COMBINATORIAL DESIGNS FOR KEY DISTRIBUTION AND SECURE RE-KEYING IN GROUP COMMUNICATION SYSTEMS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Nathaniel Karst

August 2011

COMBINATORIAL DESIGNS FOR KEY DISTRIBUTION AND SECURE RE-KEYING IN GROUP COMMUNICATION SYSTEMS

Nathaniel Karst, Ph.D.

Cornell University 2011

Combinatorial t-designs are a class of highly regular set systems subject to interesting incidence conditions. These objects have been found to be very useful in applications ranging from tournament scheduling to traffic routing in communication networks. In the work presented here, we use a particularly nicely-structured class of combinatorial designs, known as symmetric 2-designs, to solve a problem concerning secure re-keying in a wireless communication system after the ejection of one or more users from the network. We show that employing a symmetric 2-design as a key distribution in this type of system provides a number of benefits, including collusion prevention and provably light loads for the base station to execute necessary secure re-keying operations. We show that a class of symmetric 2-design key distribution allows for minimal re-keying procedures after multiple simultaneous user ejections and that this problem is NP-hard for arbitrary key distributions. For cases where the structure of symmetric 2-designs is insufficient to make these strong claims, we present a novel algorithm for identifying a collection of keys sufficient to re-key a network after any number of ejections. We provide simulation results to show that for symmetric 2-design key distributions this algorithm performs significantly better than existing solutions. To make these guarantees, we draw connections between combinatorial designs, cover-free families and various key distribution methodologies. We conclude by presenting a sample application of this machinery, namely the advanced metering infrastructure being

deployed to monitor end-user electricity consumption as part of the smart grid. The wireless sensors employed in this scheme have tight constraints on memory, computation and power, and so symmetric encryption is a natural choice for data security. The distribution of the cryptographic keys necessary for these operations is difficult, and fluid group membership further complicates the problem. The widespread adoption of AMI has the potential to significantly increase the efficiency of the power distribution network. The acceptability of AMI to consumers is directly tied to their perceived security; a robust infrastructure is necessary to assure consumers of the protection of their personal information.

BIOGRAPHICAL SKETCH

Nathaniel Joseph Karst was born to Michael and Anita in Kettering, Ohio in 1984. He grew up outside Memphis, Tennessee with his siblings Andrew, Casey and Emma. He graduated in May 2007 from Franklin W. Olin College of Engineering in Needham, Massachusetts with a B.S. in electrical and computer engineering. He completed his doctorate in applied mathematics in June 2011. He will go on to teach mathematics at Babson College. His interests include brewing, cooking, David Attenborough documentaries, gardening, hiking, science fiction and yoga.

To the many and varied teachers in my life

ACKNOWLEDGEMENTS

I've had an enormous number of people encourage and help me during my education. In the beginning, my parents' only academic requirement, that I always do my best, was as liberating as it was terrifying. Ben Cook, Karen Garrison and Sharon Proffer sparked my interest in so many subjects, including science and mathematics. Sarah Adams and John Geddes introduced me to the real pleasures of mathematical research and teaching. In doing so, they set me on my current trajectory; I really can't thank either of them enough.

I'd like to thank my committee members Ken Brown (mathematics) and Adam Bojanczyk (electrical engineering) for their helpful conversations and guidance. I'd especially like to thank my adviser Stephen Wicker. The freedom he has given me to follow my nose on a problem has been wonderful. But I think that the freedom to make my own mistakes might be even more valuable. He has always being to there to help put things back on track after I derail them. I appreciate the Center for Applied Mathematics giving me the flexibility to pursue my professional interests, even as they evolve, and for so consistently supporting inter-disciplinary research.

Thanks most of all to my family, Mom, Dad, Be, G&G, Andy, Casey and Emma for the constant encouragement and to my best buds Ben and David for all the excellent times that have made grad school so special. You all've made the tough times bearable and the easy times wonderful. Your support has meant the world to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

T-DESIGNS

1.1 Introduction

Design theory is a branch of combinatorics which is concerned with various forms of incidence among subsets drawn from some underlying support set. Over the course of its development, design theory has been shown to have remarkable connections to diverse areas of mathematics including number theory, finite geometry, linear error-correcting codes and graph theory. Given the breadth of its reach, it may be surprising to learn that the earliest roots of design theory find themselves among the recreational mathematics of the mid-1800s. Kirkman posed perhaps the first problem in design theory in an 1847 edition of *Lady and Gentleman's Diary*:

> "Fifteen young ladies in a school walk out three abreast for seven days in succession: it is required to arrange them daily so that no two shall walk twice abreast."

For obvious reasons, this statement is known as *Kirkman's schoolgirl problem.* Few branches of pure mathematics are conceived in magazines, and the fact that design theory bucks this trend is a testament both to its often seemingly simple problem statements and to its utility in modeling real-world situations. As with all problems in the field, Kirkman's school problem is concerned with a collection of elements and the exact way in which subsets of those elements are related to one another. For instance, here Kirkman stipulates that any pair of school girls appears in at most one row. If we take the school girls as our elements and define incidence to mean being located together in a row, then any pair of elements is incident at most once. To put this qualitative definition on firm mathematical footing, a more basic definition is first needed.

Definition 1.1.1. *A (finite)* **set system** *is an ordered pair $(X, \mathcal{B})$ in which X is a (finite) set and $\mathcal{B}$ is a collection of subsets of X. We will call the elements of X* **points** *and the sets of $\mathcal{B}$* **blocks**. *We say a set system is* **uniform** *if all blocks have the same cardinality and* **regular** *if all points occur in the same number of blocks.*

Note that this definition does not specify that the collection $\mathcal{B}$ is itself a set; there may exist identical blocks in $\mathcal{B}$. Set systems are an incredibly diverse and flexible class of mathematical objects. The pair $(V, \mathcal{E})$ consisting of the vertices and (hyper)edges of a (hyper)graph constitute a set system. The matroid $(M, \mathcal{I})$ consisting of a ground set and a collection of independent sets also defines a set system. A probability space $(\Omega, \mathcal{F}, P)$ can be thought of a set system in which the collection of subsets $\mathcal{F} \subseteq 2^{\Omega}$ forms a σ-algebra, together with a probability measure P. Most importantly for the work presented here is the class of set systems known as t-designs.

1.2 t-designs

The combinatorial objects known as t-designs grew out of the seminal works of 19th century mathematicians such as Kirkman and Steiner. These designs are most fundamentally uniform and regular set systems with an additional incidence condition.

Definition 1.2.1. *A set system $(X, \mathcal{B})$ is a t-(v, b, r, k, λ)* **design** *if*

(1) $|X| = v$

(2) $|\mathcal{B}| = b$

(3) every $x \in X$ *occurs in exactly* r *sets in* $\mathcal{B}$

(4) $|B| = k$ *for all* $B \in \mathcal{B}$

(5) every t*-subset of* X *appears in exactly* λ *blocks in* $\mathcal{B}$*.*

We call r and k the *replication number* and *block size*, respectively. Note that even here the property that $\mathcal{B}$ is a set, that is, that there are no repeated blocks, is not required *a priori*. Designs containing no repeated blocks are often referred to as *simple*. For the balance of this work, we will assume that every design which we introduce is simple. We will also omit from consideration the *complete design* which is composed of all k-subsets of X with $t \leq k$.

Conditions (1)-(4) impose uniformity and regularity on the set system. It is condition (5) that puts t-designs apart from all other set systems, and indeed it is not easily satisfied. Designs with $t = 1$ are redundantly defined, as condition (3) would imply that $r = \lambda$ in this case. We will therefore always assume that $t > 1$. Designs with $t = 2$ have been relatively well-studied and will be main mathematical tool used in the applications featured here. Some infinite classes of t-designs with small λ are known for $t > 2$. For instance, there is a 3-$(q^2 + 1, q + 1, 1)$ design, a so-called *Möbius* or *inversive plane*, for every prime power q [11]. However, no t-(v, b, r, k, λ) design is known to exist for any $t > 5$ and $\lambda < 4$ [11]. Determining how many if any t-$(v, k, 1)$ designs exist for large t is one of the largest open problems in design theory.

The parameters of a t-(v, b, r, k, λ) design are not independent of one another. Simple algebraic equations allow us to write any two of the parenthetical parameters in terms of the remaining three. The proofs of these well-known theorems are included here both for completeness and so that the reasoning and methods found in the original work presented later will have proper motivation.

Result 1.2.1. *The parameters of a t-(v, b, r, k, λ) design $(X, \mathcal{B})$ satisfy $bk = vr$.*

Proof. We will count the total number of points in all blocks in two ways. There are b blocks in $\mathcal{B}$, each containing exactly k points. Additionally, there are v points in X, each occurring in exactly r blocks. Hence, the total number of points in the design is $bk = vr$. □

Result 1.2.2. *The parameters of a t-(v, b, r, k, λ) design $(X, \mathcal{B})$ satisfy $\lambda(v-1) = r(k-1)$.*

Proof. Fix $x \in X$. We will count the number of pairs (x, y), $y \in X$ and $y \neq x$, occurring in all blocks in $\mathcal{B}$. For all $v - 1$ choices of suitable y there exist exactly λ blocks containing both x and y. For the right side of the equality, the point x occurs in exactly r blocks, and in each there are exactly $k - 1$ other distinct points. □

Despite the fact that Result 1.2.1 and Result 1.2.2 imply that there exists a more concise notation which describes any t-(v, b, r, k, λ) design, we will continue to use this expanded version for clarity's sake, except in one particular class that will be introduced below.

Example 1.2.1 ([11]). *Let $X = \{a, b, \ldots, o\}$ and define*

$$\begin{aligned}\mathcal{B} = \{&abc, djn, ehm, fio, gkl\\ &ahi, beg, cmn, dko, fjl\\ &ajk, bmo, cef, dhl, gin\\ &ade, bln, cij, fkm, gho\\ &afg, bhj, clo, dim, ekn\\ &alm, bik, cdg, ejo, fhn\\ &ano, bdf, chk, eil, gjm\}.\end{aligned} \tag{1.1}$$

The pair $(X, \mathcal{B})$ is a 2-$(15, 35, 7, 3, 1)$ design. If we associate each $a, b, \ldots, o$ distinctly with one of 15 school girls, take the blocks of $\mathcal{B}$ as the rows of girls and take each of the rows of this array of $\mathcal{B}$ as one of the days of the week, then this presentation of $(X, \mathcal{B})$ is a solution to Kirkman's schoolgirl problem.

Example 1.2.2. *Let $X = \{0, 1, \ldots, 6\}$ and define*

$$\begin{aligned}\mathcal{B} = \{&\{0, 1, 3\}, \{1, 2, 4\},\{2, 3, 5\}, \{3, 4, 6\},\\ &\{4, 5, 0\}, \{5,6, 1\}, \{6, 0, 2\}\}.\end{aligned} \tag{1.2}$$

One can verify that $(X, \mathcal{B})$ is 2-$(7, 7, 3, 3, 1)$ design. This design can be identified with the projective geometry PG(2,2) which will be discussed in Section 1.3.1. It is commonly known as the Fano plane.

We note that in this particular case the number of points v equals the number of blocks b, and the replication number r equals the block size k. This additional structure is indicative of a larger class of 2-(v, v, k, k, λ) designs which have been by far the most extensively investigated in design theory literature.

1.3 Symmetric 2-Designs

Fischer's inequality states that for any 2-(v, b, r, k, λ) design the number of blocks is at least the number of points, that is $b \geq v$. The class of 2-designs which meet this bound with equality are called symmetric.

Definition 1.3.1. *A* **symmetric 2-**(v, k, λ) **design** *of order* $q = k - \lambda$ *is a 2-*(v, v, k, k, λ) *design.*

We will sometimes refer to a symmetric 2-(v, k, λ) design simply as a symmetric design if the parameters are either clear or irrelevant. Here the term "symmetric" refers not to any geometrical property necessarily, but rather to the equivalence of the conditions on the numbers of blocks and points and on the block size and the replication number. For reasons that we will not delve into here, some authors prefer to use the term *square* to describe 2-(v, v, k, k, λ) designs. Symmetric designs have yet another interesting (and useful) equivalence between blocks and points: any two blocks have intersection cardinality λ. In fact, these symmetries are themselves equivalent.

Theorem 1.3.1. *Given a 2-*(v, b, r, k, λ) *design* $(X, \mathcal{B})$*, the following are equivalent:*

(1) $v = b$*,*

(2) $r = k$*,*

(3) any two blocks share exactly λ *points.*

Proof. Result 1.2.1 shows the bidirectional equivalence (1) $\Leftrightarrow$ (2). To see, (3) $\Rightarrow$ (1), consider a set system constructed in the following way. For each $x \in X$, define

$B'_x = \{B \in \mathcal{B} : x \in B\}$, and let $\mathcal{B}' = \{B'_x\}_{x \in X}$ be the collection of these sets with x ranging over X. Then the pair $(\mathcal{B}, \mathcal{B}')$ form a 2-design by hypothesis (3); it is easy to verify that 2-design axioms hold. Fisher's inequality then implies that $|\mathcal{B}| \leq |\mathcal{B}'| = |X|$. But since $(X, \mathcal{B})$ is a 2-design, Fisher's inequality also implies that $|X| \leq |\mathcal{B}|$ and hence $v = |X| = |\mathcal{B}| = b$.

For (1) $\Rightarrow$ (3), fix a block $B \in \mathcal{B}$ and let λ_i be the number of points shared between B and $B_i \in \{B_1, B_2, \ldots, B_{v-1}\}$. (We have assumed (1), as well, since we have shown (1) $\Rightarrow$ (2) previously.) Each of the k points of B occurs in exactly $k-1$ other blocks in $\mathcal{B}$, leading to the equation

$$k(k-1) = \sum_{i=1}^{v-1} \lambda_i. \tag{1.3}$$

Then by Result 1.2.2 with $r = k$, we have

$$\lambda = \frac{1}{v-1} \sum_{i=1}^{v-1} \lambda_i, \tag{1.4}$$

so that the first moment of any collection $\{\lambda_i\}$ is λ. Suppose that $\lambda = 1$. Each λ_i is a non-negative integer, and since $\lambda = 1$, every two points occurs in exactly 1 block so that $\lambda_i < 2$ for all $i = 1, 2, \ldots, v-1$. Then Equation 1.4 implies that $\lambda_i = 1$ for all $i = 1, 2, \ldots, v-1$. Since B was chosen arbitrarily, the desired result holds for $\lambda = 1$.

Now suppose that $\lambda > 1$. Each of the $\binom{k}{2}$ pairs of points from B occurs in exactly $\lambda - 1$ other blocks, giving the relation

$$\begin{aligned} (\lambda-1)\binom{k}{2} &= \sum_{i=1}^{v-1} \binom{\lambda_i}{2} && (1.5) \\ (\lambda-1)k(k-1) &= \sum_{i=1}^{v-1} \lambda_i(\lambda_i - 1). && (1.6) \end{aligned}$$

Again using Result 1.2.2, we arrive at

$$\lambda(\lambda - 1) = \frac{1}{v-1}\sum_{i=1}^{v-1}\lambda_i(\lambda_i - 1) \tag{1.7}$$

$$\lambda^2 = \frac{1}{v-1}\sum_{i=1}^{v-1}\lambda_i^2, \tag{1.8}$$

where the final line follows from the first moment result above. Hence, the second (non-central) moment of any collection $\{\lambda_i\}$ is λ^2. The first and second moment equations of $\{\lambda_i\}$ are enough to show that $\lambda_i = \lambda$ for all $i = 1, 2, \ldots, v-1$. We can define an auxiliary integer-valued variable $-\lambda \leq \delta_i \leq k - \lambda$ for each λ_i so that

$$\lambda_i = \lambda + \delta_i. \tag{1.9}$$

Then Equation 1.4 implies that $\sum_{i=1}^{v-1}\delta_i = 0$. After substitution by the auxiliary variable Equation 1.8 reads

$$\lambda^2 = \frac{1}{v-1}\sum_{i=1}^{v-1}\lambda_i^2 \tag{1.10}$$

$$= \frac{1}{v-1}\sum_{i=1}^{v-1}(\lambda + \delta_i)^2 \tag{1.11}$$

$$= \frac{1}{v-1}\sum_{i=1}^{v-1}(\lambda^2 + 2\lambda\delta_i + \delta_i^2) \tag{1.12}$$

$$= \lambda^2 + \frac{1}{v-1}\sum_{i=1}^{v-1}\delta_i^2, \tag{1.13}$$

which holds only if $\delta_i = 0$ for all $i = 1, 2, \ldots, v-1$. Hence, the arbitrarily chosen block B shares exactly λ points with any other block, and so any two blocks share exactly λ points. Having shown (1) ⇔ (2) and (1) ⇔ (3), the desired result has been proved. □

To the author's knowledge, the reasoning showing (1) ⇒ (3) is novel. The standard technique requires the introduction of incidence matrices and relies heavily on linear algebraic machinery. While powerful in its own right, this traditional approach does intentionally distance itself from the combinatorial nature of the problem. The new approach maintains contact with the combinatorial underpinnings of the subject and shows that identical results are within reach using more elementary tools.

The most powerful existence criteria concerning symmetric designs are the celebrated non-existence results due to Bruck and Ryser, who together proved the result for $\lambda = 1$ [5], and Ryser and Chowla, who later extended the result for the cases where $\lambda > 1$ [9].

Result 1.3.1 (Bruck-Ryser-Chowla)**.** *If a symmetric 2-(v, k, λ) design exists with v even, then $k - \lambda$ is a square. If a symmetric 2-(v, k, λ) design exists with v odd, then the equation $x^2 = (k - \lambda)y^2 + (-1)^{(v-1)/2}\lambda z^2$ has a nontrivial solution in the integers.*

The proof of this theorem is lengthy not especially informative to the material covered in this text; it is available in most textbooks on design theory. It is a testament to the wide mathematical connections of design theory that a question involving only incidence between finite sets should involve the existence of solutions of a Diophantine equation. In 1989, Lam, Thiel and Swiercz provided the most recent advancement past the Bruck-Ryser-Chowla thereom by proving that there does not exist a symmetric 2-(121,11,1) design via a computerized search [24].

1.3.1 Projective Planes

Singer provided the most well-known construction algorithm for symmetric 2-designs. His method deals with vector space inclusion over a finite field. Designs generated in this fashion will be used in later as key distributions for group communications systems. We will see that their properties allow for powerful results concerning the ability to re-key the system after one or more user is ejected from the network.

Result 1.3.2 (Singer, [35]). *There exists a symmetric 2-$(q^2+q+1, q+1, 1)$ design for every prime power q.*

Proof. Let V be the 3-dimensional vector space over the finite field $GF(q)$. Let the points of X be the 1-dimensional subspaces of V. For each 2-dimensional subspace W of V, define

$$B_W = \{U \setminus \{0\} : 0 \leq U \leq W \leq V\}, \tag{1.14}$$

where the symbol $\leq$ denotes vector space inclusion. Then define a set system by identifying points and blocks with each of the 1- and 2-dimensional subspaces of V, respectively, and associating point-block incidence with vector space inclusion, that is

$$\mathcal{B} = \{B_W : W \leq V, \dim(W) = 2\}. \tag{1.15}$$

It remains to verify that the conditions of a 2-design are satisfied.

Each 1-dimensional subspace of V contains exactly $q-1$ nonzero elements, and after excluding the zero vector, these 1-dimensional subspaces partition the vectors in V. Hence, the number of points is

$$\begin{aligned} |X| &= \frac{q^3-1}{q-1} && (1.16) \\ &= q^2+q+1. && (1.17) \end{aligned}$$

By construction, the number of blocks is identical to the number of distinct 2-dimensional subspaces of V. In a 3-dimensional vector space, each 2-dimensional subspace can be be uniquely identified by its 1-dimensional dual subspace. Hence,

$$\begin{aligned} |\mathcal{B}| &= |X| \quad (1.18) \\ &= q^2 + q + 1, \quad (1.19) \end{aligned}$$

and so $v = |X| = |\mathcal{B}| = b$.

To determine the number of 2-dimensional subspaces containing a fixed 1-dimensional subspace L, first note that the choice of any nonzero vector not in L uniquely determines a 2-dimensional subspace, namely their span in V. Next, note that two 2-dimensional subspaces of V which both include L must be disjoint outside of L. Hence, these these 2-dimensional subspaces partition the nonzero vectors of $V \setminus L$, and so the replication number is

$$\begin{aligned} r &= \frac{q^3 - q}{q^2 - q} \quad (1.20) \\ &= q + 1. \quad (1.21) \end{aligned}$$

Each 2-dimensional subspace W of V contains $q^2 - 1$ nonzero vectors, and the included 1-dimensional subspaces partition the nonzero vectors in W with each containing exactly $q - 1$ vectors. Hence,

$$\begin{aligned} |B_W| &= \frac{q^2 - 1}{q - 1} \quad (1.22) \\ &= q + 1 \quad (1.23) \end{aligned}$$

for every $B_W \in \mathcal{B}$. Therefore, we have $r = q + 1 = k$.

It remains to show to that the 2-design incidence condition is satisfied, namely that any two distinct 1-dimensional subspaces of V are contained in exactly $\lambda = 1$ 2-dimensional subspace of V. Since V is 3-dimensional, this inclusion relation is satisfied only by the span of the two 1-dimensional subspaces.

Consider two distinct 2-dimensional subspaces A and B of V. Then $\dim(A \cap B) < 2$ since the subspaces are distinct, but $\dim(A \cap B) \geq 1$ because

$$\dim(A \cup B) \leq \dim(V) \tag{1.24}$$

$$\dim(A) + \dim(B) - \dim(A \cap B) \leq \dim(V) \tag{1.25}$$

$$\dim(A \cap B) \geq 1. \tag{1.26}$$

Hence, we have $\dim(A \cap B) = 1$, and so any two blocks share exactly one point. □

Singer's construction can easily be extended to deal with the inclusion of one-dimensional subspaces in d-dimensional hyperplanes in the vector space $GF(q)^{d+1}$. Such a set system is called a projective geometry of order q over $GF(q)^{d+1}$ denoted $PG(d, q)$. For the work presented here, the projective planes $PG(2, q)$ will suffice. The symmetric 2-design featured in Example 1.2.2 follows from the Singer construction for a projective plane with $q = 2$.

1.3.2 Biplanes

While we have so far concentrated on symmetric designs with $\lambda = 1$, other symmetric designs do exist. These classes of symmetric designs are far less well understood that their projective plane counterparts. One class that will be featured in later applications are the symmetric designs with $\lambda = 2$.

Definition 1.3.2. *A* **biplane** *is symmetric 2-$(v, k, 2)$ design.*

There are only finitely namely biplanes known to exist, namely for $k = 4, 5, 6, 9, 11, 13$ [11]. Moreover, it is widely conjectured that for any $\lambda > 1$ there are only finitely many symmetric designs.

Example 1.3.1. *Let $X = \{0, 1, \ldots, 10\}$ and define*

$$\mathcal{B} = \{\{1,3,4,5,9\}.\{2,4,5,6,10\},\{3,5,6,7,0\},\{4,6,7,8,1\}, \tag{1.27}$$

$$\{5,7,8,9,2\},\{6,8,9,10,3\},\{7,9,10,0,4\},\{8,10,0,1,5\}, \tag{1.28}$$

$$\{9,0,1,2,6\},\{10,1,2,3,7\},\{0,2,3,4,8\}\}. \tag{1.29}$$

One can verify that $(X, \mathcal{B})$ is a biplane of order 3, a symmetric 2-(11,5,2) design.

1.4 Residuals of Symmetric Designs

It is often advantageous to think of a set system $(X, \mathcal{B})$ as a set system $(X', \mathcal{B}')$ with points and/or blocks removed. In this way, seemingly disparate classes of combinatorial objects can be linked. Existence results for one class can be applied the other, and one can investigate which properties of the original set system are preserved during the transformation. For our work here, the concept of a residual set system will be very useful.

Definition 1.4.1. *Let $(X, \mathcal{B})$ be a set system. The $\mathcal{E}$-**residual**, $\{E_1, E_2, \ldots, E_r\} = \mathcal{E} \subseteq \mathcal{B}$, is the set system $(X', \mathcal{B}')$ with*

$$X' = X \setminus \left(\bigcup_{i=1}^{r} E_i \right) \tag{1.30}$$

$$\mathcal{B}' = \left\{ C \setminus \left(\bigcup_{i=1}^{r} E_i \right) : C \in \mathcal{B} \setminus \mathcal{E} \right\}. \tag{1.31}$$

If $\mathcal{E}$ is a single set $B \in \mathcal{B}$, we will often abuse notation and identify the $\mathcal{E}$-residual of $(X, \mathcal{B})$ as the B-residual. If in addition the choice of B is understood or irrelevant, we will refer to the B-residual as simply the residual. Symmetric designs are unique among 2-designs in that their residuals are also 2-designs.

Result 1.4.1. *The B-residual of a symmetric 2-(v, k, λ) design $(X, \mathcal{B})$ is 2-$(v - k, v - 1, k, k - \lambda, \lambda)$ design for every $B \in \mathcal{B}$.*

Proof. Clearly there are $v - k$ points and $v - 1$ blocks in the residual. Every point not in B in the original design remains in exactly k blocks in the residual, so the replication number of the residual is k. By Theorem 1.3.1, every block $C \in \mathcal{B} \backslash \{B\}$ shares exactly λ points with B. Hence, every block in the residual contains exactly $k - \lambda$ points. Since every pair of points from X appear in exactly 2 blocks in $\mathcal{B}$, any pair of points in $X \setminus B$ remain in exactly λ blocks in the residual. Hence, all 2-design axioms are satisfied. □

Notice that we do not prove that residual of a symmetric 2-design is symmetric, and in fact they are not.

1.4.1 Affine Planes

The residual of a projective plane of order q is a 2-$(q^2, q^2 + q, q + 1, q, 1)$ design by Result 1.4.1. Such a design is known as an *affine plane* of order q. While an affine plane of order q can always be constructed as the residual of a projective plane of the same order, their structure is perhaps best elucidated by a construction due to Bose [3].

Result 1.4.2. *There exits a 2-$(q^2, q^2 + q, q + 1, q, 1)$ design for every prime power q.*

Proof. Let V be the 2-dimensional vector space of the finite field $GF(q)$. Associate the set of points X with the q^2 elements of V. For all one-dimensional subspaces

$W < V$ and $\alpha \in V$, define

$$W + \alpha = \{w + \alpha : w \in W\} \tag{1.32}$$

to be the translate of W by α. Define the collection of blocks to be all distinct translates of all 1-dimensional subspaces of V. Any translate has the same cardinality as the original 1-dimensional subspace, so the block size is q.

Considering the translates of a fixed 1-dimensional subspace as cosets of an additive group, it is clear that that any two of these translates are either identical or disjoint. Hence, there are $q^2/q = q$ distinct translates of any given 1-dimensional subspace. From this fact, we can derive both the replication number and the number of blocks in the candidate design.

The collection of 1-dimensional subspaces partition the set $V \setminus \{0\}$; any vector is included in its span, and any two linear subspaces intersect only at the zero vector. Hence, there are $(q^2 - 1)/(q - 1) = q + 1$ distinct 1-dimensional subspaces of V. Since any element $\beta \in V$ occurs in exactly one translate of each 1-dimensional subspace, namely $W + \beta$ for each $0 < W < V$, we find that the replication number is $|\{W : 0 < W < V\}| = q + 1$. Each linear subspace has exactly q distinct translates, and no two translates of two distinct linear subspaces can be identical. This implies that there are $(q + 1)q = q^2 + q$ total blocks.

So far we have derived all parenthetical parameters of the candidate design except λ, and it remains to show that any two vectors in V occur in exactly one block. Consider two vectors $\alpha, \beta \in V$, and let W be the span of $(\alpha - \beta)$. (Note that this construction is symmetric with respect to points, *i.e.* $\operatorname{span}(\alpha - \beta) =$

span$(\beta - \alpha)$.) Then we verify

$$\alpha = (\alpha - \beta) \cdot 1 + \beta \tag{1.33}$$

$$\beta = (\alpha - \beta) \cdot 0 + \beta, \tag{1.34}$$

so that both α and β are contained in $W + \beta$. To prove uniqueness, assume for contradiction that there exists another translate $U + \gamma$ containing both α and β. Then

$$u + \gamma = \alpha = w + \beta \tag{1.35}$$

$$u' + \gamma = \beta = w' + \beta. \tag{1.36}$$

Subtracting one equation from the other, we have

$$u - u' = w - w'. \tag{1.37}$$

But the left side is an element u'' of U, and ku'' is equal to the difference of elements in W for all $k \in GF(q)$. Varying k over the entirety of $GF(q)$ generates U. Hence, $U \subseteq W$. Considering the right side of the equation in the same manner shows $W \subseteq U$. With containment in both directions, we have shown the uniqueness of a translate containing any two vectors in V. □

Note that since the blocks of $\mathcal{B}$ represent the translates of vector subspaces, the blocks are either disjoint or have constant intersection cardinality.

As with projective planes, we can naturally extend the notion of the affine plane of order q to the affine geometry $AG(d, q)$ in which the points are the vectors of $GF(q)^d$, the blocks are correspond to translates of the $(d-1)$-dimensional subspaces of $GF(q)^d$, and point-block inclusion is taken set-wise [6].

Example 1.4.1. *Let $X = \{0, 1, 2, 5\}$ and define*

$$\begin{aligned} \mathcal{B} = \{&\{0,1\}, \{2,5\} \\ &\{0,2\}, \{1,5\} \\ &\{0,5\}, \{1,2\}\}. \end{aligned} \tag{1.38}$$

The pair $(X, \mathcal{B})$ is an affine plane of order 2. It is also the $\{3, 4, 6\}$-residual of the projective plane of order 2 featured in Example 1.2.2.

1.4.2 Residuals of Biplanes

The current understanding of biplanes is much less complete than that of their projective plane counterparts. The natural notions of geometry in the latter context allow for powerful existence and structural theorems about both the projective planes themselves and their residuals. Unfortunately, no such unifying theory has been discovered for biplanes, and so the results here remain piece-meal. For instance, biplane residuals have been documented for use in coding theory as in work of Key and Tonchev in [21]. But on the whole, these designs have received little attention due to a lack of firm intuitive footing.

CHAPTER 2

SECURE RE-KEYING IN GROUP COMMUNICATION SYSTEMS

2.1 Wireless Sensor Networks

In a wireless sensor network (WSN), a collection sensor nodes collect data and wirelessly communicate these data to other sensor nodes and/or to a base station in charge of aggregation and processing. The sensors nodes themselves are designed to be cheap and long-lasting so that a large number can be deployed simultaneously. The popular Zigbee wireless sensor platform supports network sizes of over 64,000 nodes each with a battery life of over 1 year [34]. As wireless sensor technologies have matured, WSNs have been used in a wide variety of applications, including military deployments, precision agriculture and health monitoring. For a recent survey of uses of WSNs, see [1]. There are two primary types of WSN, *hierarchical* and *distributed*. In a hierarchical WSN, a central authority (*e.g.*, base station or trust center) coordinates the network. The amount of control given to the central authority varies. We typically assume that the central authority has greater resources and is more secure than the sensor nodes. Hierarchical WSNs can be efficiently coordinated but are sensitive to the loss or compromise of the central authority, both of which result in network failure. In a distributed WSN, there is no central authority coordinating the network; information flows through the network in a distributed fashion. Distributed WSNs are more resilient to node loss but can be less efficient than their hierarchical counterparts. For our work here, we will consider a one-level hierarchical WSN, that is, a base station overseeing a collection of equally privileged sensor nodes. The base station will be tasked with broadcasting a group-wide information stream and coordinating encryption through the distribution of cryptographic keys. It may be the case that a sensor node cannot directly communicate with the base station. In this case, other nodes act as intermediaries on a *multi-hop path*. Hence, even in a hierarchical WSN, it is important that individual nodes maintain the ability to communicate with one another.

Data security during transmission is a priority in many WSN applications, and cryptography is an obvious solution. Public-key cryptography is a mainstay of modern wired communications, but the unique constraints found in WSNs make symmetric (private) key cryptography the preferred solution. While not an issue with wired devices, the computationally expensive mathematical procedures and relatively large cryptographic keys found in most public key cryptography schemes are out of place in this context. Symmetric key cryptography, in which a single cryptographic key is used in both encryption and decryption, is a better solution in most WSNs. The idea of symmetric cryptography is not a new one. Classic examples of symmetric key encryption range from the Caesar cipher to the Enigma machine. More recently, the National Institute of Standards and Technology (NIST) has approved the symmetric block ciphers included in the Advanced Encryption Standard (AES) for protecting data owned by the federal government [32]. The 128-bit flavor of the AES is the specified encryption of the Zigbee standard [2]. The relative simplicity of the encryption and decryption procedures does come at a cost, however. First, network connectivity becomes unassured, as any two users within range of one another can securely communicate if and only if both have access to at least one common cryptographic key. Second, an encrypted message is decipherable to *any* user owning the appropriate key, not just the intended recipient. In a full mesh Zigbee network operating in the commercial mode, every pair of users possesses a unique cryptographic key in order to address these concerns [30]; each user of n users then owns $O(n)$ keys. For even moderately sized networks, this number of keys represents a considerable memory overhead for the individual sensor nodes. When designing a reasonable solution to put in practice, we will have to make tradeoffs between the connectivity and the potential for eavesdropping within the network.

In a more general model for key distribution in a communication system employing symmetric encryption first introduced by Mitchell and Piper [29], each user u_i, $i = 1, 2, \ldots, b$, receives a collection of keys B_i, called a *key chain*, drawn without replacement from a network-wide *key pool* X of cardinality v. Limited memory at the sensor nodes upper bounds the number of keys a sensor node can store. It is typically advantageous for a sensor node to store as many keys as possible in order to maximize the probability that it will be able to securely communicate with its neighbors; it is therefore reasonable to assume that the key chain size k is constant across all users. Given these assumptions, the collection X of keys together with the collection of key chains $\mathcal{B} = \{B_1, B_2, \ldots, B_b\}$ forms a uniform set system $(X, \mathcal{B})$ as presented in Definition 1.1.1. The combinatorial properties of this set system can have enormous impacts on the efficacy of the sensor network.

Qualitatively speaking, large key chains relative to the size of the key pool result in higher probability that two users will share a common cryptographic key. In the extreme case, there is a single key in the key pool, and every sensor node has access to this key. Such a key distribution can be found in the Zigbee residential security mode, for instance [30]. This scheme provides a small degree of security at very low memory cost but is not preferred for critical applications, because if the single key is compromised, then no secure communication can take place between members of the network. In one of the first key distribution schemes, Eschenauer and Gligor proposed choosing the size k key chain of each of the b users of the WSN uniformly randomly and without replacement from a key pool of size v [19]. The probability that any two users can securely communicate is then determined by the relative sizes of k and v. In many applications, however, firmer guarantees as to network connectivity are needed. It is not surprising that the key distributions satisfying these tougher constraints have more mathematical structure than those of randomized schemes. The combinatorial t-designs investigated in Chapter 1 are a class of natural candidates due to their uniformity, regularity and incidence conditions. The class of 2-designs are both the most plentiful known class of t-designs and the best understood, and so we will focus our efforts on these special designs. As first document by Çamtepe and Yener [7], a 2-design $(X, \mathcal{B})$ can be used as a key distribution for a WSN by associating the points and blocks of the design with the keys and key chains of the key distribution, respectively. Unfortunately, it is frequently the case that two blocks in a arbitrary 2-design are disjoint, and so two users may not share a common cryptographic key with which to communicate securely. In this situation, their message must pass through one or more intermediaries along a multi-hop path. The presence of multi-hop paths increases both latency and overhead, as nodes are required to spend more resources communicating messages other than their own. Lee and Stinson put forward μ-common intersection designs which ensure that there exist at least μ distinct 2-hop paths between any two users which cannot communicate directly [25]. Mathematically, any pair of disjoint blocks of the design have non-empty intersections with at least μ common blocks. Chakrabarti, Maitra and Roy took a more probabilistic approach by assigning to each user the keys associated with multiple blocks from a 2-design [8]. In this way, many of the desirable properties of 2-designs are preserved while increasing the probability that two users share common key. This benefit comes at the cost of higher storage and administrative costs at the sensor nodes. In many communication scenarios, probabilistic results about network connectivity are not sufficient and strong claims about worst-case performance are necessary.

If it is required that any two nodes be able to securely communicate directly, even more structure is needed in the key distribution. The scheme of Çamtepe and Yener based on projective planes is in many senses optimal [7]: all users have the same number of keys, providing uniformity among memory requirements; each key is owned by a constant number of users, giving a constant cost if a key being leaked to an adversary; every pair of users share a single key and so possess the minimal amount of information required to communicate. These features do not come free, however. A key distribution formed by a projective plane supports far fewer users than other key distributions with identical key chain and key pool sizes. For instance, a projective plane of order q has key chain size $q + 1$, key pool size $q^2 + q + 1$ and supports $q^2 + q + 1$ users, while a randomized distribution due to Eschenauer and Gligor supports with key chain size $q + 1$ and key pool size $q^2 + q + 1$ supports $\binom{q^2+q+1}{q+1}$ users.

An attentive reader may have noticed that if the role of points and blocks were interchanged in key distribution based on a 2-design, then the final axiom of a 2-design, namely that any pair of points occurs in exactly λ blocks, would imply that any two users share exactly λ common cryptographic keys. This would be a very desirable characteristic, but it comes at a large cost. Fischer's inequality states that the number of blocks is at least as large as the number of points, and so under the alternate formulation, the number of users supported by the system is at most the number of keys. This is not competitive with many other classes of key distribution schemes which typically support many more users than they have keys.

2.2 Group Communication Systems

A group communication system is in many ways an extension of a WSN. Here, in addition to aggregating and processing the data collected by the sensor nodes under its control, the base station also broadcasts a so-called *group communication stream*. Making the same assumptions about the necessity of data security and the limitations of the sensor nodes, this group communication stream is symmetrically encrypted by the base station using a group-wide *session key* to which every user has access. Each user additionally owns a collection of *administrative keys* which are used, for instance, to distribute a new session key if the need arises. We typically use the term “key chain” in this context to refer to a user’s collection of administrative keys only. Similarly, the term “key pool” in this context is used to refer to the collection of all administrative keys.

Fluid membership makes key distribution in group communications systems a difficult task. For example, suppose that a new member joins the network. It is possible that this new member has been recording the encrypted group communication stream before her addition to the network. Hence, if the session key is not changed after her arrival, the new user can decrypt old communications that she was not at the time privileged to hear; this is known as the *backward secrecy* problem. For another case, suppose a member leaves or must be ejected from the network. Clearly, the session key must be changed, because if not the ejected member can continue decoding the private group communication to which he is no longer privileged; this is known as the *forward secrecy* problem. But member leave is more complicated than member join, because after a user ejection the replacement session key must be securely disseminated without using any administrative key owned by the ejected user; encrypting the new session key using an administrative key owned by the ejected member would be pointless, as the ejected member could simply decrypt the replacement session key and regain access to the group communication. When a user leaves the network, the base station generates a fresh session key together with new administrative keys to replace those owned by the ejected user. The new session and administrative keys are collectively known as the *re-keying message*. The base station then securely distributes the re-keying message to all remaining privileged members. The technical difficulty arises in that this secure distribution must be accomplished without using any key owned by any ejected user(s). The exact manner in which secure re-keying is achieved can vary, and some methods are far more efficient than others.

Harney and Muckenhirn were among the first to consider the problem of group key management [20]. One of the notable limitations of their Group Key Management Protocol (GKMP) is that the system cannot be re-keyed after a user ejection; a new group is formed after every such event. Wong, Gouda and Lam introduced n-ary trees as group key management structures [39]. The leaves of the tree represent keys owned by individual users, all interior nodes represent subgroup keys and the root represents the session key. Each of v users then owns all keys on the path from the associated leaf to the root, and so each user owns $\log_n v$ keys. These authors show that the base station must send $n(\log_n(v) - 1)$ separate encryptions of the re-keying message. This efficient re-keying set comes at the cost of increased number of total keys necessary to maintain a given number of users. To support v users using an n-ary tree, the base station must store

$\sum_{i=0}^{\log_n v} n^i = (n^{1+\log_n v} - 1)/(n-1)$ keys. For instance, in a binary tree supporting v users, the base station must store and administer $2v-1$ keys. Eltoweissy, Heydari, Morales and Sudborough introduced the concept of an (v, k, m) exclusion basis system (EBS) [17]. In this set system the points represent users and the blocks represent keys; note that this is the reverse of set systems we consider here. A (v, k, m) EBS supports v users $\{1, 2, \ldots, v\}$ and can be re-keyed after the ejection of single user e with m sets such that their union is $\{1, 2, \ldots, e-1, e+1, \ldots, v\}$. The authors show that this scheme supports $v \leq \binom{k+m}{k}$ users where k is the key chain size. We will see that this class of structures is a subset of a larger class of combinatorial objects. Their formulation provides for support for a large number of users, but is susceptible to collusion attacks. In the original work that follows, we will present a scheme that is in some sense a compromise between these two systems.

2.2.1 Cover-free Families

For a concrete example of how the re-keying problem dictates the structure of the underlying administrative key distribution, first consider the case in which a single user is ejected from the network. In a default re-keying solution, the base station sequentially encrypts the re-key message with *every* administrative key not owned by the ejected user and broadcasts the encrypted message. For instance, this is the solution put forward by Eltoweissy *et al.* to re-key an exclusion basis system [17]. Any privileged user remaining after the single ejection can gain access to the replacement keys if and only if he has access to at least one administrative key not owned by the ejected user. Hence, the system can be securely re-keyed after the ejection of one user if and only if no key chain is identical to or a proper subset of any other. In symbols, $B \nsubseteq C$ for every pair of distinct sets B and C in $\mathcal{B}$.

In general, we are concerned with the ability of the system to eject more than one user simultaneously. If a group of users is found to be colluding to compromise the network, they must be ejected together; if even one of the colluders receives the replacement session key, he can distribute it to his cohort and so neutralize the re-keying operation. Suppose again that the base station encrypts the re-key message sequentially with each administrative key not owned by any of the ejected members and broadcasts to the group. Then a remaining user can decrypt the replacement keys if and only if she has access to at least one key not owned by *any* of the ejected users. Mathematically speaking, the base station can successfully re-key the network after any r simultaneous ejections if and only if no user's key chain is included in the union of r other key chains. The formulation of this requirement is captured exactly in the definition of an r-cover-free family.

Definition 2.2.1. *An* r-cover-free family *is a set system* $(X, \mathcal{B})$ *in which any distinct blocks* $B_1, B_2, \ldots, B_r$ *and* A *in* $\mathcal{B}$ *satisfy*

$$A \nsubseteq \bigcup_{i=1}^{r} B_i. \tag{2.1}$$

In words, no union of r *blocks covers any other block.*

The case in which $r = 1$ was investigated by Sperner in the mid-1920s [35]. Here, the cover-free family is an anti-chain in the poset under inclusion of the power set of X. Sperner bounded the size of a 1-cover-free family of an ambient v-set X at

$$|\mathcal{B}| \leq \binom{v}{\lceil v/2 \rceil}. \tag{2.2}$$

Notice that this agrees with preliminary intuition that the collection of all subsets of size $\lceil v/2 \rceil$ is maximal under non-inclusion.

The subject of cover-free families lay dormant until the 1960s when Kautz and Singleton later investigated the more general case of $r > 1$ in a coding theoretic context [23]. In data retrieval systems it is often advantageous to classify files according to certain descriptors which are drawn form a system-wide dictionary. A file is then identified by the collection of descriptors it satisfies. In the framework considered by Kautz and Singleton, the descriptors themselves are associated with binary words of length n, and a file is associated with the component-wise OR of the descriptors it satisfies. Under what circumstances is this mapping well-defined? For an error case, consider a file which satisfies a collection of descriptors $\mathcal{D}$ but not some fixed descriptor $D \notin \mathcal{D}$. If D is logically included in the component-wise OR of the descriptors in $\mathcal{D}$, then the retrieval system would incorrectly deduce that the file also satisfies D. Hence, if we require that the retrieval system support up to r descriptors per file without ambiguity, no descriptor may be logically included in the component-wise OR of any r other descriptors. Considering each component in the length n binary word to be an element in a support set X and each descriptor as a set containing the points from X to be the components in which the descriptor's binary word is equal to 1 gives the modern formulation of a r-cover-free family.

It is somehow appropriate that the first modern usage of cover-free families was developed in application to a real world problem. Indeed, cover-free families have been re-discovered several times exactly because their structure is so utilitarian. Desmet *et al.* proposed using cover-free families to protect spread spectrum communications systems from insider adversaries [15]. Cover-free families were put forward by Colbourn, Ling and Syrotiuk for transmission scheduling in mobile ad hoc networks without knowledge of the network topology [12]. Wang and Pieprzyk used 2-cover-free families for anonymous membership broadcast schemes in which a base station can broadcast a message and only the intended recipient can deduce the message's destination [38]. Staddon, Stinson and Wei introduced cover-free families as traceability codes [36]; here illegally distributed content can be traced back to the users who colluding to pirate the content.

Cover-free families have received attention from theoreticians, as well. Erdös, Frankl and Füredi gave one of the first thorough treatments of the subject from a theoretical perspective [18]. They documented construction techniques, bounded the cardinality of a cover-free family with given parameters and investigated the interplay between the size of an r-cover-free family and the cover parameter r. More recently, Stinson and Wei have produced a generalization of the concept cover-freedom [37]: a $(w, r; d)$-cover-free family is a collection of subsets $\mathcal{B}$ of a support set X such that any w blocks $B_1, B_2, \ldots, B_w \in \mathcal{B}$ and r other blocks $A_1, A_2, \ldots, A_r \in \mathcal{B}$ satisfy

$$\left| \left(\bigcap_{i=1}^{w} B_i \right) \setminus \left(\bigcup_{j=1}^{r} A_j \right) \right| \geq d. \tag{2.3}$$

The traditional notion of a r-cover-free family is included here as the class of $(1, r; 1)$-cover-free families. In the applications that follow, the added power provided by this generalization will not be needed, and we will always take the term cover-free family to mean the original, limited definition. Ling, Wang and Xing give an excellent review of both cover-free family theory and applications [27].

As early as the late 1980s, cover-free families were recognized for their ability to prevent collusion [29]. In the context of a WSN, if no user's key chain lies in the union of r other key chains, then no r users can collude to forge the key chain of any other user. More recently, however, Xu, Chen and Wang have proposed using an r-cover-free family as the key distribution in a group communication system supporting up to r simultaneous user ejections [40]. In the re-keying phase, Xu *et al.* stipulate that the base station first determine a collection of keys in which every non-ejected owns at least one of these selected keys, and no ejected member owns any; the authors give no construction for this collection. We will simply call any such collection of keys a *re-keying set* where the exact key distribution and number of ejected users under consideration are clear. Then to re-key the network, the base station must encrypt the re-key message with each key in the re-keying set and broadcast the result to the network. It is obviously desirable to minimize the size of this collection, because the no secure communication can take place until the network has been re-keyed. At the same time, the method for finding a minimal re-keying set should be as computationally inexpensive as possible to minimize load at the base station. Balancing re-keying latency and the computational complexity of determining a suitably small re-keying set is integral to the success of a re-keying solution.

2.2.2 NP-hardness of Minimal Re-keying

Manufacturing a minimal re-keying set is related to Hitting Set, one of Karp's original NP-complete problems [22]. Hitting Set can be phrased as an optimization problem in the following way: given a set X and a collection $\mathcal{B}$ of subsets of X, what is the minimal cardinality of $H \subseteq X$ such that H "hits" every set in $\mathcal{B}$, that is $H \cap B \neq \emptyset$, for every $B \in \mathcal{B}$. We can formalize the relationship between Hitting Set and secure re-keying in the following theorem.

Theorem 2.2.1. *The problem of finding a minimal re-keying set after any number of user leaves in a group communication system with key distribution* $(X', \mathcal{B}')$ *is NP-hard.*

Proof. We will reduce from Hitting Set, meaning we will show that every instance of

Hitting Set corresponds to at least one instance of the re-keying problem; therefore, the re-keying problem is at least as difficult as Hitting Set. Let X be the support set and $\mathcal{B}$ the collection of subsets of X for which we want to find a minimal hitting set. Let E be a set disjoint from X, and suppose that E represents the union of the key chains of the ejected users. Let

$$\begin{aligned} B_i &= B_i' \setminus E \\ \mathcal{B}' &= \{B_i' : B_i \in \mathcal{B}\} \\ X' &= X \cup E \end{aligned}$$

so that $(X', \mathcal{B}')$ is a key distribution of the network before the user ejection(s). This mapping can clearly be performed in polynomial time. Then the Hitting Set problem on the set system $(X, \mathcal{B})$ is equivalent to the minimal re-keying problem on the key distribution $(X', \mathcal{B}')$ in the case where any user(s) owning the key collection E is ejected from the network. Hence, Hitting Set is many-one reducible to the minimal re-keying set problem. □

We note for completeness that Wong, Gouda and Lam independently stated (but did not prove) the NP-hardness of the re-keying problem using a Set Cover reduction [39]. The statement of the theorem makes no assumptions about the exact nature of the key distribution. And in general the results of a NP-hardness theorem should be always taken with a grain of salt. The theorem does not claim that finding a minimal re-keying set is always hopeless endeavor, but rather that finding a minimal re-keying set for an *arbitrary* key distribution should be expected to be difficult. Therefore, the theorem naturally points towards only considering key distributions with a large amount of combinatorial and/or algebraic structure and hoping that this structure is enough to circumvent the natural difficulty of the problem.

2.3 Secure Re-keying Using Symmetric 2-Designs

We have seen in Chapter 1 that t-designs and symmetric 2-designs in particular have nice combinatorial structure. Perhaps this structure is enough to allow for strong results as minimal hitting set cardinalities of these set systems and their residuals. In this way, we could guarantee that symmetric 2-designs would be suitable as key distributions in group communication systems with non-static membership. First we must prove that after a given number of ejections a hitting set exists. As previously discussed, for a group communication system to support secure re-keying after any r simultaneous user leaves it is necessary and sufficient that the key distribution of the network forms an r-cover-free family.

Result 2.3.1 ([18]). *A symmetric 2-(v, k, λ) design $(X, \mathcal{B})$ forms a $\lfloor (k-1)/\lambda \rfloor$-cover-free family.*

Proof. Fix a block $B \in \mathcal{B}$. Because the design is symmetric, any block not equal to B shares exactly λ points with B. Suppose there is a collection of blocks whose cardinality λ intersections with B are disjoint from one another. Then since B has k points, and each block in the selected collection covers exactly λ, the covering collection must have at least $\lceil k/\lambda \rceil$ members, and clearly no fewer will suffice to cover B. Then in particular no collection of $\lceil k/\lambda \rceil - 1 = \lfloor (k-1)/\lambda \rfloor$ blocks will cover B. Since B was chosen arbitrarily, the pair $(X, \mathcal{B})$ forms a $\lfloor (k-1)/\lambda \rfloor$-cover-free family. □

Theorem 1.3.1 and Result 2.3.1 shows that a symmetric 2-(v, k, λ) design is a natural candidate for a key distribution of a group communication system since it supports full mesh connectivity and $O(k)$ simultaneous ejections. Projective planes are both the most plentiful known class of symmetric 2-designs and by Result 2.3.1

support the largest number of simultaneous user leaves relative to key chain size k, and so we will focus our attention here.

2.3.1 Projective Plane Key Distributions

In 1978, Brouwer and Schrijver unknowingly settled the problem of minimal rekeying in a projective plane key distribution after a single user leave. As we discussed in Section 1.4, the residual of a projective plane of order q is an affine plane of order q, and so the collection of keys and key rings found in a projective plane key distribution after a single user leave form the points and blocks of an affine plane. Brouwer and Schrijver bounded the number of points needed to meet every hyperplane in $AG(2, q)$.

Result 2.3.2 ([4]). *Let $(X, \mathcal{B})$ be a projective plane of order q formed by the points and lines of $PG(2, q)$. Then the cardinality of the minimal hitting set of the E-residual of $(X, \mathcal{B})$ is $2q - 1$ for any $E \in \mathcal{B}$.*

Brouwer and Schrijver were investigating blocking and hitting sets of projective planes and their residuals in a purely theoretic context; yet again, pure mathematics research finds its way into an applied context decades later. This result gives a firm foundation on which to base optimality results for a single ejection from a projective plane key distribution. The following theorem allows us to extend optimality results to cases involving multiple simultaneous ejections.

Theorem 2.3.1. *Let h_0 be the cardinality of any minimal hitting set of a set system $(X, \mathcal{B})$. Then for any $\{E_1, \ldots, E_n\} = \mathcal{E} \subseteq \mathcal{B}$ the cardinality of the minimal* *hitting set of the $\mathcal{E}$-residual set system $(X', \mathcal{B}')$ with*

$$X' = X \setminus \left(\bigcup_{i=1}^{n} E_i \right) \tag{2.4}$$

$$\mathcal{B}' = \left\{ C \setminus \left(\bigcup_{i=1}^{n} E_i \right) : C \in \mathcal{B} \setminus \mathcal{E} \right\}. \tag{2.5}$$

as in Definition 1.4.1 is at least $h_0 - n$.

Proof. Let S_0 be any minimal hitting set of $(X, \mathcal{B})$. Consider removing any 2 points from S_0 and tracking how many blocks are no longer hit by the resulting set. Suppose that only one block is uncovered by this process. Stated differently, there is a unique block that contains as its hitting set elements the 2 chosen points; all other blocks contain at least one element from the hitting set that is not one of the selected points. Hence, we could remove either of the selected points from S_0 and create a hitting set with cardinality strictly less than $|S_0|$. This contradicts the minimality of S_0, so removing two points from the hitting set necessarily uncovers at least 2 blocks. So the contrapositive is also true, namely that removing fewer than 2 blocks results in removing fewer than 2 points from any minimal hitting set. Hence, a hitting set of $(X \setminus B, \mathcal{B} \setminus \{B\})$ for some $B \in \mathcal{B}$ must have at least $|S_0| - 1$ points. Repeated application of this case gives the desired result. □

It may seem counterintuitive that removing blocks from a set system might act to increase the minimal hitting set cardinality, but there is an example of such behavior close at hand.

Example 2.3.1. *Let $(X, \mathcal{B})$ be a projective plane of order q. Since any two blocks in a projective plane share exactly one point, any block is a hitting set of the set system. Hence, the size of a minimal hitting set is at most $h_0 \leq q + 1$. (In fact, this is exactly the minimum hitting set cardinality, a fact which we will not prove* *here.) Now consider the residual of $(X, \mathcal{B})$ after the removal of a single block. Result 2.3.2 states that the minimal hitting set of the set system has cardinality $2q-1$. Hence, removing a block from the set system acts to* increase *the minimal hitting set cardinality for any $q > 1$.*

There are several constructions of hitting sets of cardinality $2q-1$ for the residual of a projective plane of order q, and Result 2.3.2 shows that all such hitting sets are optimal; we will present one such construction in the following theorem. But theoretical investigation of these hitting sets stops here; to the author's knowledge, no work has been done on determining the minimal hitting set of the residual of a projective plane after two or more block removals. Theorem 2.3.1 provides a framework for discussing optimality hitting sets in these scenarios. We next show that a minimal hitting set of size $2q-2$ of a projective plane after any two block removals is in fact achievable.

Theorem 2.3.2. *Given a key distribution $(X, \mathcal{B})$ based on a projective plane of order q, a minimal re-keying set after the ejection of a single member contains $2q-1$ points, and a minimal re-keying set after the simultaneous ejection of two members contains $2q-2$ points.*

Proof. Fix $E \in \mathcal{B}$ as the user that is to be ejected. Choose any other user $B \in \mathcal{B} \setminus \{E\}$ and let $x_e = B \cap E$. Now, B shares exactly 1 point with every remaining block, and it is either the case that this point is x_e or it isn't. There are $q-1$ blocks in the first class, all of them necessarily disjoint from each other. We can use as a rekeying set $B \setminus \{x_e\}$ together with one point from each of these $q-1$ blocks. Hence, the minimal rekeying set size is at most $q + (q-1) = 2q-1$. The work of Brouwer and Schrijver featured in Result 2.3.2 shows that this number of points is also necessary.

Now suppose that users $E_1, E_2 \in \mathcal{B}$ leave the system. Since $q \geq 2$, there exists at least $B \in \mathcal{B}$ such that $E_1 \cap B = x_e = B \cap E_2$. Every block in $\mathcal{B}$ shares exactly one element with $\mathcal{B}$ and in only $q-2$ blocks besides E_1 and E_2 is this point x_e, and these $q-3$ blocks are necessarily disjoint outside of x_e. Hence, it suffices to take as a re-keying set $B \setminus \{x_e\}$ and any one point from each of the $q-2$ remaining blocks which contain x_e. Thus, a re-keying set of cardinality $2q-2$ is sufficient to re-key a projective plane key distribution after two simultaneous ejections. Result 2.3.2 together with Theorem 2.3.1 shows that this number of points is also necessary. $\square$

One can imagine making similar arguments as those in the proof of Theorem 2.3.2 for the cases of more than two simultaneous ejections. However, due to the structure of symmetric 2-designs not all ejection cases are isomorphic, that is, the properties of the minimal re-keying set will be dependent on exactly which users were ejected together. For instance, in the case of three simultaneous user leaves from a projective plane key distribution, it may or may not be the case that all three users share a common key. One can easily verify using arguments similar to those above that the size of the minimal re-keying set is dependent on whether this property is satisfied. Rather than enumerate a growing list of cases for increasing number of ejections, we will point out that a re-keying set is guaranteed by Result 2.3.1 for any collection of $k-1$ or fewer simultaneous user leaves. In Section 2.4 we will present an algorithm which produces optimal results for the cases of one and two simultaneous ejections and significantly smaller re-keying sets for larger numbers of user leaves than the current available technologies are able to produce.

Table 2.1: Comparison of binary tree, projective plane and exclusion basis system key distributions parameters for fixed key chain size k

	Binary tree	Projective plane	Exclusion basis system
Users supported	2^k	$k^2 - k + 1$	$\binom{k+m}{k}$
Key pool size	$2^{k+1} - 1$	$k^2 - k + 1$	$k + m$
Key chain size	k	k	k
Re-key messages	$k - 1$	$2k - 3$	m

2.3.2 Biplane Key Distributions

Standing by themselves, it is difficult to tell whether the optimal re-keying sets from Theorem 2.3.2 are “good”. It may be the case that a slight alteration of the structure of the key distribution could give better re-keying performance at little or no cost to the other desirable qualities of the communication system. One natural place to begin a comparison is with the symmetric 2-$(v, k, 2)$ biplanes.

Theorem 2.3.3. *Given a key distribution $(X, \mathcal{B})$ based on a biplane of order $q \geq 2$, a minimal re-key set after the ejection of a single member contains at most q keys, and a minimal re-keying set after the simultaneous ejection of two members contains at most q keys.*

Proof. We will upper bound the rekeying load in this scenario by explicit construction of suitable rekeying sets.

For the case of a single ejection, let $E \in \mathcal{B}$ be the key chain of the ejected user. Choose any $B \in \mathcal{B} \setminus \{E\}$, and define $X_e = B \cap E$. Every pair of blocks shares exactly two common points, and every pair of points occurs in exactly two blocks. Hence, the only blocks containing X_e are E and B themselves, and every other block contains at least one point in $B \setminus X_e$. Thus, the rekeying load for this scenario is at most $k - 2 = q$.

For the case of two simultaneous ejections, let $E_1, E_2 \in \mathcal{B}$ be the key chains of the ejected user. The replication number and the order of a biplane are related by $r = q + 2$. Since $q \geq 2$, we have $r \geq 4$. Hence it is possible to choose $B \in \mathcal{B} \setminus \{E_1, E_2\}$ such that B contains a point x_e found in both E_1 and E_2. The collection $B \cap (E_1 \cup E_2)$ must contain exactly 3 points. If it contained 2, then the two points $E_1 \cap E_2$ would occur in more than two blocks, a contradiction of the

2-design axiom. If it contained 4, then the intersections $B \cap E_1$ and $B \cap E_2$, each having cardinality 2, would be disjoint, a contradiction on our choice of B.

Any block $B' \in \mathcal{B} \setminus \{E_1, E_2\}$ is hit by $B \setminus (E_1 \cup E_2)$ unless both of the points shared between B' and B occur in $B \cap (E_1 \cup E_2) = \{x_e, y, z\}$. Without loss of generality, suppose that the pair $\{x_e, y\}$ occurs in both B and E_1 and that the pair $\{x_e, z\}$ occurs in both B and E_2. The remaining pair $\{y, z\}$ appears in B and one other block U. Hence, this block U is the only block not hit by the set $B \setminus (E_1 \cup E_2)$. It suffices to take one point from U to complete the re-keying set. Hence, there exists a re-keying set of size $|B| - B \cap (E_1 \cup E_2) + 1 = k - 3 + 1 = k - 2 = q$. □

Unfortunately, there is no known biplane equivalent of Result 2.3.2, and so far a proof has alluded the author. We can, however, establish that in some cases the bounds from Theorem 2.3.3 are tight.

An order $q = 2$ biplane is a 2-(7,4,2) design. To improve on the bound in this case, a hitting set of cardinality 1 of a E-residual or $\{E_1, E_1\}$-residual of the biplane must be found. But this is impossible, since ever point occurs in exactly 4 blocks. Hence, one point is not sufficient to hit the 6 blocks remaining after one ejection or the 5 blocks left after two ejections.

Next, take for instance the case of ejections from the biplane of order $q = 3$ featured in Example 1.3.1. To improve on the result above, a re-keying set with fewer than $q = 3$ keys must be constructed. Any key hits $k = 5$ key chains, any single pair of keys occurs in exactly two blocks. Hence, any pair of points hits $2k - 2 = 8$ key chains by inclusion-exclusion. There are $v = 11$ users Clearly no set of $q - 1 = 2$ keys is sufficient for either the one or two ejection case, and so the bounds from Theorem 2.3.3 above are tight when $q = 3$.

Suppose now that the key distribution is based on the biplane of order $q = 4$, so that $k = 6$. A re-keying set improving on Theorem 2.3.3 has at most 3 keys. By inclusion exclusion, any three 3 keys hits at most $3k - 2\binom{3}{2} + \binom{3}{3} = 18 - 6 + 1 = 13$ key chains. There are 16 blocks in the original biplane, and so a set of any 3 keys is insufficient to re-key the system after the ejection of one or two users, and so the bounds Theorem 2.3.3 are tight when $q = 4$.

We can encapsulate these ideas in the following theorem.

Theorem 2.3.4. *Given a key distribution $(X, \mathcal{B})$ based on a biplane of order $2 \leq q \leq 2$, a minimal re-key set after the ejection of a single member contains q keys, and a minimal re-keying set after the simultaneous ejection of two members contains q keys. Hitting sets of these cardinalities are achievable.*

This is where this type of reasoning about minimal hitting sets in residuals of biplanes runs dry. The tactic used by Brouwer and Schrijver to determine cardinality of a minimal hitting set of $AG(d, q)$ relies heavily on the vector space interpretation of the set system. Unfortunately, there is no straight-forward equivalent in the context of biplanes; these set systems are defined by axioms that are not in agreement with those of finite geometry, and to date there has been no successful work in connecting biplanes of arbitrary order outside of their axiomatic definition. We will see later in the form of simulation results that the bound presented in the previous theorems can certainly be improved for larger orders.

2.4 Hitting Set Constructions

While in general finding a minimal hitting set is hard, there do exist approximation algorithms which guarantee that the hitting set produced will be at most a multiplicative factor times the minimal cardinality. Take for instance the following greedy algorithm for finding a hitting set of the residual set system $(X, \mathcal{B})$.

```
Greedy algorithm

  (0) $H \leftarrow \emptyset$; $\mathcal{B}' \leftarrow \mathcal{B}$; $X' \leftarrow X$
  (1) $H \leftarrow H \cup \{x\}$ such that $x \in X'$ and $|\{B \in \mathcal{B}' : x \in B\}|$ is maximal
  (2) $X' \leftarrow X' \setminus \{x\}$
  (3) $\mathcal{B}' \leftarrow \{B \setminus X' : B \in \mathcal{B}'$ and $B \cap H = \emptyset\}$
  (4) if $H$ is a hitting set of $\mathcal{B}$, return $H$; else, go to (1).
```

We note for clarity that if in Step (1) there are multiple valid choices for x, we select one uniformly at random. Chvatal showed that this algorithm produces hitting sets that are at most $\ln(k)$ times the minimal hitting set cardinality [10]. We will see that in practice, the algorithm performs much better than this bound for residuals of symmetric designs.

The default re-keying solution is to take as the re-keying set all keys not owned by any ejected user. One can imagine that such a scheme does not provide competitive results. So instead we compare the greedy algorithm to a randomized construction in which keys chosen uniformly randomly without replacement from the key pool and until the accumulated collection forms a re-keying set.

```
Randomized algorithm

  (0) $H \leftarrow \emptyset$

  (1) $H \leftarrow H \cup \{x\}$ with $x \in X \setminus H$ chosen uniformly randomly

  (2) if $H$ is a hitting set of $\mathcal{B}$, return $H$; else, go to (1).
```

The data presented in Table 2.2 and Figure 2.4 detail the performance of the greedy algorithm when applied to the key distribution based on a projective plane of order 11. For each data point, users where chosen uniformly randomly without replacement and ejected from the network. The greedy and randomized algorithms were then performed on the residual set system and the cardinalities of these re-keying sets were recorded. (Note that since the projective plane of order 11 forms an 11-cover-free according to Result 2.3.1, both algorithms are guaranteed to terminate documented data points.) The procedure was repeated $N = 1000$ times. Error bars in Figure 2.4 have width 4 times the standard error $\sigma/\sqrt{N}$, where σ is the sample standard deviation. Result 2.3.2 and Theorem **??** imply that the minimal cardinalities of a re-keying sets after one ejection and two simultaneous ejections for this key distribution are 21 and 20, respectively, and indeed the greedy algorithm produces such minimal re-keying sets for all N trials. For greater than 2 simultaneous ejections, the greedy algorithm exhibits near constant performance, requiring roughly separate encryptions of the replacement keys in order to secure the network after the simultaneous ejection of 3 to 11 users. Note that the default re-keying scheme is competitive with neither the randomized nor greedy algorithm.

Table 2.3 features sample mean and standard deviation data from $N = 1000$ trials of the greedy and randomized algorithms applied to the key distribution based on a biplane of order 7. (Note that biplane of order 7 forms 4-cover-free family by Result 2.3.1.) Figure 2.4 displays this data; error bars in Figure 2.4 have width 4 times the standard error $\sigma/\sqrt{N}$, where σ is the sample standard deviation. The results are similar to those observed in the projective plane example with a few notable exceptions. For the cases of one ejection and two simultaneous ejections, the greedy algorithm sometimes produces hitting sets of cardinality 6. This performance beats the upper bounds guaranteed by Theorem 2.3.3. We have shown, however, that Theorem 2.3.3 is tight for $q \leq 6$. Together, these facts imply that the minimal hitting sets of biplanes have more interesting structure than those of projective planes. For more than 2 simultaneous ejections, the greedy algorithm has near constant performance. Note that the default re-keying scheme is not competitive with either of the other two algorithms.

Table 2.2: Comparison of the sample mean μ (sample standard deviation σ) cardinalities with $N = 1000$ for greedy, random and default re-keying set construction algorithms for a projective plane of order 11

No. ejections	Greedy	Random	Default
1	21.00 (0.00)	45.61 (7.59)	121.00 (0.00)
2	20.00 (0.00)	44.13 (6.76)	110.00 (0.00)
3	20.18 (0.43)	43.16 (6.36)	99.93 (0.25)
4	20.05 (0.53)	42.27 (6.43)	90.70 (0.50)
5	20.06 (0.62)	40.84 (5.73)	82.25 (0.70)
6	20.07 (0.65)	39.68 (5.40)	74.54 (0.92)
7	20.09 (0.72)	38.61 (4.94)	67.51 (1.11)
8	20.10 (0.76)	37.03 (4.91)	61.01 (1.39)
9	20.20 (0.79)	35.90 (4.35)	55.19 (1.53)
10	20.31 (0.86)	35.13 (4.04)	49.90 (1.65)
11	20.42 (0.92)	33.82 (3.68)	45.03 (1.80)

Table 2.3: Comparison of the sample mean μ (sample standard deviation σ) cardinalities with $N = 1000$ for greedy, random and default re-keying set construction algorithms for a biplane of order 7

No. ejections	Greedy	Random	Default
1	6.97 (0.16)	11.66 (2.14)	28.00 (0.00)
2	6.99 (0.19)	10.66 (1.69)	21.00 (0.00)
3	6.91 (0.40)	9.75 (1.38)	15.57 (0.49)
4	6.85 (0.54)	8.91 (1.07)	11.48 (0.75)

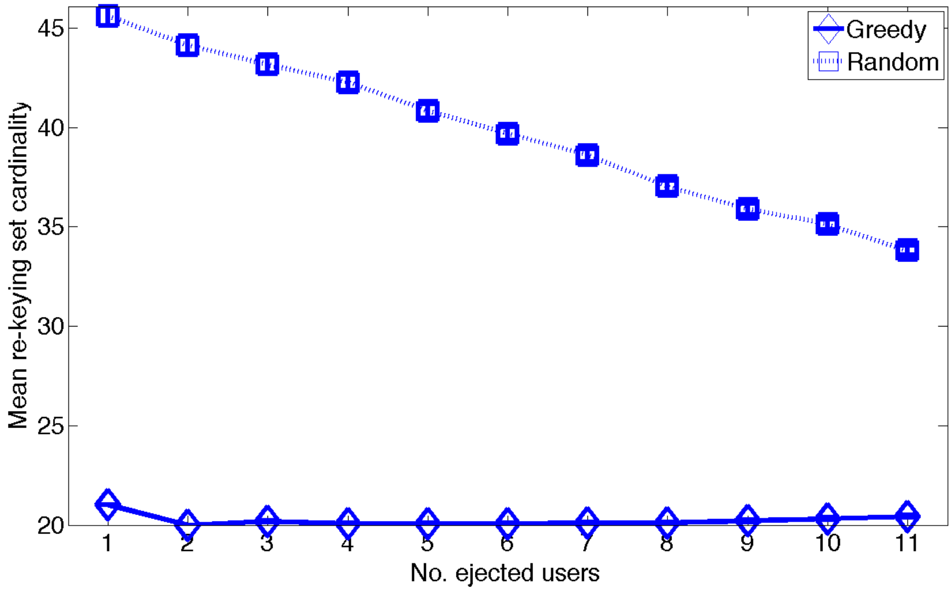

Figure 2.1: Plot of the data featured in Table 2.2 with error bar width representing 4 times the standard error $\sigma/\sqrt{1000}$

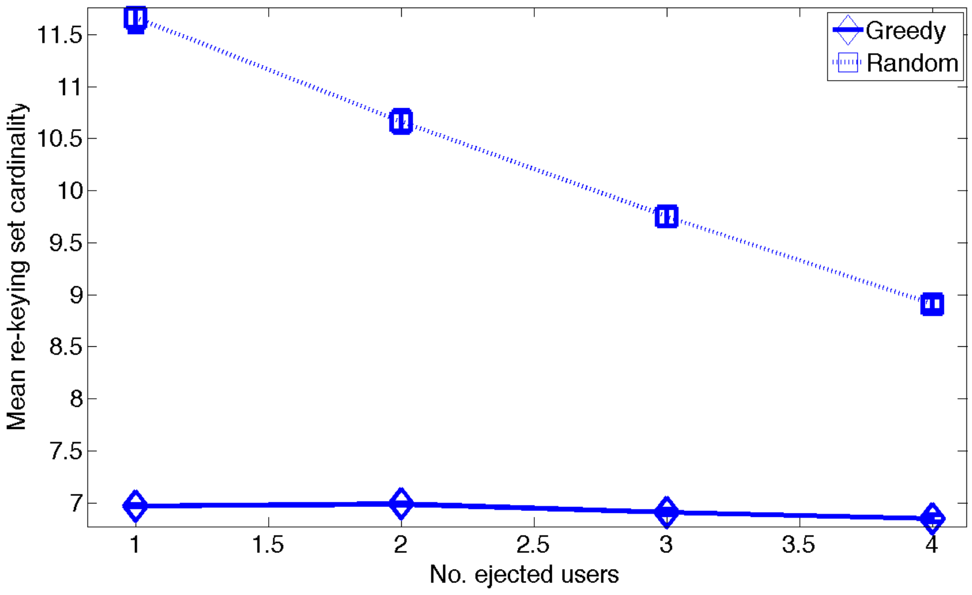

Figure 2.2: Plot of the data featured in Table 2.3 with error bar width representing 4 times the standard error $\sigma/\sqrt{1000}$

CHAPTER 3

ADVANCED METERING INFRASTRUCTURE

3.1 Neighborhood Area Networks

The increasing availability of accurate, real-time electricity consumption data has the potential to substantially increase both the efficiency and the quality of service offered by utility providers. With knowledge of current conditions, power companies and their business partners can reduce waste by bringing generators online only when inferred demand begins to outstrip current supply. Real-time demand statistics also enable service providers to price electricity on a hourly or finer timescale. Customers who are provided with this pricing information can make better-informed choices as to when to perform energy-intensive tasks. These local cost-saving behaviors, known as *economic demand response*, organically flatten global demand. A reduced peak-to-average demand ratio further benefits electricity providers and their customers by eliminating the need to support generators which may only operate in extreme usage scenarios. The ensemble of technologies used to gather and analyze the electricity consumption data necessary to make demand response possible is known as *advanced metering infrastructure* (AMI). The Federal Energy Regulatory Commission (FERC) has estimated that as of 2010, AMI penetration in the United States reached 8.7% nationwide and over 13% in some areas, with over 500 groups offer demand response services [14]. This level of participation represents an 85% increase in a two year period. The advancement of AMI is also receiving significant fiscal support from the federal government. The American Recovery and Reinvestment Act has so far awarded over $790 million dollars for AMI development and deployment [33], and in addition AMI projects have received funding through federal commitments to the modernization of the nation's electrical grid. We note that while the work presented here will focus on the applications of AMI to distribution of electricity, it is becoming more prevalent to use similar advanced metering technologies and techniques in water and natural gas utilities.

In the deployment of a distribution system featuring AMI, homes are first retrofitted with "smart" meters as replacement for traditional mechano-electric meters. These smart meters monitor consumption in the usual way on an hourly or sub-hourly basis and in addition transmit usage data to a neighborhood-wide collection station at least once daily [14]. A collection station oversees the neighborhood area network (NAN) by aggregating consumption statistics and sending a summary to the electricity provider. Transmissions from smart meters to collection stations are typically wireless to facilitate easy installation, while the backhaul from the collection station to the electricity provider is typically a wired connection. The 2010 FERC AMI definition specifies that in return for the consumption data gathered from smart meters, electricity providers must supply customers with current utility price information at least once daily; the definition does not specify the mode in which this information be conveyed, however. On top of this base level of service, one can imagine entrusting further functionality to the AMI, including providing fine grain (*e.g.* sub-daily or more frequent) pricing data, sending pricing data directly to the home via the smart meter, emergency consumption reduction for outage avoidance, quality of service monitoring and remote disconnect capabilities.

The potential for AMI to be a transformative set of technologies is directly tied to its perceived security. Accurate prediction of system-wide demand requires large-scale participation, and universal adoption of a non-secure system is not realistic. Customers are less likely to accept a system which does not closely their guard personal information or does not provide strong assurances that the pricing data and commands received at the home smart meter can be trusted. Privacy issues related to utility consumption data have been receiving attention at both the federal and state levels. The 2010 Guidelines for Smart Grid Cyber Security published by the National Institute of Standards and Technology (NIST) documents the need for the security of consumer data to remain a priority as smart grid technologies advance and are deployed [31]. The California Public Utilities Commission has recently dealt with privacy issues related to the availability of customer utility consumption data both to the customers themselves and to “other interested persons” [13]. More recently, President Obama has ordered a 60 day cyber-security review, including the security of the nation’s electric grid. These concerns are not unfounded. Lisovich, Mulligan and Wicker have shown that even the coarse-grain consumption information provided by AMI can be used to infer details as to what is taking place within a household [28]. Lerner and Mulligan have detailed potential types of abuse stemming from unencrypted AMI data and discussed Fourth Amendment implications to AMI data availability [26]. Without encryption on the provider-to-customer link, a system with AMI is vulnerable to attacks in which an adversary impersonates the utility provider. Depending on the level of control over individual homes given to the provider through a smart meter, such an impersonation could result in consequences ranging from incorrect pricing information to termination of services.

At its core, AMI is a wireless sensor network (WSN) as described in Section 2.1. Advanced metering systems which provide feedback to consumers through smart meters have a level of infrastructure on top of the underlying WSN. In addition to communicating securely with the base station, each consumer must be able to decode the group-wide pricing information and commands being transmitted via the NAN collection station. In other words, this additional functionality forms a group communication system as discussed in Section 2.2.

3.2 Home Area Networks

Next generation utility technology will also focus on intra-home applications. A base station within the home, perhaps taking the form of the smart meter itself, will collect consumption information from individual appliances using non-intrusive load monitor (NILM) technology. In the past, NILM technology has centered on feature-detection at the home-wide level, that is, extracting appliance load signatures from a single home-wide monitoring point [16]. This is a complicated problem as the load monitor must be trained to recognize the power consumption signatures of individual appliances from aggregate data. Suitably sophisticated techniques including neural network training and cluster detection have been put forward as possible solutions; see the work of Zeifman and Roth for a recent overview [41]. Many consumer devices (*e.g.*, OWL, TED, PowerCost) incorporate some or all aspects of this technology.

Another appliance-level monitoring strategy involves placing sensors on individual appliances. Each sensor then transmits the power drawn by its associated appliance to a collection station. In this way, the signatures are already disaggregated. For instance, the Chinese firm Sailwider has developed an appliance-level monitoring system in which a sensor is placed between the appliance and a standard electrical outlet. In another system, General Electric is developing a collection of appliances that will interface with the company's Brillion consumption monitoring platform. The end-to-end control of this system provides the home-wide station the ability to alter the functional behavior of appliances based on current utility prices. This bi-directional communication constitutes a group communication system. Such systems present the same security risks as their AMI equivalents discussed above and perhaps even more so. Since each appliance is transmitting its consumption, an eavesdropper could potentially determine the number, make, model and operational schedule of a home's monitored appliances. Moreover, the collection of appliances in a home changes over time, and the event of membership change in the network must be considered. Intra-home appliance-level monitoring and control technologies are still in their infancy, however, and no common platform has been converged upon. It is unclear whether wireless sensor motes with limited computational and batter life (like the ones considered in this text) will come out the winners.

CHAPTER 4
CONCLUSION

We have presented in this work a scheme for a cryptographic key pre-distribution and secure re-keying based on symmetric 2-designs. These combinatorial objects are natural candidates due to their highly regular structure and well-documented construction algorithms. We provided results showing that a symmetric 2-(v, k, λ) design forms a $\lfloor (k-1)/\lambda \rfloor$-cover-free family. This prevents up $\lfloor (k-1)/\lambda \rfloor$ users from pooling their key chains in order to impersonate another user. Moreover, this same structure allows secure re-keying after up to $\lfloor (k-1)/\lambda \rfloor$ simultaneous user ejections. We showed that in general the problem of finding a collection of keys suitable for re-keying a group communication system after even one user ejection is NP-hard. Fortunately, the combinatorial structure of symmetric 2-designs allowed us to circumvent the natural difficult of this problem. We provided an algorithm for constructing a re-keying set after one or two simultaneous user ejections from a projective plane key distribution, and produced a combination of known results and novel work to show that these re-keying sets are minimal. We provided a similar construction for biplane key distributions; we proved the re-keying sets here are minimal in some cases. For more than two simultaneous ejections in either of these key distributions, we cited a well-known approximation algorithm for Hitting Set. We provided simulation results documenting that this algorithm performs significantly better than existing solutions. We compared these symmetric 2-design schemes to existing key distributions on a number of different metrics.

We presented a sample application of this technology in the form of advanced metering infrastructure (AMI), on both the home and appliance level. We documented existing literature showing that the information being passed from consumer to utility provider through the AMI presents a potential privacy hazard, and we concluded the wireless sensor networks being deployed to monitor end-user utility consumption as part of the AMI should be symmetrically encrypted. One such solution has been presented in the main body of this work.

BIBLIOGRAPHY

[1] I Akyildiz, W Su, Y Sankarasubramaniam, and E Cayirci. Wireless sensor networks: a survey. *Computer Networks*, 38(4):393–422, Jan 2002.

[2] ZigBee Alliance. Zigbee 2007 standard. Jan 2007.

[3] R Bose. An affine analogue of singer's theorem. *Journal of the Indian Mathematical Society*, 6:1–15, 1942.

[4] AE Brouwer and A Schrijver. The blocking number of an affine space. *Journal of Combinatorial Theory, Series A*, 24(2):251–253, Jan 1978.

[5] R Bruck and H Ryser. The nonexistence of certain finite projective planes. *Canadian Journal of Mathematics*, 1(1):88–93, Dec 1949.

[6] P Cameron and J van Lint. Designs, graphs, codes and their links. *London Mathematical Society Student Texts*, Jan 1991.

[7] S Çamtepe and B Yener. Combinatorial design of key distribution mechanisms for wireless sensor networks. *IEEE/ACM Transactions on Networking*, 15(2):346–358, Apr 2007.

[8] D Chakrabarti, S Maitra, and B Roy. A key pre-distribution scheme for wireless sensor networks: merging blocks in combinatorial design. *International Journal of Information Security*, 5(2):105–114, Jan 2006.

[9] S Chowla and H Ryser. Combinatorial problems. *Canadian Journal of Mathematics*, 2(1):93–99, Jan 1950.

[10] V Chvatal. A greedy heuristic for the set-covering problem. *Mathematics of operations research*, 4(3):233 –235, Jan 1979.

[11] C Colbourn and J Dinitz. The crc handbook of combinatorial designs. *Chapman and Hall*, Jan 1996.

[12] C Colbourn, A Ling, and V Syrotiuk. Cover-free families and topology-transparent scheduling for manets. *Designs*, 32(1-3):65–95, Jan 2004.

[13] California Public Utilties Commission. Assigned comissioner and administrative law judge's joint ruling inviting comments on proposed policies and

findings pertaining to the smart grid policies established by the energy information and security act of 2007. 2009.

[14] Federal Energy Regulatory Commission. 2010 assessment of demand response and advanced metering.

[15] Y Desmedt, R Safavi-Naini, H Wang, L Batten, and C Charnes. Broadcast anti-jamming systems. *Computer Networks*, 35(2-3):223–236, Jan 2001.

[16] S Drenker and A Kader. Nonintrusive monitoring of electric loads. *Computer Applications in Power, IEEE*, 12(4):47 – 51, 1999.

[17] M Eltoweissy, M Heydari, L Morales, and I Sudborough. Combinatorial optimization of group key management. *Journal of Network and Systems Management*, 12(1):33–50, Jan 2004.

[18] P Erdös, P Frankl, and Z Füredi. Families of finite sets in which no set is covered by the union of r others. *Israel Journal of Mathematics*, 51:79–89, Jan 1985.

[19] L Eschenauer and V Gligor. A key-management scheme for distributed sensor networks. *Proceedings of the 9th ACM conference on computer and communications security*, pages 41–47, Nov 2002.

[20] H Harney and C Muckenhirn. Rfc2094: Group key management protocol (gkmp) architecture. *RFC Editor United States*, Jan 1997.

[21] James William Peter Hirschfeld, Spyros Simos Magliveras, and Marialuisa J. De Resmini. Geometry, combinatorial designs, and related structures: proceedings of the .... page 258, Jan 1997.

[22] R Karp. Reducibility among combinatorial problems. *50 Years of Integer Programming 1958-2008*, pages 219–241, Jan 2010.

[23] W Kautz and R Singleton. Nonrandom binary superimposed codes. *IEEE Transactions on Information Theory*, 10(4):363 – 377, 1964.

[24] C Lam, L Thiel, and S Swiercz. The non-existence of finite projective planes of order 10. *Canadian Journal of Mathematics*, 41:1117–1123, Jan 1989.

[25] J Lee and D Stinson. A combinatorial approach to key predistribution for

distributed sensor networks. *IEEE Wireless Communications and Networking Conference*, 2:1200–1205, Jan 2005.

[26] J Lerner and D Mulligan. Taking the 'long view' on the fourth amendment: stored records and the sanctity of the home. *Stanford Technology Law Review (STLR)*, 3, Jan 2008.

[27] S Ling, H Wang, and C Xing. Cover-free families and their applications. *Computer and Network Security*, 1:75–98, Jan 2007.

[28] M Lisovich, D Mulligan, and S Wicker. Inferring personal information from demand-response systems. *Security & Privacy, IEEE*, 8(1):11 – 20, 2010.

[29] C Mitchell and F Piper. Key storage in secure networks. *Discrete Applied Mathematics*, 21(3), Oct 1988.

[30] S Nguyen and C Rong. Zigbee security using identity-based cryptography. *Proceeedings of the 4th International Conference on Autonomic and Trusted Computing*, pages 3–12, Jan 2007.

[31] NIST. Introduction to nistir 7628 guidelines for smart grid cyber security. *http://csrc.nist.gov/publications/nistir/ir7628/introduction-to-nistir-7628.pdf*.

[32] NIST. Federal information processing standards publication 197. *http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf*, 2001.

[33] Department of Energy. Advanced metering infrastructure. *http://www.smartgrid.gov/smartgrid_projects?category=4*.

[34] L Ruiz-Garcia, L Lunadei, P Barreiro, and I Robla. A review of wireless sensor technologies and applications in agriculture and food industry: state of the art and current trends. *Sensors*, 9(6):4728–4750, 2009.

[35] E Sperner. Ein satz über untermengen einer endlichen menge. *Mathematische Zeitschrift*, Dec 1928.

[36] J Staddon, D Stinson, and R Wei. Combinatorial properties of frameproof and traceability codes. *IEEE Transactions on Information Theory*, 47(3):1042–1049, 2001.

[37] D Stinson and R Wei. Generalized cover-free families. *Discrete Mathematics*, 279:463–477, Jan 2004.

[38] H Wang and J Pieprzyk. A combinatorial approach to anonymous membership broadcast. *Computing and Combinatorics: Lecture Notes in Computer Science*, 2387:387–413, 2002.

[39] C Wong, M Gouda, and S Lam. Secure group communications using key graphs. *IEEE/ACM Transactions on Networking*, 8(1), Feb 2000.

[40] L Xu, J Chen, and X Wang. Cover-free family based efficient group key management strategy in wireless sensor network. *Journal of Communications*, Jan 2008.

[41] M Zeifman and K Roth. Nonintrusive appliance load monitoring: Review and outlook. *IEEE Transactions on Consumer Electronics*, 57(1):76–84, Jan 2011.