Information Sciences 179 (2009) 3673–3684

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Optimal subset-difference broadcast encryption with free riders ☆

Murat Ak *, Kamer Kaya, Ali Aydın Selçuk

Department of Computer Engineering, Bilkent University, Ankara 06800, Turkey

ARTICLE INFO

Article history:
Received 16 October 2007
Received in revised form 19 February 2009
Accepted 23 May 2009

Keywords:
Broadcast encryption
Digital rights management
Group key management
Subset-difference scheme
Free riders

ABSTRACT

Broadcast encryption (BE) deals with secure transmission of a message to a group of receivers such that only an authorized subset of receivers can decrypt the message. The transmission cost of a BE system can be reduced considerably if a limited number of free riders can be tolerated in the system. In this paper, we study the problem of how to optimally place a given number of free riders in a subset-difference (SD)-based BE system, which is currently the most efficient BE scheme in use and has also been incorporated in standards, and we propose a polynomial-time optimal placement algorithm and three more efficient heuristics for this problem. Simulation experiments show that SD-based BE schemes can benefit significantly from the proposed algorithms.

1. Introduction

Today's secure multimedia applications such as pay-TV, content protection, secure audio streaming and Internet multicasting usually require a broadcast encryption (BE) scheme which enables data transmission to a large set of receivers such that only an authorized subset can decrypt it. This is typically achieved by pre-establishing a set of long-term keys at each receiver device, which is later used to support or revoke selected sets. The particular design of a BE system varies according to the system characteristics, such as size of the user domain, required security level, available bandwidth, and hardware capabilities. In the traditional setting, the amount of long-term storage is very limited as it has to be tamper resistant, the communication channel is one way, and the devices are stateless in the sense that no additional long-term storage is possible.

Although recent advances in the technology, such as the availability of two-way communication channels, have reduced the pay-per-view TV systems' reliance on BE schemes, new application areas have emerged that greatly benefit from BE, such as content protection [18,23], multicasting promotional material and low cost pay-per-view events [2,16], multi-certificate revocation/validation [3] and dynamic group key management [24,25,5,8,19].

Two important performance parameters in evaluating a BE system are the key storage and transmission overheads incurred. Some of the most efficient BE schemes today are the subset-difference (SD) scheme of Naor et al. [20] and its variants [12,13]. The SD scheme has become popular in applications recently and is already implemented in the next-generation DVD standard [1].

In the traditional BE model, it is assumed that all unauthorized receivers must be excluded in a broadcast. Abdalla et al. [2] observed that this model is unnecessarily strict for most practical applications and the cost of a BE system can be reduced significantly when some free riders can be tolerated.

☆ This work is supported in part by the Turkish Scientific and Technological Research Agency (TÜBİTAK), under Grant number 108E150.
* Corresponding author. Present address: Bilkent University, Computer Engineering Department, Bilkent, Ankara, Turkey. Tel.: +90 (312) 290 1350; fax: +90 (312) 266 4047.
E-mail addresses: muratak@cs.bilkent.edu.tr (M. Ak), kamer@cs.bilkent.edu.tr (K. Kaya), selcuk@cs.bilkent.edu.tr (A.A. Selçuk).

0020-0255/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2009.05.025

1.1. Related work

After Berkovits [4] introduced the idea of BE in 1991, Fiat and Naor [11] presented their model, which is the first formal work in the area. They introduced the *resiliency* concept, and defined *k-resilience* to mean being resilient against a coalition of up to k revoked users. Their best scheme required every receiver to store $O(k \log k \log n)$ keys and the center to broadcast $O(k^2 \log^2 k \log n)$ messages where n is the total number of users.

Wallner et al. [24] and Wong et al. [25] independently proposed the logical key hierarchy (LKH) for secure Internet multicast. LKH was not a broadcast encryption scheme, but its key distribution idea was very useful for broadcast encryption. The idea was to relate the receivers with the leaves of a tree, associate a unique key with each node of the tree, and give each receiver the keys of the nodes on the path from the corresponding leaf to the root. With this approach, key storage complexity became logarithmic in terms of the number of receivers, $O(\log n)$.

In [20], which is another milestone in broadcast encryption research, Naor, Naor and Lotspiech proposed two schemes, the complete subtree (CS) and subset-difference (SD). The CS scheme was mainly an adaptation of the LKH ideas to BE and has a transmission cost of $O(r \log(n/r))$, r denoting the number of revoked users. The SD scheme decreased the transmission overhead to $O(r)$ at the expense of increasing the key storage to $O(\log^2 n)$. The SD scheme was the most efficient scheme at the time of its proposal, and most of the recent schemes proposed since then are still based on the SD scheme.

The first significant variant of SD was the layered subset-difference (LSD) scheme, which was proposed by Halevy and Shamir [13]. Optimized LSD has a transmission overhead of $O(\log n \log\log n)$ and a key storage of $O(r \log\log n)$. Goodrich et al. [12] introduced the stratified subset-difference (SSD) scheme, which has $O(r \log n / \log\log n)$ transmission overhead and $O(\log n)$ key storage complexity. An analysis of [11,13,20] can be found in [14].

In the last few years, a number of different approaches have been introduced in BE research. A work on public key cryptographic solutions by Boneh et al. [6] uses bilinear maps and the bilinear decision Diffie–Hellman exponent problem. They achieve constant size ciphertext and a trade-off between ciphertext and public key sizes, whose product is linear in number of receivers. Another recent work by Boneh and Hamburg provides a framework for identity-based broadcast encryption schemes [7]. Recently, there has been an increasing amount of interest in the public key BE framework and it has been the subject of several new studies [9,10,15,17,21].

The idea of allowing some free riders in the system in order to get better performance was introduced by Abdalla et al. [2]. This work was also the first to adapt the key distribution idea of the LKH scheme to broadcast encryption. They investigated the efficient usage of free riders in depth and developed the basic intuitions about the effective assignment of free riders. To minimize the transmission overhead, Ramzan and Woodruff [22] recently proposed an algorithm to optimally choose the set of free riders to be allowed in the CS scheme. Their algorithm was based on a dynamic programming approach that decides the free rider assignment in a tree recursively in a bottom-up fashion.

1.2. Contributions

In this paper, we study how the transmission cost of an SD scheme can be minimized by the effective placement of a limited number of free riders. The contribution is twofold: First, we give a polynomial-time algorithm which computes the optimal placement for a given number of free riders in an SD scheme. We then propose three heuristic methods which work in a greedy fashion. Experimental results show that significant cost reductions are possible in the SD scheme by these algorithms. They also show that the heuristic methods yield nearly optimal solutions most of the time, with a running time dramatically better than that of the optimal algorithm.

1.3. Organization

After describing the SD scheme in Section 2, we formalize the problem in Section 3. Section 4 gives the optimal algorithm and Section 5 describes the proposed heuristics for the problem. After presenting the experimental results in Section 6, we conclude the paper in Section 7.

2. Subset-difference scheme

The SD scheme [20], like many other BE schemes, organizes the set of users in the system as leaves of a binary tree. The basic notations regarding this tree are summarized in Table 1. The nodes in the tree are organized into subsets, and an encryption key is assigned to each subset. A user is given the keys of the subsets of which he is a member. The SD scheme is distinguished by the way it defines these subsets: For every non-leaf node x, and every descendant y of x, a subset is defined as

$$S_{x,y} = \{v | v \in T(x) \text{ and } v \notin T(y)\}.$$

The collection of the $S_{x,y}$ subsets is denoted by $\mathscr{S}$. An example subset-difference and an example cover are illustrated in Fig. 1.

In the broadcast phase of the scheme, to send an encrypted message to a set of privileged users P, the center finds a collection $\mathscr{C} \subseteq \mathscr{S}$ that exactly covers P,

Table 1
Notations regarding the SD tree.

$L(x)$	Immediate left child of x
$R(x)$	Immediate right child of x
$d(x)$	Depth of x; the distance between x and the root
$T(x)$	Subtree rooted at node x
$r(x)$	Number of revoked users in $T(x)$
$p(x)$	Number of privileged users in $T(x)$

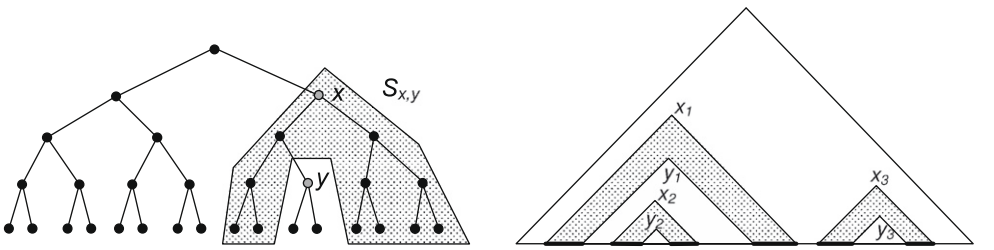

Fig. 1. Example of subset difference and cover.

$$P = \bigcup_{S_{x,y} \in \mathscr{C}} S_{x,y}.$$

A message encryption key k is used to encrypt the transmitted packet. For each subset $S_{x,y} \in \mathscr{C}$, a separate copy of k is encrypted under that subset's key and transmitted along with the message in the header. The transmission cost of the broadcast is defined as the number of these encryptions, i.e., the cardinality of the cover $|\mathscr{C}|$.

3. Problem statement

As observed by Abdalla et al. [2], in many cases it may be preferable to allow a limited number of free riders in a BE system in order to reduce the transmission cost. Given the number of free riders that can be tolerated, the question becomes how to utilize this quota most efficiently.

In our treatment, U denotes the set of all receivers, and P and $R = U - P$ denote the set of privileged and revoked receivers, respectively, where $n = |U|$, $p = |P|$, $r = |R|$. We denote the tree of all users in the system by $\mathscr{T}$. The free rider quota allowed is denoted by f, and c_f denotes the *free rider ratio* f/p. The problem is to find a cover $\mathscr{C} \subseteq \mathscr{S}$, $P \subseteq \bigcup_{S_{x,y} \in \mathscr{C}} S_{x,y}$ with $|\bigcup_{S_{x,y} \in \mathscr{C}} S_{x,y} - P| \leqslant f$, such that $|\mathscr{C}|$ is minimum.

Definition 1 (*i-point, e-point*). We call a node x an *inclusion point (i-point)* and y an *exclusion point (e-point)* in an SD configuration where $S_{x,y}$ is in the cover $\mathscr{C}$.

Definition 2 (*meeting point*). A node x is called a *meeting point* if both $T(L(x))$ and $T(R(x))$ contain revoked leaves, or if x itself is a revoked leaf.

A "meeting point" is a point where a branch occurs in the Steiner tree induced by the revoked users in $\mathscr{T}$, which is the minimum subtree in $\mathscr{T}$ that covers all revoked leaves. As in other works [20,13,22], this Steiner tree is of particular interest for the optimization algorithms we will discuss. We will denote the highest meeting points in the left and right subtrees of a node x in this tree, i.e., the "meeting point children" of x, by $L_{mp}(x)$ and $R_{mp}(x)$, respectively.

By definition, there are r meeting points that are leaves. Since every other meeting point is a common ancestor of two other meeting points, there are $r - 1$ internal meeting points. Thus, there are $2r - 1$ meeting points in total. Also note that the highest meeting point does not have to be the root of the whole binary tree. If one of the root's children does not have any revoked users under it, then the root will not be a meeting point.

4. Optimal algorithm

In this section, we describe a dynamic programming solution for the SD optimization problem with free riders. The approach is based on the dynamic programming approach of Ramzan and Woodruff [22] for the CS scheme. However, a

completely different formulation is needed here due to the complicated relationship between the recursive subproblems in the SD scheme. For the same reason, the approximation algorithm of [22] is also not applicable.

Let x be a meeting point and let $(x;f_x)$ denote the problem instance where exactly f_x free riders are to be placed in $T(x)$. Let $Cost(x,f_x)$ denote the cost of the optimal solution to this problem. Let the left and right meeting point children of x be $y = L_{mp}(x)$ and $z = R_{mp}(x)$. Consider the case where f_y of the free riders are to be assigned under y and $f_z = f_x - f_y$ of them are to be assigned under z. Then, as proven in Section 4.1, the optimal cost for this partition can be expressed in terms of the optimal solutions of $(y;f_y)$ and $(z;f_z)$ as

$$Cost(y,f_y) + Cost(z,f_z) + C_l + C_r, \tag{1}$$

where C_l denotes the additional cost of covering the path between x and y (by the addition of either $S_{x,y}$ or $S_{L(x),y}$, as we explain in detail below) and C_r denotes its counterpart between x and z. Accordingly, the cost of the optimal solution to the problem $(x;f_x)$ can be expressed as

$$Cost(x,f_x) = \min_{\substack{f_y,f_z \geqslant 0 \\ f_y+f_z=f_x}} \{Cost(y,f_y) + Cost(z,f_z) + C_l + C_r\}. \tag{2}$$

Now consider C_l, the cost of the subset that will be added between x and y. First of all, if $f_y = r(y)$, the subtree $T(y)$ and consequently, the whole left subtree of x will be privileged, and no subsets will be needed on the left side of x.

Given that $T(y)$ is not fully privileged, $S_{x,y}$ will be added to the cover if and only if $f_z = r(z)$; i.e., the right subtree of x is fully privileged.

Given that $T(z)$ is not fully privileged either (i.e., $f_z < r(z)$), the only possible addition on the left side of x is $S_{L(x),y}$, which will take place if and only if $L(x) \neq y$ (i.e., y is not the immediate left child of x).

The addition of $S_{x,y}$ or $S_{L(x),y}$ to the cover may or may not bring an additional cost. If y is an i-point in the optimal solution to $(y;f_y)$, the new set will be merged with the existing set under y, and again we will have $C_l = 0$.

Hence the value of C_l is determined as

$$C_l = \begin{cases} 1, & \text{if } f_y < r(y) \text{ and } (f_z = r(z) \text{ or } d(y) - d(x) \geqslant 2) \text{ and } y \text{ is not an } i\text{-point.} \\ 0, & \text{otherwise.} \end{cases}$$

The value of C_r is determined similarly.

If there are more than one solutions that give the minimum cost at (2), the solution that makes x an i-point is selected for the possibility of a later merger.

4.1. Optimal substructure property

Theorem 4.1 below states the optimal substructure property of the SD optimization with free riders problem.

Theorem 4.1. *Let x be a meeting point in an SD tree $\mathscr{T}$, and $y = L_{mp}(x)$ and $z = R_{mp}(x)$. Consider the problem of placing f_x free riders under x optimally, where f_y of them are to be placed under y. An optimal solution to this problem exists that is based on the optimal solutions of $(y;f_y)$ and $(z;f_z)$, where $f_z = f_x - f_y$.*

Proof. Assume to the contrary that the optimal solution to the problem at x gives a suboptimal configuration at either y or z (w.l.o.g., assume it is suboptimal at y); and assume no equivalent solution exists that is based on some optimal solutions at y and z. Let $cost'_y$ denote the cost of the suboptimal configuration at $T(y)$ induced by the optimal solution at x. Similarly, let $cost'_z$, C'_l, and C'_r denote the costs it induces at subtree $T(z)$, and on the paths x–y and x–z respectively. Let $cost_y$ and $cost_z$ be the cost of the optimal solutions of $(y;f_y)$ and $(z;f_z)$, and C_l and C_r denote the associated costs on the paths x–y and x–z in the solution to $(x;f_x)$ based on these optimal solutions at y and z. Hence, we have

$$cost'_y + cost'_z + C'_l + C'_r < cost_y + cost_z + C_l + C_r, \tag{3}$$

$$cost'_y > cost_y, \tag{4}$$

$$cost'_z \geqslant cost_z. \tag{5}$$

Given that C_l and C_r are either 0 or 1, the situation above is possible only with $C_l = C_r = 1$ and $C'_l = C'_r = 0$. The case $C_l = C_r = 1$ is possible only when (i) $T(y)$ and $T(z)$ are not fully privileged; (ii) y and z are not an immediate child of x; and (iii) y and z are not i-points in the optimal solutions of $(y;f_y)$ and $(z;f_z)$. Under conditions (i) and (ii), the assumption that $C'_r = 0$ is possible only when z is an i-point in the corresponding solution in $T(z)$. Given that z was not an i-point in the optimal solution of $(z;f_z)$, this implies $cost'_z > cost_z$. Therefore,

$$cost'_y + cost'_z + C'_l + C'_r \geqslant cost_y + cost_z + C_l + C_r. \quad \square$$

4.2. Algorithm OptimalAssign

Algorithm 1. OptimalAssign $(\mathscr{T}, P, f)$

```
1: MP ← FIND MEETINGPOINTS(R)
2: for i = 1 to r do
3:    x ← MP[i]
4:    C_x[0], C_x[1], I_x[0], I_x[1] ← 0
5: for i = r + 1 to 2r − 1 do
6:    x ← MP[i]; y ← L_mp(x); z ← R_mp(x)
7:    for f_x = 0 to min(r(x), f) for
8:       C_x[f_x] ← ∞
9:       for f_y = max(f_x − r(z), 0) to min(r(y), f_x) do
10:         f_z ← f_x − f_y
11:         tcost ← C_y[f_y] + C_z[f_z] + C_l + C_r
12:         if tcost < C_x[f_x] or (tcost = C_x[f_x] and (r(y) = f_y or r(z) = f_z))
            then
13:            C_x[f_x] ← tcost
14:            L_x[f_x] ← f_y
15:            I_x[f_x] ← 0
16:            if f_y = r(y) or f_z = r(z) then
17:               I_x[f_x] ← 1
18: root_MP ← MP[2r − 1]
19: (result, f_act) ← FindCost(root_MP)
20: 𝒞 ← FindCover(root_MP, f_act)
```

Algorithm 2. findcost $(root_{MP})$

```
1: result ← ∞;
2: for f_root_MP ← 0 to f do
3:    rcost ← C_root_MP[f_root_MP]
4:    if d(root_MP) ≠ 0 then
5:       if I_root_MP[f_root_MP] ≠ 1 then
6:          rcost ← rcost + 1
7:    if result > rcost then
8:       result ← rcost
9:       f_act ← f_root_MP
10: return (result, f_act)
```

Algorithm 1 shows the optimal algorithm based on the dynamic programming formulation given in (2). The *MP* array, which is initialized on line 1, contains a list of the meeting points in $\mathscr{T}$. This array is generated by the FindMeetingPoints procedure such that a meeting point is always listed before its parent. Hence, as the array is processed in order, the program proceeds from the leaves towards the root. In the course of the algorithm, a two-dimensional cost array $C_x[f_x]$ is filled in a bottom-up fashion where a cell $[x, f_x]$ stores the cost of the optimal solution for the subtree of x when f_x free riders are used.

In addition to the cost array, the arrays I_x and L_x are used to maintain the critical information regarding the optimal solution obtained for each problem instance $(x; f_x)$. In the algorithm, $I_x[f_x]$ holds whether x is an i-point in that optimal solution and $L_x[f_x]$ holds how many of the f_x free riders in that optimal solution are assigned to the left subtree of x.

In Algorithm 1, two more procedures are used: The first one, findcost, called on line 19, is described in Algorithm 2. It traverses the cost array filled in the dynamic programming part and finds the optimal cost. The second procedure, findcover, uses I_x and L_x arrays to find the $S_{x,y}$s used in the optimal solution. As described above, the array $I_x[f_x]$ holds whether x is an i-point (i.e., $S_{x,y} \in \mathscr{C}$ for some $y \in T$), and $L_x[f_x]$ holds how many of the f_x free riders are assigned to the left subtree of x in the optimal solution. Note that for an i-point x, the corresponding e-point y is the first descendant of x such that y has more revoked nodes in its subtree than free riders, and also, if y is not a leaf node itself, both children of y have more revoked nodes in their subtrees than free riders. Hence, FindCover can construct $\mathscr{C}$ with a breadth-first search in $O(r)$ time.

The main body of the algorithm OptimalAssign consists of the three nested loops between lines 5 and 17. The first for loop, on line 5, iterates $r-1$ times; the second loop, on line 7, iterates $\min(r(x), f)$ times; and the last one, on line 9, iterates $O(\min(r(y), f))$ times. Hence, a straightforward analysis gives the time complexity of the algorithm as $O(rf^2)$. However, as the following theorem proves, a tighter bound can be found as $O(rf + r \log\log n)$. The proof is along the same lines as that of the dynamic programming algorithm given for the CS scheme in [22].

Theorem 4.2. *The time complexity of the algorithm* OPTIMALASSIGN *is* $O(rf + r\log\log n)$.

Proof. Let *iMP* denote the set of internal meeting points in $\mathscr{T}$. For a meeting point $x \in iMP$, we will use y and z to denote $L_{mp}(x)$ and $R_{mp}(x)$ such that $r(y) \leqslant r(z)$. Then, the total complexity of the three nested loops on lines 5–17 is bounded by

$$t = \sum_{x \in iMP} \min(r(x), f) \cdot \min(r(y), f). \tag{6}$$

The terms that contribute to this summation will be analyzed in three classes:

(1) $x \in iMP$ such that $r(y), r(z) < f$.
(2) $x \in iMP$ such that $r(y) \leqslant f < r(z)$.
(3) $x \in iMP$ such that $f \leqslant r(y), r(z)$.

We will denote these classes by MP_1, MP_2, MP_3, and their contributions to summation (6) by t_1, t_2, t_3, respectively. First consider MP_1 and t_1:

$$t_1 = \sum_{x \in MP_1} r(x)r(y) = \sum_{x \in MP_1} r(y)r(y) + \sum_{x \in MP_1} r(z)r(y) \tag{7}$$

Let t'_1 and t''_1 denote the first and the second halves of summation (7). Since, by definition, $r(y) \leqslant r(z)$, we have $t'_1 \leqslant t''_1$, and therefore, $t_1 \leqslant 2t'_1$.

To compute a bound on t''_1, we will define a formal variable X_u for each revoked user u and set all of these formal variables to 1. By using these variables, we can write $r(y) = \sum_{u \in R \cap T(y)} X_u$ and $r(z) = \sum_{u \in R \cap T(z)} X_u$; hence,

$$r(z)r(y) = \sum_{\substack{u \in R \cap T(y) \\ v \in R \cap T(z)}} X_u X_v,$$

where every X_i equals 1.

Now consider the question of how many monomials $X_u X_v$ a particular revoked user u contributes to the summation t''. Let $\mathscr{T}'$ denote the Steiner tree consisting of the meeting points in $\mathscr{T}$, where a meeting point x and its meeting point children $L_{mp}(x)$ and $R_{mp}(x)$ are linked directly. Let x be the highest ancestor of u in $\mathscr{T}'$ that is in MP_1. Consider the path $u = u_0, u_1, \ldots, u_k = x$ in $\mathscr{T}'$. Let v_i be the sibling of u_i for $0 \leqslant i < k$. Since $T(v_i)$ and $T(v_j)$ are disjoint for all $i \neq j$, there are $\sum_{i=0}^{k-1} |r(v_i)|$ monomials containing X_u and each of them has coefficient 1. So the number of monomials containing X_u can be no more that $2f$ since $x \in MP_1$ and $T(x)$ contains at most $2f$ revoked users. Given that there are r revoked users in total, we have $t'_1 = O(rf)$, and consequently, $t_1 = O(rf)$.

Second, consider MP_2 and t_2:

$$t_2 = \sum_{x \in MP_2} \min(r(x), f) \cdot \min(r(y), f) = \sum_{x \in MP_2} f r(y)$$

Note that any $x \in MP_2$ cannot be a descendant of any other $x' \in MP_2$; hence the $T(y)$, $T(y')$ subtrees are disjoint for any distinct $x, x' \in MP_2$. Therefore, we have

$$t_2 = f \sum_{x \in MP_2} r(y) \leqslant rf.$$

Third and last, consider MP_3 and t_3. Consider the subtree $\mathscr{T}'' \subset \mathscr{T}'$ consisting only of the meeting points in MP_3 and their left and right children. Since there are r revoked users in total, there can be at most r/f leaves in $\mathscr{T}''$. So, the number of the meeting points in MP_3 is no more than $r/f - 1$. Note that the contribution of a meeting point in MP_3 to t_3 is f^2; hence $t_3 = f^2 O(r/f) = O(rf)$.

Since each of t_1, t_2, and t_3 is $O(rf)$, we have $t = O(rf)$. Besides, finding the meeting points at the beginning of the algorithm takes $O(r\log\log n)$ time [22]. Hence, the overall time complexity of the algorithm OPTIMALASSIGN is $O(rf + r\log\log n)$. □

5. Greedy heuristics

When a faster solution is needed, a heuristic algorithm that gives nearly optimal solutions in a shorter time can be preferred. In this section we describe three heuristic methods for this purpose, two greedy algorithms and a third combined method, which return near-optimal results with a running time significantly faster than that of the optimal algorithm.

5.1. Top-down heuristic

The first heuristic searches the user tree in a top-down fashion to identify the $S_{x,y}$ subsets to cover a given receiver set P, such that each subset taken satisfies in itself the free rider ratio $c_f = f/p$.

Note that an SD tree cannot be searched greedily by just looking at single nodes because the $S_{x,y}$ subsets are defined by two nodes having a descendant–ascendant relationship. We define an exclusion point $e(x)$ for every node x to be the descendant of x with the largest subtree under it that is completely revoked. The TopDownAssign heuristic first calls the FindEPoints procedure, which identifies $e(x)$ for a node x recursively, beginning from the root of the Steiner tree, i.e., the highest meeting point. Then TopDownCover is called, which searches the tree from top to bottom for subsets that satisfy the free rider ratio c_f.

TopDownCover(x) takes $S_{x,e(x)}$ into the cover if it satisfies the free rider ratio. Otherwise, if x is a meeting point, the procedure is called recursively on $L(x)$ and $R(x)$. If x is not a meeting point, then a subset that covers all privileged descendants of x until the first meeting point is added to the cover, and the procedure is repeated, beginning from that meeting point. One can also see that we indeed do not need the e-points between an immediate child of a meeting point and its first meeting point descendant. Hence, FindEPoints only finds the e-points of the meeting points and those of their immediate children.

Algorithm 3. TopDownAssign($\mathscr{T}, P, f$)

```
1: MP ← FindMeetingPoints(R)
2: root_MP ← MP[2r − 1]
3: c_f ← f/p
4: if root = root_MP then
5:    𝒞 ← ∅
6: else
7:    𝒞 ← {S_{root,root_MP}}
8: FindEPoints(root_MP)
9: TopDownCover(root_MP)
```

Algorithm 4. FindEPoints(x)

```
1: if r(x) > 0 then
2:    if p(x) = 0 then
3:       e(x) ← x
4:    else
5:       y ← e(L(x)) ← FindEPoints(L_mp(x))
6:       z ← e(R(x)) ← findepoints(R_mp(x))
7:       if r(y) > r(z) then
8:          e(x) ← y
9:       else
10:         e(x) ← z
11:   return e(x)
12: else
13:   return null
```

Algorithm 5. TopDownCover(x)

```
1: if (r(x) − r(e(x)))/(p(x) − p(e(x))) ⩽ c_f then
2:    𝒞 ← 𝒞 ∪ {S_{x,e(x)}}
3: else
4:    if r(L(x)) > 0 and r(R(x)) > 0 then
5:       TopDownCover(L(x))
6:       TopDownCover(R(x))
7:    else
8:       if r(R(x)) = 0 then
9:          𝒞 ← 𝒞 ∪ {S_{x,L_mp(x)}}
10:         TopDownCover(L_mp(x))
11:      if r(L(x)) = 0 then
12:         𝒞 ← 𝒞 ∪ {S_{x,R_mp(x)}}
13:         TopDownCover(R_mp(x))
```

The TopDownAssign heuristic has two main subroutines: FindEPoints and TopDownCover. Both subroutines are recursive methods called once for each meeting point, and do a constant amount of work at each call, hence have a complexity of $O(r)$. The complexity of the algorithm also includes the cost of finding meeting points, which is $O(r \log\log n)$. Hence, the overall time complexity of TopDownAssign is $O(r \log\log n)$.

5.2. Bottom-Up Heuristic

The free rider quota can be utilized more efficiently by a targeted free rider placement heuristic that places the free riders on an existing solution to merge the subsets in the cover $\mathscr{C}$ as efficiently as possible: One can remove an existing $S_{x,y}$ subset from $\mathscr{C}$ by saturating $T(y)$ with free riders. Then $T(x)$ will become fully privileged and has to be covered. Consequently, the subset $S_{parent(x),sibling(x)}$ will be temporarily added to the cover, and it will be determined whether it can be merged with any other subsets or not. Note that if $parent(x)$ is an e-point in the current cover (i.e., $S_{x',parent(x)} \in \mathscr{C}$ for some x'), the newly saturated $T(x)$ will be merged with $S_{x',parent(x)}$, replacing $S_{x',parent(x)}$ by $S_{x',sibling(x)}$. Similarly, if $sibling(x)$ is an i-point in the current cover (i.e., $S_{sibling(x),y'} \in \mathscr{C}$ for some y'), then $T(x)$ will be merged with $S_{sibling(x),y'}$. Hence, there are three possibilities regarding the reduction in the cover size $|\mathscr{C}|$:

- **0:** There will be no reduction if the subset $S_{parent(x),sibling(x)}$ cannot be merged with any other subset. This happens when neither $parent(x)$ is the e-point nor $sibling(x)$ is the i-point of any other subset in $\mathscr{C}$.
- **1:** A reduction of 1 will be obtained when the subset $S_{parent(x),sibling(x)}$ can only be merged with either $S_{x',parent(x)}$ or $S_{sibling(x),y'}$ for some x' or y'.
- **2:** As the best case, a reduction of 2 will be obtained when $S_{parent(x),sibling(x)}$ can be merged with both $S_{x',parent(x)}$ and $S_{sibling(x),y'}$, for some x', y'.

To decide which subset to remove next, the BottomUpAssign heuristic uses the *rate of return*, defined as the reduction in the cover size divided by the number of free riders needed. The heuristic dynamically maintains a priority queue SL of subsets in the current cover ordered according to their rate of return. Whenever a subset is to be removed, the first one in the queue is selected.

Algorithm 6. BottomUpAssign($\mathscr{T}, P, f$)

```
1: C ← SDExactAssign(T, P)
2: SL ← GetPQ(C, f)
3: while SL ≠ ∅
4:    repeat
5:       (x, y) ← ExtractFirst(SL)
6:    until r(y) ⩽ f
7:    C ← C − {S_{x,y}}
8:    Saturate(y)
9:    (x_new, y_new) ← Merge(C, SL, x)
10:   C ← C ∪ {S_{x_new, y_new}}
11:   Insert(SL, S_{x_new, y_new})
12:   f ← f − r(y)
```

Algorithm 7. Merge($\mathscr{C}, SL, x$)

```
1: if S_{x',parent(x)} ∈ C for some x' then
2:    x_new ← x'
3:    C ← C − {S_{x',parent(x)}}
4:    Remove(SL, S_{x',parent(x)})
5: else
6:    x_new ← parent(x)
7: if S_{sibling(x),y'} ∈ C for some y' then
8:    y_new ← y'
9:    C ← C − {S_{sibling(x),y'}}
10:   Remove(SL, S_{sibling(x),y'})
11: else
12:   y_new ← sibling(x)
13: return (x_new, y_new)
```

The GetPQ procedure produces the priority queue SL of $S_{x,y}$ subsets with $r(y) \leqslant f$, ordered according to their rate of return. The ExtractFirst procedure extracts the first subset $S_{x,y}$ in SL and returns the corresponding indices. The saturate procedure updates the r and rate of return values of all ascendants of y, rearranging SL accordingly.

Regarding the time complexity of the BottomUpAssign heuristic, finding the initial cover with the SDExactAssign procedure, which is Naor, Naor and Lotspiech's exact SD assignment algorithm, takes $O(r \log n)$ time [20]. Then, creation of the priority queue SL takes $O(r \log r)$ time. In the while loop, the ExtractFirst routine is called $O(r)$ times in total, among which at most f

lead to a set merger. The calls not leading to a merger will be completed in $O(r \log r)$ time in total. For the calls that lead to a merger, a run of INSERT, REMOVE and SATURATE may be needed per merger. INSERT and REMOVE take $O(\log r)$ time. SATURATE includes $O(\log n)$ decrease key operations, each of which may take $O(\log r)$ or $O(1)$ time depending on whether a binary or Fibonacci heap is used for implementing *SL*, making the total cost of the set merger operations $O(f \log n \log r)$ or $O(f \log n)$ accordingly. Therefore, the overall complexity of BOTTOMUPASSIGN is $O(r \log n + f \log n \log r)$ with a binary heap implementation and $O(r \log n)$ with a Fibonacci heap implementation of the priority queue *SL*.

5.3. Hybrid heuristic

The running time of the BOTTOMUPASSIGN heuristic increases significantly when the amount of the free rider quota to be placed is high. This problem can be solved by using the TOPDOWNASSIGN procedure to obtain an initial configuration and running BOTTOMUPASSIGN on top of it, instead of starting BOTTOMUPASSIGN with an exact SD cover and placing all free riders one by one. This combined method, which we call HYBRIDASSIGN, returns near-optimal solutions significantly faster than the original BOTTOMUPASSIGN.

6. Experimental results

We tested the practical performance of the algorithms in a series of simulation experiments, conducted with the parameters $n = 1024$, $1 \leqslant p \leqslant 1024$, and $0 \leqslant c_f \leqslant 2$. We summarize the results in this section. Each data point in the plots is averaged over 50,000 runs. At each run, a set of p users are selected randomly to be the privileged user set *P*. The free riders are chosen according to that *P* by the algorithm being tested. Then the SD cover is computed for the resulting receiver set, and that cover's cardinality is taken into account as the transmission cost for that run.

Figs. 2 and 3 compare the transmission costs obtained by the proposed algorithms against that of the basic SD scheme. Fig. 2 presents the results according to the privileged set size p for a set of selected c_f values. Fig. 3 presents the results according to the free rider ratio c_f.

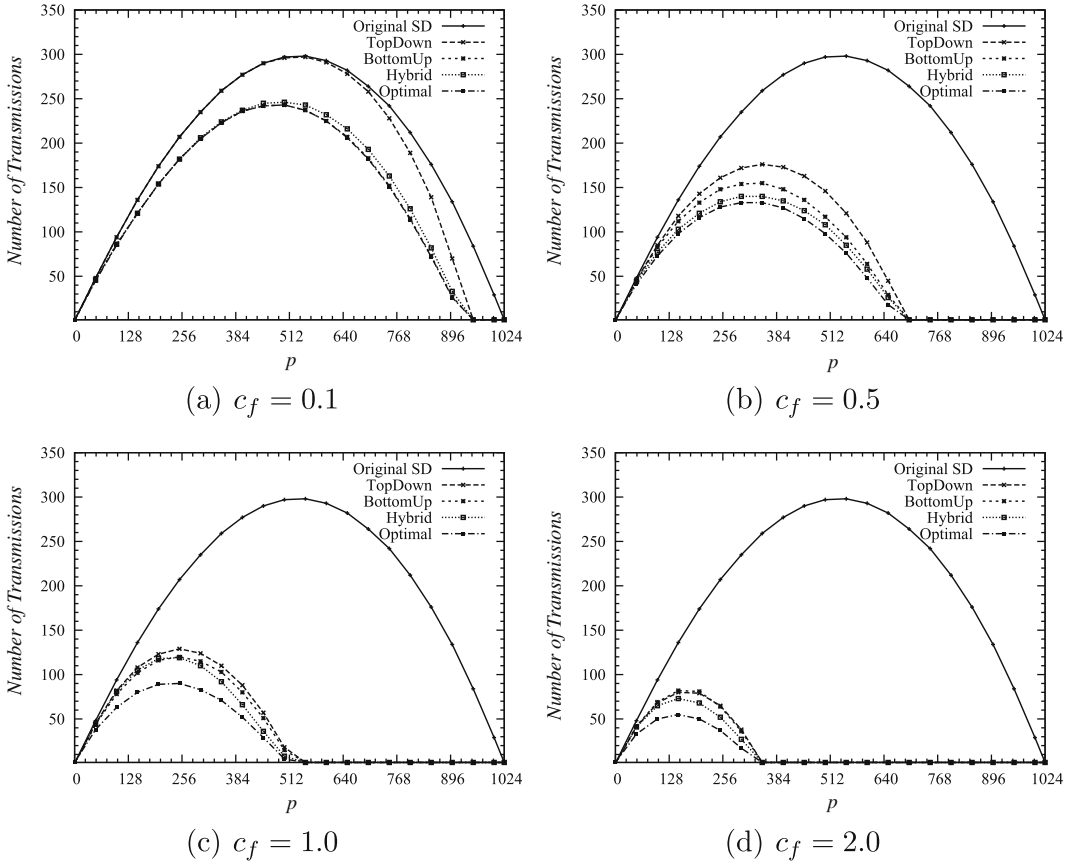

Fig. 2. Transmission costs of the algorithms with respect to p.

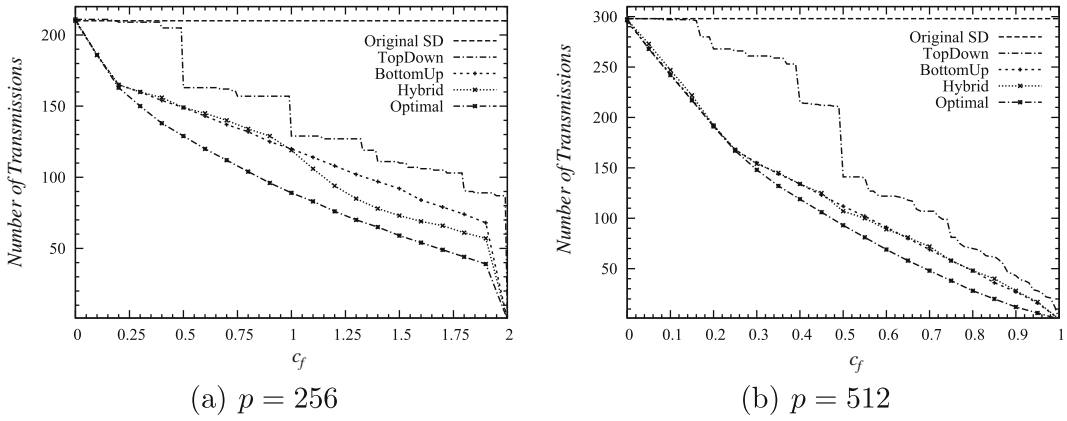

Fig. 3. Transmission costs of the algorithms with respect to c_f.

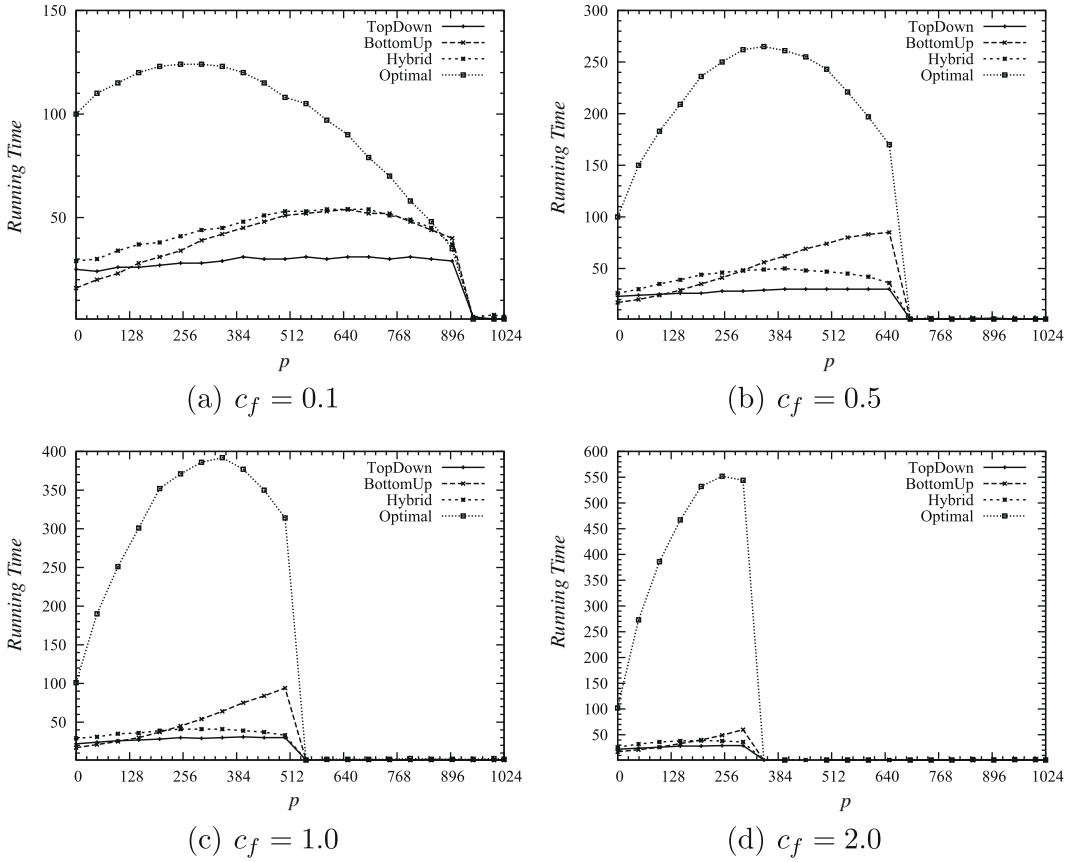

Fig. 4. Execution time of the algorithms in seconds. The figures are the total time of the 50,000 runs taken for each data point.

The results show that significant gains are possible by the proposed algorithms. With a limited free rider ratio such as 0.1, a 20% or greater reduction can be obtained; and when larger values of c_f are tolerable, a reduction of 80% or more is possible. The experiments also show that the results returned by the HYBRIDASSIGN heuristic are usually very close to the results obtained by the optimal algorithm. In the experiments, we also observed that if the distribution of the revoked users is uniform, then the distribution of the free riders is as well.

Fig. 4 compares the running times of our algorithms. The results show that HYBRIDASSIGN turns out to have the best cost-benefit performance among the heuristic methods. Its running time is only slightly more than that of TOPDOWNASSIGN, while its performance matches that of BOTTOMUPASSIGN and sometimes approaches that of the optimal algorithm.

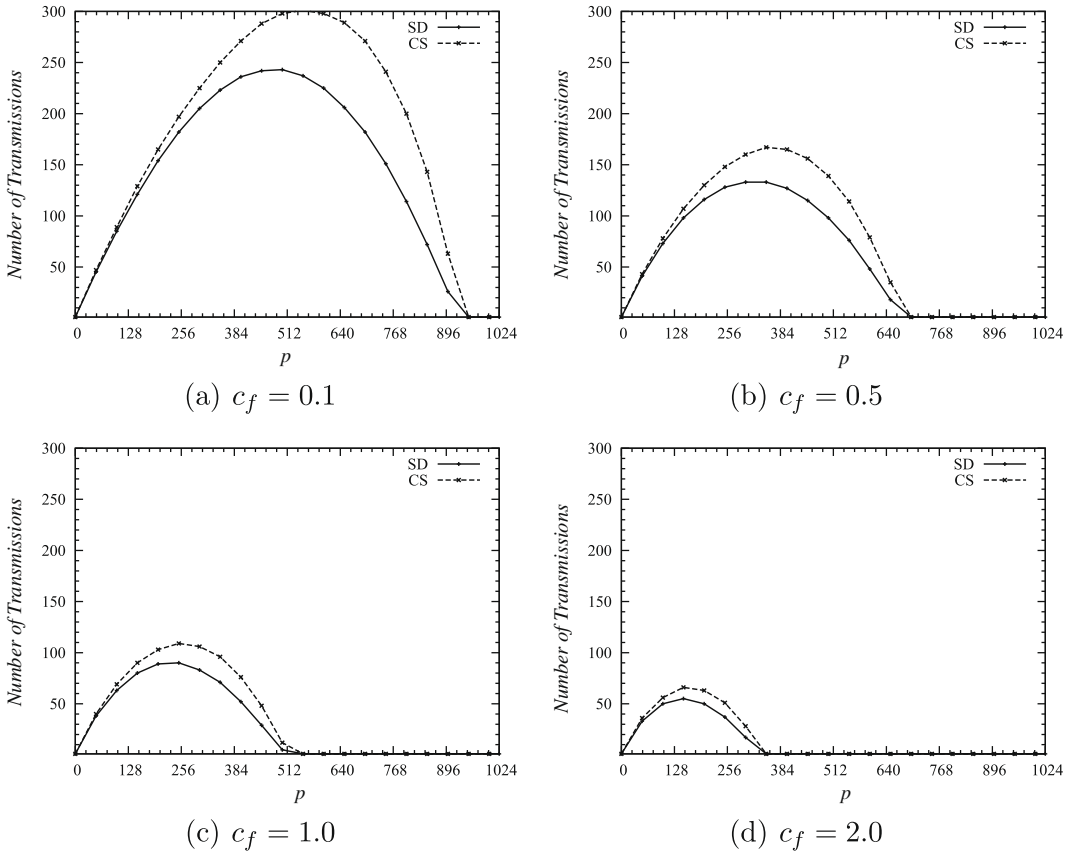

Fig. 5. Transmission costs obtained by the optimal algorithms for the CS and the SD schemes.

6.1. Comparison with the CS Scheme

An optimal free rider assignment algorithm for the CS scheme was given by Ramzan and Woodruff [22]. We also implemented this algorithm and compared it to our optimal algorithm for the SD scheme. Fig. 5 compares the performance of the two optimal algorithms in terms of the transmission cost. The results show that, with the same number of free riders allowed, the SD scheme can give a transmission cost of 20% less than that of the CS scheme.

7. Conclusion

The SD scheme is one of the most efficient BE schemes today. In this paper, we studied the problem of improving the performance of an SD scheme by allowing a limited number of free riders in the system. We first proposed an optimal algorithm based on a dynamic programming approach, which finds the best free rider placement that leads to the minimum transmission overhead. Subsequently, we proposed three heuristics for the same problem, that return near-optimal solutions with a faster running time. The TOPDOWNASSIGN heuristic works extremely fast, but it may not utilize all the available free rider quota, or it may spend a large amount of it fast and carelessly, possibly missing configurations that are more efficient. These drawbacks were solved in the BOTTOMUPASSIGN heuristic, which uses a targeted placement approach, placing the free riders slowly and carefully, and using all the available quota. However, this procedure gets slower as the free rider quota to be placed increases. Noting the advantages and disadvantages of the two procedures, we offered a third heuristic, HYBRIDASSIGN, that combines the advantages of the two approaches.

The experimental results show that the optimal placement algorithm and the three heuristics proposed provide significant reductions in the transmission cost of the SD scheme.

Besides the basic SD scheme, these algorithms can also be applied to its variants, such as LSD [13] and SSD [12]. These variants differ from the basic SD in the way they generate the keys of the tree, but they are exactly the same as the basic SD scheme as far as cover finding is concerned. Hence, the systems based on these SD variants can benefit equally from the proposed algorithms.

References

[1] AACS-Advanced Access Content System, 2007, <http://www.aacsla.com>.
[2] M. Abdalla, Y. Shavitt, A. Wool, Key management for restricted multicast using broadcast encryption, I EEE/ACM Transactions in Networking 8 (4) (2000) 443–454.
[3] W. Aiello, S. Lodha, R. Ostrovsky, Fast digital identity revocation, in: CRYPTO'98, vol. 1462, LNCS, Springer, 1998, pp. 137–152.
[4] S. Berkovits, How to broadcast a secret, in: EUROCRYPT'91, vol. 547, LNCS, Springer-Verlag, 1991, pp. 535–541.
[5] C. Blundo, A. Cresti, Unconditional secure conference key distribution schemes with disenrollment capability, Information Sciences 120 (1–4) (1999) 113–130.
[6] D. Boneh, C. Gentry, B. Waters, Collusion resistant broadcast encryption with shorter ciphertexts and private keys, in: CRYPTO'05, vol. 3621, LNCS, Springer-Verlag, 2005, pp. 258–275.
[7] D. Boneh, M. Hamburg, Generalized identity based and broadcast encryption schemes. In ASIACRYPT'08, vol. 5350, LNCS, Springer-Verlag, 2008, pp. 455–470.
[8] J.-T. Chung, C.-M. Li, T. Hwang, All-in-one group-oriented cryptosystem based on bilinear pairing, Information Sciences 177 (24) (2007) 5651–5663.
[9] V. Daza, J. Herranz, P. Morillo, C. Ráfols, Ad-hoc threshold broadcast encryption with shorter ciphertexts, Electronic Notes in Theoretical Computer Science 192 (2) (2008) 3–15.
[10] C. Delerableé, D. Pointcheval, Dynamic threshold public key broadcast encryption, in: CRYPTO'08, vol. 5157, LNCS, Springer-Verlag, 2008, pp. 317–334.
[11] A. Fiat, M. Naor, 1993, Broadcast encryption, in: CRYPTO'93, vol. 773, LNCS, Springer-Verlag, 1993, pp. 480–491.
[12] M.T. Goodrich, J.Z. Sun, R. Tamassia, Efficient tree based revocation in groups of low-state devices, in: CRYPTO'04, vol. 3152, LNCS, Springer-Verlag, 2004, pp. 511–527.
[13] D. Halevy, A. Shamir, The LSD broadcast encryption scheme, in: CRYPTO'02, vol. 2442, LNCS, London, UK, Springer-Verlag, 2002, pp. 47–60.
[14] J. Horwitz. A survey of broadcast encryption, Manuscript, 2003.
[15] M. Kusakawa, H. Hiwatari, T. Asano, S. Matsuda, Efficient dynamic broadcast encryption and its extension to authenticated dynamic broadcast encryption, in: CANS'08, vol. 5339, LNCS, Springer-Verlag, 2008, pp. 31–48.
[16] S.-T. Li, A platform-neutral live IP/TV presentation system, Information Sciences 140 (1–2) (2002) 33–52.
[17] Y.R. Liu, W.G. Tzeng, Public key broadcast encryption with low number of keys and constant decryption time, in: PKC'08, vol. 4939, LNCS, Springer-Verlag, 2008, pp. 380–396.
[18] J. Lotspiech, S. Nusser, F. Pestoni, Broadcast encryption's bright future, Computer 35 (2002) 57–63.
[19] J. Nam, J. Paik, U.M. Kim, D. Won, Resource-aware protocols for authenticated group key exchange in integrated wired and wireless networks, Information Sciences 177 (23) (2007) 5441–5467.
[20] D. Naor, M. Naor, J. Lotspiech, Revocation and tracing schemes for stateless receivers, in: CRYPTO'01, vol. 2139, LNCS, Springer-Verlag, 2001, pp. 41–62.
[21] J.H. Park, H.J. Kim, M.H. Sung, D.H. Lee, Public key broadcast encryption schemes with shorter transmissions, IEEE Transactions on Broadcasting 54 (3) (2008) 401–411.
[22] Z. Ramzan, D. Woodruff, Fast algorithms for the free riders problem in broadcast encryption, in: CRYPTO'06, vol. 4117, LNCS, Springer-Verlag, 2006, pp. 308–325.
[23] C.B.S. Traw, Protecting digital content within the home, Computer 34 (2001) 42–47.
[24] D.M. Wallner, E.J. Harder, R.C. Agee, Key management for multicast: issues and architectures, Internet Draft, 1999.
[25] C.K. Wong, M. Gouda, S.S. Lam, Secure group communication using key graphs, in: SIGCOMM'98, 1998, pp. 68–79.