$, so the union will be at most $r+1$ times larger. Define,

$$d = \Theta(\log n / \log r^2), \text{ and let } q = \Theta(r^2 d) \text{ be prime.}$$

Then for an appropriate choice of constants.

- $q^{d+1} \geq n$, and
- for *any* r degree-d polynomials in $\mathbb{F}_q[X]$ there is a point in $\mathbb{F}_q$ on which the polynomials all differ.

We construct q different coordinate systems. For each system we treat the set $[n]$ as a collection of distinct univariate polynomials and represent them by their evaluation on $d+1$ points in $\mathbb{F}_q$. To identify the coordinate system, we choose the evaluation on the field element i to be the first coordinate of members in the ith system, and when broadcasting to a set $[n] \setminus R$, we use a system for which the polynomials R all differ on their first coordinate.

Each of the q coordinate systems will correspond to a set system which is the union of q^2 *exclusive* set systems. When broadcasting to a set R, we choose a coordinate system for which each element of R has a different first coordinate in $\mathbb{F}_q$. For each coordinate system, for each interval

$[i,j]$ with $i \leq j \in \mathbb{F}_q$, we restrict to points whose first coordinate lies in $[i,j]$. Let ρ, τ be integer parameters to be determined. The idea is to partition the line $[1,q]$ into intervals each containing $r_i \in \{\rho-1, \rho\}$ elements of R and consuming $t_i \in \{\tau, \tau+1\}$ encryptions in the broadcast. For an appropriate choice of ρ and τ, this allows us to use the exclusive set systems of section 2 independently on each interval.

The number of points in a given interval may be as small as $\rho-1$, but is certainly less than n. We make the simplifying assumption that it is exactly n by artificially increasing the universe size. When we take the union over all of these set systems, we delete these extra points.

Since the numbers of sets for a given interval i has the form $\mathrm{poly}(r_i,t_i)n^{r_i/t_i}$ for $r_i \in \{\rho-1,\rho\}$ and $t_i \in \{\tau, \tau+1\}$, we will choose $r_i/t_i \approx r/t$. If it is already the case that $r^4 t < n^{\frac{1}{2t}}$, we may use the scheme of either theorem 12 or theorem 13. Otherwise, if possible, we will choose ρ and τ to satisfy

$$\Omega\left(n^{\frac{1}{4\tau}}\right) \;\leq\; \rho^4\,\tau \;<\; n^{\frac{1}{2\tau}}, \tag{1}$$

subject to the constraint

$$\frac{\rho-1}{\tau} \;<\; \frac{r}{t} \;\leq\; \frac{\rho}{\tau}. \tag{2}$$

The idea is that when bounding the total number of sets, this will help us pull out an extra factor of $n^{1/\tau}$ down into the $\mathrm{poly}(r,t)$ factor. Note that due to integrality constraints our bound may be of the form $\mathrm{poly}(r,t)n^{r/t+1/\tau}$, and thus we will need $n^{1/\tau}$ to be polynomial in r and t. We now give an efficient algorithm Generate for doing this. The basic idea behind Generate is to keep decreasing τ and ρ until they satisfy constraints 1 and 2, noting that as ρ and τ decrease together, it is more likely that constraint 1 holds.

Generate(r,t):

1. Set integer variables $\rho = r$ and $\tau = t$.
2. $i = 2$.
3. While $\tau > 1$,
   (a) If $\rho^4\tau < n^{\frac{1}{2\tau}}$, then exit.
   (b) Else,
      i. $\tau = \lfloor t/i \rfloor$.
      ii. Choose ρ so that $\frac{\rho-1}{\tau} < \frac{r}{t} \leq \frac{\rho}{\tau}$.
      iii. $i = i+1$.

Lemma 14 *If* Generate *outputs* $(\rho,\tau) \neq (r,t)$ *and* $\tau \neq 1$, *then* ρ, τ *satisfy constraints 1 and 2.*

Proof: Suppose $\tau \neq 1$. Then in some iteration we have $\rho^4\tau < n^{1/(2\tau)}$. If this occurs in the first iteration, then we have $(\rho,\tau) = (r,t)$. Otherwise, consider the last time for which $\rho^4\tau \geq n^{1/(2\tau)}$. Suppose $\tau = \lfloor t/i \rfloor$, and let $\tau' = \lfloor t/(i+1) \rfloor$ be the value of τ in the next iteration. Note that $\tau, \tau' > 1$. Then $\tau'/\tau = \lfloor t/(i+1) \rfloor / \lfloor t/i \rfloor$. Suppose $\lfloor t/(i+1) \rfloor = c$. Then

$$t \leq (c+1)(i+1) - 1,$$

so that

$$\lfloor t/i \rfloor \leq \lfloor \left(\frac{i+1}{i}\right)(c+1) - \frac{1}{i} \rfloor = \lfloor c+1+c/i \rfloor.$$

Thus,

$$\tau'/\tau \geq c/(c+1+c/i) \geq 1/(1+1/c+1/i) \geq 1/2,$$

since $c, i > 1$ are integers. We also claim that $\rho' \geq \rho/4$. Indeed, if $\rho \leq 4$, this follows from the fact that ρ' is a positive integer. On the other hand, if for $\rho > 4$ we had $\rho' < \rho/4$, then

$$\frac{\rho'}{\tau'} \;<\; \frac{\rho/4}{\tau/2} \;=\; \frac{\rho/2}{\tau} \;<\; \frac{\rho-1}{\tau} \;\leq\; \frac{r}{t},$$

contradicting constraint 2, which holds because of step 3(b)ii. Thus,

$$(\rho')^4\tau' \geq \frac{\rho^4\tau}{2\cdot 4^4} \;\geq\; \frac{n^{\frac{1}{2\tau}}}{2\cdot 4^4} = \Omega(n^{\frac{1}{4\tau'}}),$$

which shows that constraint 1 holds. ■

We can now, for instance, apply the explicit construction of theorem 12.

Theorem 15 *Let n, r, t be positive integers and suppose n is sufficiently large. There is an explicit*

$$(n, \mathrm{poly}(r,t)n^{r/t}, r, t)$$

-exclusive set system. Broadcasting can be done in $\mathrm{poly}(r,t,\log n)$ *time.*

Proof: To broadcast with a revoked set R, we first find a coordinate system for which the polynomials R all differ on their first coordinate. Then, we run Generate to obtain ρ and τ. If $(\rho,\tau) = (r,t)$, then we run the protocol of theorem 12, which gives the desired complexity.

Otherwise, we arbitrarily partition the line $[1,q]$ into intervals such that each interval contains $\rho-1$ or ρ revoked points. We also arbitrarily allocate either τ or $\tau+1$ encryptions in the broadcast to each interval, subject to the constraint that their sum is t.

Finally, for each interval, we use the exclusive set system of theorem 12 with n points, $\rho-1$ or ρ revoked points, and broadcast size τ or $\tau+1$. Since the intervals for a given coordinate system concern disjoint points, the correctness of this scheme follows from that of the exclusive set system

of theorem 12. Moreover, since $\rho \leq r$ and $\tau \leq t$, it is easy to see that broadcasting can be done in $\mathrm{poly}(r, t, \log n)$ time given that the schemes used in each interval have this property.

It remains to derive the size of our set system. There are q coordinate systems. For each system, there are q^2 intervals. Each interval corresponds to an exclusive set system generated by theorem 12 on n points with the number of revoked points $r_i \in \{\rho - 1, \rho\}$ and the broadcast size $t_i \in \{\tau, \tau + 1\}$. To analyze the number of sets per interval, we divide the output of Generate into two cases (recall that at this point we need only consider $(\rho, \tau) \neq (r, t)$).

Case 1: $\tau \neq 1$. In this case the number of sets per interval is at most

$$\begin{aligned} \mathrm{poly}(r_i, t_i) n^{r_i/t_i} &\leq \mathrm{poly}(r, t) n^{\rho/\tau} \leq \\ \mathrm{poly}(r, t) n^{r/t} n^{1/\tau} &\leq \mathrm{poly}(r, t) n^{r/t}, \end{aligned}$$

where the second inequality follows by constraint 2 and the third by constraint 1.

Case 2: $\tau = 1$. Then by the analysis in lemma 14, we have $\rho^4 = \Omega(n^{1/4})$. We have the same sequence of inequalities as in case 1, where the second inequality follows by constraint 2, but now the third inequality follows from the fact that $r \geq r_i = n^{\Omega(1)}$, so that $\mathrm{poly}(r) = n^{1/\tau}$.

Thus, we have $q^3 \mathrm{poly}(r, t) n^{r/t} = \mathrm{poly}(r, t) n^{r/t}$ sets. ■

Corollary 16 *Let n, r, t be positive integers with $r = O(t)$ and n sufficiently large. There is an explicit*

$$(n, \mathrm{poly}(r, t) \binom{n}{r}^{1/t}, r, t)$$

-exclusive system with $\mathrm{poly}(r, t, \log n)$ broadcasting time.

Proof: Because we have $r = O(t)$,

$$n^{r/t} = r^{r/t} \left(\frac{n}{r}\right)^{r/t} \leq r^{r/t} \binom{n}{r}^{1/t} \leq \mathrm{poly}(r) \binom{n}{r}^{1/t},$$

and thus the system above has $\mathrm{poly}(r, t) \binom{n}{r}^{1/t}$ sets. ■

We defer a formal analysis which reduces the degree of the $\mathrm{poly}(r, t)$ factor to the full version of this paper. The idea there is to use randomness to spread the n points evenly across the line $[1, q]$ so that when broadcasting each interval in the partition of $[1, q]$ has roughly the same number of points. Although we use randomness, this randomness will preserve efficient broadcast and description size. We also modify constraint 1, basing it on theorem 13 instead of theorem 12, leading to slightly better bounds. We suspect that similar techniques can be used to achieve an exclusive set system with $\mathrm{poly}(r, t) \binom{n}{r}^{1/t}$ sets even when $r = \omega(t)$, though we have not worked out the details.

Open Questions: Can one achieve $\mathrm{poly}(r, t) \binom{n}{r}^{1/t}$ family size when $r = \omega(t)$ (versus our $\mathrm{poly}(r, t) n^{r/t}$)? For small r and t, is our $O(rt \binom{n}{r}^{1/t})$ bound on the family size tight? The known lower bound is $\Omega(t \binom{n}{r}^{1/t})$. How small can the $\mathrm{poly}(r, t)$ factor be for general n, r, and t?

Acknowledgment: The third author thanks Andrew C. Yao and the students at Tsinghua University for support and hospitality while performing some of this research.

References

[1] W. Aiello, S. Lodha, and R. Ostrovsky. Fast digital revocation. In *Crypto*, pages 137–152, 1998.

[2] N. Alon, J. Bruck, J. Naor, M. Naor, and R. Roth. Construction of asymptotically good, low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on Information Theory*, 38 (192):509–516, 1992.

[3] R. Baker, G. Harman, and J. Pintz. The difference between consecutive primes ii. *London Math. Soc*, (3) 83:532–562, 2001.

[4] A. Fiat and M. Naor. Broadcast encryption. In *Crypto*, pages 480–491, 1993.

[5] E. Gafni, J. Staddon, and Y. Yin. Efficient methods for integrating traceability and broadcast encryption. In *Crypto*, pages 372–387, 1999.

[6] D. Halevy and A. Shamir. The lsd broadcast encryption scheme. In *Crypto*, pages 47–60, 2002.

[7] J. Hastad, S. Jukna, and P. Pudlak. Top-down linear bounds for depth-three circuits. *Computational Complexity*, 5:99–112, 1995.

[8] S. Jukna. *Extremal Combinatorics with Applications in Computer Science*. Springer-Verlag, 2001.

[9] R. Kumar, S. Rajagopalan, and A. Sahai. Coding constructions for blacklisting problems without computational assumptions. In *Crypto*, pages 609–623, 1999.

[10] R. Kumar and A. Russell. A note on the set systems used for broadcast encryption. In *SODA*, pages 470–471, 2003.

[11] J. Lotspiech, D. Naor, and N. Naor. Revocation and tracing schemes for stateless receivers. In *Crypto*, pages 41–62, 2000.

[12] A. Lubotzky, R. Phillips, and P. Sarnak. Explicit expanders and the ramanujan conjectures. *Combinatorica*, 8:261–277, 1988.

[13] M. Luby and J. Staddon. Combinatorial bounds for broadcast encryption. In *Eurocrypt*, 1998.

[14] D. Stinson, T. Trung, and R. Wei. Secure fremeproof codes, key distribution patters, group testing algorithms and related structures. *Journal of Statistical Planning and Inference*, 86:595–617, 2000.

[15] O. Telelis and V. Zissimopoulos. Absolute o(log m) error in approximating random set covering: an average case analysis. *Information Processing Letters*, 94:4:171–177, 2005.

[16] B. Weidge. Statistical methods in algorithm design and analysis. *PhD thesis, Carnegie Mellon University*, 1978.

A. A lower bound on the keys per user

Theorem 12 of [13] gives a lower bound on the number of keys per user for the broadcast encryption schemes we consider. It is also a lower bound on the number of sets in an (n, k, r, t)-exclusive set system that some integer i must occur in, and we restate their theorem in this language.

Theorem 17 *([13], restated) In any (n, k, r, t)-exclusive set system, there is an integer i which occurs in at least $(\binom{n}{r}^{1/t} - 1)/(rt)$ sets.*

Whereas [13] use the Sunflower lemma [8] to prove their bound, we use the following relaxation of a sunflower to a flower, and a corresponding lemma.

Definition 18 *A set system $\mathcal{F} = \{F_1, \ldots, F_M\}$ is a flower with core Y and k petals if there is no set of size less than k which intersects every element in the family $\mathcal{F}_Y = \{F - Y \mid F \in \mathcal{F},\ Y \subseteq F\}$.*

Note that both the set Y and the intersecting set are allowed to be arbitrary. We use lemma 7.3[3] of [8], which was first discovered in [7]:

Lemma 19 *Let $\mathcal{D}$ be a family of sets, each of cardinality at most s. If $|\mathcal{D}| > (k-1)^s$, then $\mathcal{D}$ contains a flower with k petals.*

Theorem 20 *In any (n, k, r, t)-exclusive set system, there is an integer i which occurs in at least $(\binom{n}{r}^{1/t} - 1)/r$ sets.*

Proof: Let $\mathcal{C}$ be an (n, k, r, t)-exclusive set system. Define the collection $\mathcal{C}'$ as follows. For each set $[n] \setminus R$ with $|R| = r$, find a set T of at most t sets in $\mathcal{C}$ whose union equals $[n] \setminus R$. Add T to $\mathcal{C}'$. It follows that $\mathcal{C}'$ is a collection of size $\binom{n}{r} > (\binom{n}{r}^{1/t} - 1)^t$, each of whose elements is a set of size at most t.

It follows that $\mathcal{C}'$ contains a flower $\mathcal{F}$ with some core Y and $\binom{n}{r}^{1/t}$ petals. Fix some set $F \in \mathcal{F}$. Then the union of sets in F is $[n] \setminus R$ for some set R of size r. We claim that for any $F' \neq F$ with $F' \in \mathcal{F}$, $F' - Y$ contains some set $S_{F'}$ which intersects R. Indeed, the union of elements in F' equals $[n] \setminus R'$ for some $R' \neq R$ with $|R'| = |R|$, so F' contains a set $S_{F'}$ that contains at least one element of R, and this set cannot occur in Y since Y is a subset of F and the union of sets in F does not intersect R.

Now suppose the number of sets in $\mathcal{C}$ intersecting R were less than $\binom{n}{r}^{1/t} - 1$. Then, taking all of these sets together with any element of $F - Y$ gives a set of less than $\binom{n}{r}^{1/t}$ sets which intersects $F'' - Y$ for every $F'' \in \mathcal{F}$. This contradicts the fact that $\mathcal{F}$ is a flower with core Y and $\binom{n}{r}^{1/t}$ petals. Hence, there are at least $\binom{n}{r}^{1/t} - 1$ sets in $\mathcal{C}$ which intersect R. It follows that some element in R occurs in $(\binom{n}{r}^{1/t} - 1)/r$ sets. This concludes the proof. ∎

[3] We note that it is easy to modify the proof to handle families of sets each of cardinality *at most* s, even though the original lemma is stated for sets of size exactly s.

B. Intuition for the scheme in section 2

Here we give some intuition for the basic scheme of section 2. Suppose that $n^{1/t}$ is an integer (so in particular, this intuition only works for $t = O(\log n)$), and that we only consider sets R of size exactly r, for some $r = o(n^{1/t})$. Then we can associate $[n]$ with points in $[n^{1/t}]^t$. Suppose, for the moment, we are given a set R such that for every coordinate i, the points in R have distinct ith coordinates. Then the following construction has size $O(t\binom{n^{1/t}}{r} n^{r/t})$, which for small r, t, is off by about a factor of $\binom{n^{1/t}}{r} \approx \binom{n}{r}^{1/t}$ from that in section 2.

For every choice of r distinct 1st coordinates, we form a set which includes all points except those which have one of these r distinct 1st coordinates. The number of these sets is $\binom{n^{1/t}}{r}$. Next, for each coordinate i, $1 \leq i < t$, we form a set by choosing r pairs of ith and $(i+1)$st coordinates (c_j, d_j), where the c_j are all distinct and the d_j are all distinct. We exclude exactly those points which have ith coordinate and $(i+1)$st coordinate equal to some (c_j, d_j) pair. The number of these sets is $\binom{n^{1/t}}{r} n^{1/t}(n^{1/t} - 1) \cdots (n^{1/t} - (r-1))$. So in total, the family size is $O(tn^{r/t}\binom{n^{1/t}}{r})$.

Then it is easy to find t sets to exclude a given set R with distinct values for every coordinate i, and these t sets are unique. Conversely, a simple induction shows that any other point is included in one of these t sets.

The main idea of the first scheme in section 2 is to get around this roughly quadratic blowup in size by using polynomials. The source of the blowup is that we choose r values for coordinate i, and map them to r values for coordinate $i+1$, while all other values for the ith and $(i+1)$st coordinates are *unconstrained.* In the polynomial approach, we also choose r values for coordinate i and map them to r values for coordinate $(i+1)$, but *one polynomial can be used to implement many of these "sets of pairs" constraints at once.* This is because the polynomial simultaneously constrains the other coordinates. Thus, we still have $\binom{n^{1/t}}{r}$ sets for the 1st coordinate, but to implement constraints between coordinate i and coordinate $i+1$, we only have roughly $n^{r/t}$ sets, corresponding to the number of degree $r-1$ polynomials over a field of size roughly $n^{1/t}$.

We then proceed as in section 2 to handle the case when points in R agree on their ith coordinates for certain $i \in [t]$.