Optimal jumping patterns

Steve Butler* Ron Graham† Nan Zang†

Abstract

We consider the problem of finding optimal "jumping" patterns from 1 to N where there is a cost associated with each jump. This will be done for two cost functions, in the first case the cost of jumping from a to b will be $(1-q^b)/a$ for $0<q<1$, while the second cost function will be b/a. For the first cost function we will show that all the jump lengths, except possibly the last jump, are between $\sqrt{2}$ and $19/4$. This will imply that the number of jumps in this case is of order $\Theta\big(\min(\ln N, -\ln\ln\frac{1}{q})\big)$. For the second cost function we will give some basic properties including bounds for the total cost of jumping from 1 to N.

Keywords: jumping sequences, cost function, key management

AMS classification code: 11B37

1 Introduction

Because of the tremendous increase in bandwidth in recent years, group-oriented broadcast services are becoming increasingly popular. These group broadcast network services, such as teleconferencing and pay-per-view TV, have a large number of subscribers and a central group controller (GC) that periodically broadcasts messages to all subscribers over an insecure channel. To guarantee that confidential information can be accessed only by the group members, a group key shared by all members of the group is needed. Messages sent by the GC will be encrypted with the group key and each subscriber will use the group key to read the messages.

For security reasons, the GC must maintain the group keys dynamically. For example, premium TV stations might stop subscribers from watching their encrypted content if they do not pay their subscription fees. However, when subscribers change frequently, the method for updating the group keys whenever a subscriber joins or leaves may be prohibitive. Li [3] suggested rekeying be done only periodically instead of immediately after each membership change (the rekeying period can be set by the security requirements). After a period of time,

*Department of Mathematics, University of California, San Diego, La Jolla, CA 92093 (sbutler@math.ucsd.edu).

†Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA 92093 (graham@ucsd.edu, nzang@cs.ucsd.edu).

the GC will generate some new keys as necessary and send encrypted messages to notify the remaining subscribers of the new keys. The rekeying cost is defined as the number of rekeying messages to be sent to accomplish this.

A natural problem which comes up is how to manage the keys so that the rekeying cost will be minimized. In the literature, a key tree structure is believed to be an efficient way to maintain the keys [3]. In a key tree structure, the GC is represented by a root, and the subscribers are represented by leaves. Associated with every node of the tree is an encryption key. The key associated with the root is used by the GC to broadcast confidential messages, also called the Traffic Encryption Key. All the other keys associated with non-root nodes are used to encrypt the rekeying messages, also called Key Encryption Keys. Each subscriber possesses all the keys along the path from the leaf representing the subscriber to the root.

Zhu et al. [5] introduced a new model based on some very popular network services. For those network services, they can only serve at most n subscribers and many customers wait to be served. After a period of time, the system will kick out a current subscriber with probability p. So, in this new model, the number of subscribers joining is assumed to be equal to the number of subscribers leaving during a batch updating period and every current subscriber has a probability p of being replaced by a new subscriber.

The rekeying cost in this model is defined as the total expected number of rekeying messages which have to be sent after the batch updating period. For a key tree structure T, each node v is associated with a key, Key_v, which is possessed by all the leaves under this node. We denote the number of leaves under v by $N(v)$. After a batch period, Key_v will change with probability $1-(1-p)^{N_v}$. So there are $d_v(1-q^{N_v})$ expected rekeying messages required to update the key Key_v, where d_v is the number of children of v and $q = 1-p$ (for the detailed sending procedure of the rekeying messages, see [2]). So, the rekeying cost is $\sum_{v\in V} d(v)(1-q^{N_v})$ where the sum is taken over all the nodes in T. It is convenient to remove the factor d_v from the formula by associating the weight $w(e) = 1-q^{N_v}$ with each edge $e = (v,u)$, where v is the parent of u. Then the cost of the tree T is $C(T) = \sum_{e\in E} w(e)$. The key tree management problem under this model is then to find a tree T with minimum weight. The minimum possible cost of such a tree with n leaves is defined as $OPT(p,n)$

To obtain a lower bound for the optimal tree cost, Li et al. [2] rewrote the cost function $C(T)$ in another form. Suppose tree T has n leaves. We distribute the weight of $C(T)$ to each leaf, and assign each leaf u a weight $c(p,u)$, which represents the cost per leaf. Then $c(p,u) = \sum_{e\in p(u,r)} w(e)$, where $p(u,r)$ is the (unique) path from u to the root r. Let $p(u,r)$ be $u = p_0, p_1 \dots p_{k-1}, r = p_k$. Using this we can rewrite $c(p,u)$ as

$$c(p,u) = \sum_{i=0}^{k-1} \frac{1-q^{N_{p_{i+1}}}}{N_{p_i}}.$$

We now see that $c(p,u)$ is uniquely determined by the sequence of numbers $\{N_{p_0}, N_{p_1}, \dots, N_{p_k}\}$, where $N_{p_0} = 1$ and $N_{p_k} = n$. It is easy to see that the minimum value of $c(p,u)$ will give a lower bound of minimum value of

$$OPT(p,n) \geq n \min c(p,u).$$

Li et al. [2] gave a lower bound for $c(p,u)$ when $p \to 0$ (equivalently, as $q \to 1$). In this paper, we reconsider the problem of bounding $c(p,u)$ by thinking of the N_{p_i} as giving a "jumping pattern". To better understand the behavior of the jumps we will remove the integer restriction, but we point out that the lower bound given by the real numbers will also give a lower bound in the case with integer restriction.

For a given value $0 < q < 1$ ($q = 1-p$), we want to minimize the cost of going from 1 to N (possibly ∞), where "moves" are jumps from a to b. Analogously to the $c(p,u)$ cost we will let the cost of a jump from a to b be given by $(1-q^b)/a$. So if our sequence of jumps is given by the (real) numbers $1 = a_0 < a_1 < a_2 < \cdots < a_k < N = a_{k+1}$ then the total cost of moving from 1 to N is given by

$$\begin{aligned} F(q, N; a_1, a_2, \ldots, a_k) &= \frac{1-q^{a_1}}{1} + \frac{1-q^{a_2}}{a_1} + \cdots + \frac{1-q^{a_k}}{a_{k-1}} + \frac{1-q^N}{a_k} \\ &= \sum_{i=0}^{k} \frac{1-q^{a_{i+1}}}{a_i}. \end{aligned} \tag{1}$$

We will denote the minimal cost of jumping from 1 to N by

$$F(q,N) = \min_k \min_{1<a_1<a_2<\cdots<a_k<N} F(q, N; a_1, a_2, \ldots, a_k). \tag{2}$$

We will proceed as follows. In Section 2 we will give some basic properties of $F(q,N)$ that follow from showing that the ratio of two consecutive terms of any jump sequence is bounded (this fact which will be established in Section 3). In Section 4 we consider a cost function associated with the limiting case of $q = 1$ and establish some basic properties of optimal jump sequences for that case. Finally we give some concluding remarks in Section 5

2 Basic properties of jump sequences

From the definition of jump sequences given in the introduction it is not clear what, if anything, can be said about consecutive jumps. Our main result is to show that with the exception of the last jump (i.e., the jump to N), that these jumps are both upper and lower bounded.

Theorem 1. *Let $1 < a_1 < a_2 < \cdots < a_k < N$ be an optimal jump sequence for a given value of q and N. Then for $i = 0, 1, \ldots, k-1$*

$$\sqrt{2} \le \frac{a_{i+1}}{a_i} < \frac{19}{4}.$$

In addition, if $N \le 2$ or $q < e^{-1/e}$ then the optimal thing to do is to jump straight from 1 to N. On the other hand if $N \ge 5$ and $q \ge 0.9$ then there will be at least one intermediate jump between 1 and N.

The proof of Theorem 1 will be given in Section 3. In this section we will give some consequences of this result. The first consequence is that the bound on jumps gives a natural lower bound for the cost. Namely, $F(q,N)$ is at least as much as the cost of the first jump. This gives us the following result.

Corollary 2. *Let $N \geq \sqrt{2}$, and $0 \leq q < 1$, then $F(q,N) \geq 1 - q^{\sqrt{2}}$.*

Optimal jumping patterns have a natural recursive relationship. This is helpful as it allows us to reduce the problem of how to make a jump to the problem of how to jump from 1.

Proposition 3. *Let $1 < a_1 < a_2 < \cdots < a_k < N$ be an optimal jump sequence associated with q and N. Then for any $1 \leq i \leq k$, the sequence $1 < a_{i+1}/a_i < a_{i+2}/a_i < \cdots < a_k/a_i < N/a_i$ is an optimal jump sequence for $q' = q^{a_i}$ and $N' = N/a_i$.*

Proof. The proof follows by noting that the cost function is optimal, and can be rewritten as

$$\begin{aligned} F(q,N) &= \sum_{j=0}^{k} \frac{1-q^{a_{j+1}}}{a_j} = \sum_{j=0}^{i-1} \frac{1-q^{a_{j+1}}}{a_j} + \sum_{j=i}^{k} \frac{1-q^{a_{j+1}}}{a_j} \\ &= \sum_{j=0}^{i-1} \frac{1-q^{a_{j+1}}}{a_j} + \frac{1}{a_i} \sum_{j=i}^{k} \frac{1-(q^{a_i})^{(a_{j+1}/a_i)}}{(a_j/a_i)} \end{aligned} \tag{3}$$

The second term in (3) is the cost function of the jump sequence associated with $1 < a_{i+1}/a_i < a_{i+2}/a_i < \cdots < a_k/a_i < N/a_i$ for $q' = q^{a_i}$ and $N' = N/a_i$. If this were not optimal we could replace it with the optimal cost function thus lowering $F(q,N)$, but this of course contradicts that we started with an optimal jump sequence. □

One consequence of this is that we can recursively construct $F(q,N)$ by finding the first jump. In the special case when $N = \infty$ we get the following recursive relationship.

Corollary 4. *The function $F(q,\infty)$ satisfies the following recursive relationship*

$$F(q,\infty) = \begin{cases} 1 & \text{if } q < e^{-1/e}; \\ \min\limits_{\sqrt{2}<a<19/4} \big(1 - q^a + \frac{1}{a} F(q^a,\infty)\big) & \text{otherwise.} \end{cases}$$

Perhaps the most important consequence of Theorem 1 is that it gives a natural bound on the number of jumps that can be taken. In particular it can be shown that the number of jumps from 1 to N is $\Theta\big(\min(\ln(N), -\ln\big(\ln(\frac{1}{q})\big))\big)$. So if we fix N and let $q \to 1$ then the number of jumps is of order $\ln(N)$. While if we fix q and let $N \to \infty$ then the number of jumps is of order $-\ln\big(\ln(\frac{1}{q})\big)$ (which is still finite, and indeed has a bounded maximum jump value).

Corollary 5. *Let $N < \infty$ and $q < 1$, and let k be the number of intermediate terms that are visited in the optimal jump sequence from 1 to N. Then*

$$\begin{aligned} k &\geq \min\big(0.64\ln(N) - 1.04, -0.65\ln\big(\ln(\frac{1}{q})\big) - 1.46\big), \\ k &\leq \min\big(2.89\ln(N) - 1, -2.88\ln\big(\ln(\frac{1}{q})\big) - 1.88\big). \end{aligned}$$

Proof. If $1 < a_1 < a_2 < \cdots < a_k < N$ is the optimal jump sequence then it follows from Theorem 1 that $\sqrt{2}^k \leq a_k \leq (4.75)^k$. To find a lower bound we note that by Theorem 1 we must continue jumping until either $(4.75)^k > N/5$ or $q^{(4.75)^k} < 0.9$. Solving for k in the first equation we have that $k > \big(\ln(N) - \ln(5)\big)/\ln(4.75) > 0.64\ln(N) - 1.04$. Solving for k in the second equation we first have that $(4.75)^k \ln(q) < \ln(0.9)$ or $(4.75)^k > \ln(0.9)/\ln(q)$, so that $k > \big(\ln\big(\ln(10/9)\big) - \ln\big(\ln(\frac{1}{q})\big)\big)/\ln(4.75) > -0.65\ln\big(\ln(\frac{1}{q})\big) - 1.46$

To find an upper bound we note that by Theorem 1 it is possible to continue jumping as long as $(\sqrt{2})^{k-1} < N/2$ and $q^{(\sqrt{2})^{k-1}} \geq e^{-1/e}$. Solving for k in the first equation we have that $k < \big(\ln(N) - \ln(2)\big)/\ln(\sqrt{2}) + 1 < 2.89\ln(N) - 1$. Solving for k in the second equation we first have that $(\sqrt{2})^{k-1} \leq 1/\big(e\ln(\frac{1}{q})\big)$, so that $k \leq \big(-\ln\big(\ln(\frac{1}{q})\big) - 1\big)/\ln(\sqrt{2}) + 1 < -2.88\ln\big(\ln(\frac{1}{q})\big) - 1.88$. □

3 Proof that jumps are bounded

In this section we will give the proof of Theorem 1. We will break the proof into a series of claims that will establish the upper and lower bounds as well as some special cases. The main technique is to compare costs of different sequences and show that one is always better, the difficulty will usually lie in showing what happens when q is close to 1 which requires we be careful with our analysis since the costs of all jumping sequences are near 0. We will first start with an observation which will be useful for proving the special cases.

Observation 1. *For a fixed q if the optimal jump sequence of length m beats the optimal jump sequence of length $m+1$, then the optimal jump sequence of length m beats the optimal jump sequence of length ℓ for each $\ell \geq m+1$.*

This observation follows by noting that for the optimal jump sequence of length ℓ the associated cost of the first $m+1$ terms is bounded by the total cost but is also at least as large as the cost of the optimal jump sequence of length $m+1$. The result then follows.

Claim 1. *If $N \leq 2$ or $q < e^{-1/e}$ then the optimal thing to do is to jump from 1 straight to N.*

Proof. From Observation 1 it suffices to show that the cost of jumping from 1 straight to N is better than the cost of jumping from 1 to x to N for all $1 < x < N$. Now suppose that $N \leq 2$, then this is equivalent to showing that

$$\frac{1-q^N}{1} < \frac{1-q^x}{1} + \frac{1-q^N}{x} \qquad \text{or} \qquad 0 < 1 - xq^x + (x-1)q^N.$$

We need to show $h_x(q) = 1 - xq^x + (x-1)q^N$ is positive in the range. Start by noting that $h_x(1) = 0$, it therefore suffices to show that $h'_x(q) < 0$ for $q < 1$ to establish the result. A calculation shows that

$$h'_x(q) = -x^2 q^{x-1} + N(x-1)q^{N-1} = xq^{N-1}\big(N - \frac{N}{x} - xq^{x-N}\big).$$

Now since $x - N < 0$ then $q^{x-N} > 1$ and we have

$$N - \frac{N}{x} - xq^{x-N} < N - \frac{N}{x} - x \le N - 2\sqrt{N} < 0,$$

where the last step is a simple minimization problem and uses $N \le 2$.

Now suppose that $q < e^{-1/e}$, then it is easy to check that $xq^x \le -1/e \ln q < 1$. So $0 < 1 - xq^x + (x-1)q^N$ is easily satisfied. □

Claim 2. *If $N \ge 5$ and $q \ge 0.9$ then there will be at least one intermediate jump between 1 and N.*

Proof. It suffices to show jumping from 1 to 2 to N gives a lower cost then jumping from 1 to N. This will hold if

$$\frac{1-q^N}{1} > \frac{1-q^2}{1} + \frac{1-q^N}{2} \qquad \text{or} \qquad 1 - 2q^2 + q^N < 0.$$

Since $1 - 2q^2 + q^N \le 1 - 2q^2 + q^5$ it suffices to show this holds when $N = 5$. Let $f(q) = 1 - 2q^2 + q^5$, we have that $f(0.9) = -0.02951$ while $f(1) = 0$ and $f''(q) = -4 + 20q^3 \ge 10.58$ for $0.9 \le q < 1$. Combining these establishes the result. □

We now turn to establishing the upper and lower bound. By Proposition 3 it will suffice to show that the bounds hold for the first jump.

Claim 3. *If an optimal jump sequence starts $1 < a < b < \cdots$ then $b > 2$.*

Proof. Since we are making an intermediate jump, by Claim 1 we may assume that $q \ge e^{-1/e} > 0.69$. Now suppose $b \le 2$, comparing the costs of jumping from 1 to b and from jumping 1 to a to b we have that the first sequence has lower cost if

$$\frac{1-q^b}{1} < \frac{1-q^a}{1} + \frac{1-q^b}{a}.$$

[All other terms associated with the cost are equal and drop out.] This is equivalent to $(a-1)q^b - aq^a + 1 > 0$. Since we assumed $b \le 2$ and $a - 1 > 0$ it suffices to show this holds when $b = 2$, i.e., it suffices to show

$$f(a) = (a-1)q^2 - aq^a + 1 > 0 \quad \text{for} \quad 0.69 \le q < 1, \text{ and } 1 < a < 2.$$

We have

$$f'(a) = q^2 - q^a - a\ln(q)\, q^a \quad \text{and} \quad f''(a) = -\ln(q)\, q^a(2 + a\ln(q)).$$

For the range given for a and q it is easy to check that $f''(a) > 0$, since we also have that $f(1) = 1 - q > 0$ it suffices to show that $f'(1) > 0$ and the result follows. Plugging in we have

$$g(q) = f'(1) = q^2 - q - q\ln(q).$$

Since $g(1) = 0$ to show this is positive we again do a similar trick. Namely by calculus we have

$$g'(q) = 2q - 2 - \ln(q) \quad \text{and} \quad g''(q) = 2 - \frac{1}{q}.$$

Since $g'(1) = 0$ and $g'' > 0$ in the range of q we are interested in the point is minimal, i.e., $g(q) > 0$ for $0.69 < q < 1$ and so the result follows, in particular skipping over a lowers the cost which is a contradiction. □

Claim 4. *If an optimal jump sequence starts* $1 < a < b < \cdots$ *then* $a \geq \sqrt{2}$.

Proof. Again we may assume $q \geq 0.69$. Now suppose that $a < \sqrt{2}$, comparing the costs of jumping from 1 to a to b with jumping from 1 to $\sqrt{2}$ to b we have that the second sequence has lower cost if

$$\frac{1 - q^{\sqrt{2}}}{1} + \frac{1 - q^b}{\sqrt{2}} < \frac{1 - q^a}{1} + \frac{1 - q^b}{a}.$$

[Again all other terms drop out.] By Claim 3, it again suffices to show this holds for $b = 2$, and so we want to show

$$f(a) = a\sqrt{2}(q^{\sqrt{2}} - q^a) + (\sqrt{2} - a)(1 - q^2) > 0 \quad \text{for} \quad 0.69 \leq q < 1, \text{ and } 1 < a < \sqrt{2}.$$

Using calculus we have

$$f'(a) = \sqrt{2}(q^{\sqrt{2}} - q^a) - a\sqrt{2}\ln(q)\, q^a - 1 + q^2 \quad \text{and} \quad f''(a) = -\sqrt{2}\ln(q)\, q^a(2 + a\ln(q)).$$

Again we have that $f'' > 0$ in the range that we are interested in. Since $f(\sqrt{2}) = 0$ it suffices to show that

$$g(q) = f'(\sqrt{2}) = -2\ln(q)q^{\sqrt{2}} - 1 + q^2 < 0.$$

Since $g(1) = 0$ to show that this is negative we again use calculus to get

$$g'(q) = 2q - 2\sqrt{2}\ln(q)q^{\sqrt{2}-1} - 2q^{\sqrt{2}-1} \quad \text{and}$$
$$g''(q) = 2 - 2\sqrt{2}(\sqrt{2} - 1)\ln(q)q^{\sqrt{2}-2} - (4\sqrt{2} - 2)q^{\sqrt{2}-2}.$$

Since $g'(1) = 0$ and $g'' < 0$ (the term with the $\ln(q)$ makes an insignificant contribution and the other terms then easily give a negative term) in the range that we are interested in the result follows, showing we would never take a first jump below $\sqrt{2}$. □

For the upper bound we will break it into two cases, namely when q is "small" and when q is "large". Of course we already know by Claim 1 that there are no intermediate jumps for $q < e^{-1/e}$ and so the upper bound holds vacuously in that range.

Claim 5. *If $q \leq 0.88$ then an optimal jump sequence makes at most one intermediate jump between 1 and n, and further such a jump is bounded above by 4.75.*

Proof. By Claim 1 we can assume $q \geq 0.69$. To show that there is at most one intermediate jump it suffices to show that the optimal cost of two intermediate jumps will never beat the optimal cost of one intermediate jump. To find the optimal two intermediate jump sequence and one intermediate jump sequence we minimize

$$q(x,y) = \frac{1-q^x}{1} + \frac{1-q^y}{x} + \frac{1-q^N}{y} \quad \text{and} \quad r(z) = \frac{1-q^z}{1} + \frac{1-q^N}{z}$$

respectively. In particular we need,

$$\begin{aligned} q_y(x,y) &= \frac{-\ln(q)q^y}{x} - \frac{1-q^N}{y^2} = \frac{-\ln(q)}{xy^2}\left(y^2q^y - \frac{x(1-q^N)}{-\ln(q)}\right) = 0, \\ r'(z) &= -\ln(q)q^z - \frac{1-q^N}{z^2} = \frac{-\ln(q)}{z^2}\left(z^2q^z - \frac{1-q^N}{-\ln(q)}\right) = 0. \end{aligned}$$

Given that we know $q \geq 0.69$ and the shape of the curve t^2q^t there will be two possible solutions for y and z, since we are trying to minimize we will want to find the *first* such solution in each case. In particular since $x > 1$ it follows that we have $y > z$.

We then have that the one jump sequence always beats out a two jump sequence if for any x and y we can find a z so that

$$\frac{1-q^x}{1} + \frac{1-q^y}{x} + \frac{1-q^N}{y} > \frac{1-q^z}{1} + \frac{1-q^N}{z},$$

or rearranging

$$\left(\frac{1-q^x}{1} + \frac{1-q^y}{x} + \frac{1}{y}\right) - \left(\frac{1-q^z}{1} + \frac{1}{z}\right) > \underbrace{q^N\left(\frac{1}{y} - \frac{1}{z}\right)}_{<0}.$$

So it is certainly sufficient to show that the left hand side is > 0. This reduces it down to the case $N = \infty$. By using a computer algebra system it can be checked that this will hold for $q \leq 0.88$.

Finally, the optimal jump will be the first solution to $z^2q^z + (1-q^N)/\ln(q) = 0$ which will certainly occur before the solution to $z^2q^z + 1/\ln(q) = 0$. Plotting $(4.75)^2q^{4.75} + 1/\ln(q)$ in the range $0.69 \leq q \leq 0.88$ we see that it is positive indicating that the solution occurs before 4.75, i.e., the optimal jump has length bounded above by 4.75. □

In Claim 5 we needed to show that if $1 < x < y < N$ was an optimal jump sequence with two intermediate jumps and $1 < z < N$ was an optimal jump sequence with one intermediate jump then $z < y$. Its seems reasonable that this should generalize as follows.

Conjecture 1. *If $1 < a_1 < \cdots < a_m < N$ is an optimal jump sequence with m intermediate jumps and $1 < b_1 < \cdots < b_{m+1} < N$ is an optimal jump sequence with $m+1$ intermediate jumps then $b_i < a_i < b_{i+1}$ for $i = 1, \ldots, m$.*

Claim 6. *If an optimal jump sequence starts $1 < b < \cdots$ and $q \geq 0.88$ then $b < 4.75$.*

Proof. Now suppose that an optimal jump sequence starts by jumping from 1 to b where $b > 4.75$. Then we claim that it is better to start by jumping from 1 to 2 to b. This last statement holds if

$$\frac{1-q^2}{1} + \frac{1-q^b}{2} < \frac{1-q^b}{1}.$$

[Again all other terms drop out.] Rearranging, this will be equivalent to showing $0 < 2q^2 - q^b - 1$. It suffices to show that this last statement holds for $b = 4.75$, i.e., it suffices to show

$$g(q) = 2q^2 - q^{4.75} - 1 > 0 \quad \text{for} \quad 0.88 \leq q < 1.$$

Note that $g(0.88) = 0.00393\ldots > 0$ and $g(1) = 0$. Since $g'(q) = 4q - 4.75q^{3.75}$ and $g''(q) = 4 - 17.8125q^{2.75}$ we have that the graph is concave down for our range of q, the result then follows. This shows that we would never jump more than 4.75 in an optimal jump sequence. □

4 The limiting case as $q \to 1$

As $q \to 1$ then $F(q, N) \to 0$ (i.e., since all jumps now have cost 0). So for this limiting case we should consider a different type of cost function. To determine which cost function, we start by recalling $q = 1 - p$ and note that $1 - q^x = px + O(p^2)$ by the binomial theorem. So that (1) becomes

$$F(q, N; a_1 a_2, \ldots, a_k) = \sum_{i=0}^{k} \frac{1 - q^{a_{i+1}}}{a_i} = p \sum_{i=0}^{k} \frac{a_{i+1}}{a_i} + O(p^2).$$

The obvious candidate is to use the first order term as our new cost function, i.e., so that the cost of a jump from a to b will be given by b/a. Given a sequence of jumps, the total cost of moving from 1 to N will be given by

$$\begin{aligned} G(N; a_1, a_2, \ldots, a_k) &= \frac{a_1}{1} + \frac{a_2}{a_1} + \cdots + \frac{a_k}{a_{k-1}} + \frac{N}{a_k} \\ &= \sum_{i=0}^{k} \frac{a_{i+1}}{a_i}, \end{aligned} \tag{4}$$

and we will denote the minimal cost of jumping from 1 to N by

$$G(N) = \min_k \min_{1 < a_1 < a_2 < \cdots < a_k < N} G(N; a_1, a_2, \ldots, a_k). \tag{5}$$

4.1 Jumping along the reals

We begin by noting that by the arithmetic-geometric mean inequality, for any jump sequence we have

$$G(N;a_1,a_2,\ldots,a_{k-1}) \geq k\sqrt[k]{\frac{a_1}{1}\frac{a_2}{a_1}\cdots\frac{a_{k-1}}{a_{k-2}}\frac{N}{a_{k-1}}} = k\sqrt[k]{N},$$

with equality holding if and only if $a_1/1 = a_2/a_1 = a_3/a_2 = \cdots$. Note that for any $N > 1$ the function $f(x) = xN^{1/x}$ is concave up for $x > 0$ and so has a unique minimum which occurs at $f(\ln N) = e\ln N$.

Proposition 6. *When jumping along the reals we have that*

$$G(N) = \min\left\{\lfloor\ln N\rfloor N^{1/\lfloor\ln N\rfloor}, \lceil\ln N\rceil N^{1/\lceil\ln N\rceil}\right\} = e\ln N + O\big(\frac{1}{\ln N}\big).$$

While the optimal jumping patterns are formed by geometric sequences with ratio $e^{1+o(1)}$.

Proof. The form of $G(N)$ and the formation of a geometric series with a ratio of either $N^{1/\lfloor\ln N\rfloor} = e^{\ln N/\lfloor\ln N\rfloor} = e^{1+o(1)}$ or $N^{1/\lceil\ln N\rceil} = e^{\ln N/\lceil\ln N\rceil} = e^{1+o(1)}$ follows from the statements preceding the proposition. It remains to verify the asymptotic behavior. So suppose that $\lfloor\ln N\rfloor = \ln N - \alpha$ then we have

$$\begin{aligned}\lfloor\ln N\rfloor N^{1/\lfloor\ln N\rfloor} - e\ln N &= \big(\ln N - \alpha\big)e^{\ln N/(\ln N-\alpha)} - e\ln N\\ &= e\ln N\big(e^{\alpha/(\ln N-\alpha)} - 1\big) - \alpha e\, e^{\alpha/(\ln N-\alpha)}\\ &= e\ln N\left(\frac{\alpha}{\ln N-\alpha} + O\big(\frac{1}{\ln N^2}\big)\right) - \alpha e + O\big(\frac{1}{\ln N}\big) = O\big(\frac{1}{\ln N}\big).\end{aligned}$$

A similar statement holds for when $\lceil\ln N\rceil = \ln N + \alpha$, the proof now follows. □

4.2 Jumping along the integers

We will let $G_{\mathbb{Z}}(N)$ denote the minimal cost for the case of when we restrict jumps to the integers. Clearly, $G_{\mathbb{Z}}(N) \geq G(N) = e\ln N + O(1/\ln N)$. However, with the restriction of jumping only on the integers we should expect the cost to go up. In this section will look at the asymptotic behavior of $G_{\mathbb{Z}}(N)$. First though we will introduce an important integer sequence, and some of its properties, that will play a role in the behavior of the function.

Lemma 7. *Let $a(n)$ be the integer sequence such that $a(0) = 1$ and $a(n) = \lfloor e\cdot a(n-1)+0.5\rfloor$ for $n \geq 1$ (i.e., to get the next term multiply by e and round). Then the following holds:*

- *$a(n) = \lfloor\gamma e^n + 0.5\rfloor$ for $\gamma = 1.098002099366832827899136351\ldots$*
- *$\lim\limits_{n\to\infty}\left(\sum\limits_{i=1}^{n}\dfrac{a(i)}{a(i-1)} - e\ln\big(a(n)\big)\right) = \alpha = 0.014357537447198206167909857\ldots$*

This sequence, which starts $\{1, 3, 8, 22, 60, 163, 443, 1204, 3273, 8897, \ldots\}$, is `A024581` in the `OEIS`. Essentially what it does is try to best approximate jump lengths of e where after every jump forward we reset (as compared to taking the nearest integer to powers of e which would be the sequence $\{1, 3, 7, 20, 55, 148, 403, 1097 \ldots\}$).

Proof. For the first part we let $b(n) = a(n)/e^n$. We first show that this sequence converges. For $n \geq 1$,

$$\begin{aligned} |b(n) - b(n-1)| &= \left| \frac{a(n)}{e^n} - \frac{a(n-1)}{e^{n-1}} \right| \\ &= \left| \frac{\lfloor e{\cdot}a(n-1) + 0.5 \rfloor}{e^n} - \frac{e{\cdot}a(n-1)}{e^n} \right| \\ &= e^{-n} \big| \lfloor e{\cdot}a(n-1) + 0.5 \rfloor - e{\cdot}a(n-1) \big| \;\leq\; \frac{1}{2} e^{-n}, \end{aligned}$$

where the last step follows from noting that no number is more than $1/2$ away from the nearest integer (this is what the inside of the absolute value is expressing). This implies the sequence is Cauchy and so must converge. Let $\gamma = \lim_{n \to \infty} b(n)$. Then note that for any n

$$\big|b(n) - \gamma\big| \leq \sum_{k=n+1}^{\infty} \big|b(k) - b(k-1)\big| \leq \sum_{k=n+1}^{\infty} \frac{1}{2} e^{-k} = \frac{\frac{1}{2} e^{-(n+1)}}{1 - \frac{1}{e}} < \frac{0.3}{e^n}.$$

Multiplying both sides by e^n, this implies that $\big|a(n) - \gamma e^n\big| < 0.3$. Since $a(n)$ is an integer and γe^n is less than 0.3 away it must be that $a(n)$ is the nearest integer to γe^n. From this it can be shown that $1.05e^n < a(n) < 1.15e^n$.

For the second part let

$$c(n) = \sum_{i=1}^{n} \frac{a(i)}{a(i-1)} - e \ln\big(a(n)\big),$$

and consider the following

$$\begin{aligned} c(n) - c(n-1) &= \frac{a(n)}{a(n-1)} - e \ln\left(\frac{a(n)}{a(n-1)}\right) \\ &\leq \frac{e{\cdot}a(n-1) + 0.5}{a(n-1)} - e \ln\left(\frac{e{\cdot}a(n-1) - 0.5}{a(n-1)}\right) \\ &= \frac{1}{2a(n-1)} - e \ln\left(1 - \frac{1}{2e{\cdot}a(n-1)}\right) \\ &\leq \frac{1}{2a(n-1)} + \frac{1}{a(n-1)} \;\leq\; \frac{3e}{2e^n}. \end{aligned}$$

(Here we used the fact that for $0 \leq x \leq 1/2e$ that $-\ln(1-x) \leq 2x$.) On the other hand it is easy to see that $c(n)$ is increasing, i.e., since $x - e \ln x$ has a minimum value of 0 at $x = e$, so $c(n) - c(n-1) \geq 0$. Combining we have $\big|c(n) - c(n-1)\big| \leq 3e/2e^n$ from which

the convergence of $c(n)$ easily follows. This shows that $c(n)$ converges to α with an error of order e^{-n}, a little more careful analysis can show that the error is at most $0.05e^{-2n}$.

The numerical approximation for γ and α can be found by the computing the corresponding terms for sufficiently large n. □

Theorem 8. *When jumping along the integers we have that*

$$G_{\mathbb{Z}}(N) \leq e \ln N + \alpha + O(\frac{1}{\ln N}),$$

where α is the constant introduced in Lemma 7.

Proof. We construct an approximate optimal sequence for N by letting $\ell = \lfloor \ln \ln N \rfloor$ and then take the first ℓ terms from the integer sequence from Lemma 7 to form the initial part of the sequence, $a(1), a(2), \ldots, a(\ell)$. From the proof of Lemma 7 we have that $0.38 \ln N \leq a(\ell) \leq 1.15 \ln N$, i.e., we have that $a(\ell) = \Theta(\ln N)$. Now starting at $a(\ell)$ find the optimal jump sequence jumping to N along the *real* numbers and then round each term to the nearest integer, to form the rest of the sequence. We denote this remaining part of the sequence by $b(0) = a(\ell), b(1), b(2), \ldots, b(k), b(k+1) = N$.

We now need to bound the cost of the jump sequence $a(1), \ldots, a(\ell), b(1), \ldots, b(k)$.

From Lemma 7, the first part of the jump sequence has cost bounded by $e \ln(a(\ell)) + \alpha - o(1)$. It is also easy to adapt the proof of Proposition 6 to see that the cost of the optimal jump sequence along the reals from $a(\ell)$ to N is $e \ln \big(N/a(\ell)\big) + O(1/\ln N)$. It remains to show that the error from rounding the terms to the nearest integer is of order $O(1/\ln N)$.

By Proposition 6 we know that the sequence $b(i) = \lfloor a(\ell)\beta^i + 0.5 \rfloor$ where $\beta \approx e$. A simple calculation shows that

$$\left| \beta - \frac{b(i+1)}{b(i)} \right| \leq \frac{\beta + 1}{2b(i)} \leq \frac{\beta + 1}{a(\ell)\beta^i}.$$

From this it follows that the error we get from rounding to the nearest integer is bounded by

$$\sum_{i=0}^{k} \left| \beta - \frac{b(i+1)}{b(i)} \right| \leq \sum_{i=0}^{\infty} \frac{\beta + 1}{a(\ell)\beta^i} = \frac{\beta(\beta+1)}{a(\ell)(\beta - 1)} = O(\frac{1}{\ln N}).$$

□

Looking at the statement of Theorem 8, a natural question to ask is whether we need the α term. Clearly, by the definition of α we know that for every $N = a(i)$, i.e., every N which shows up in the specified integer sequence, we have that the total cost of jumping to N is $< e \ln N + \alpha$. On the other hand, the smallest $N \neq a(i)$ and with cost of jumping to N below $e \ln N + \alpha$ is $N = a(212) + 1$, or,

$$1291318384051937580859421657451339543637942426407958928241564048452309868616554970528509821 85.$$

Using this along with the rate of convergence to α we have the following result, which indicates that if α is not the correct term for Theorem 8, it is close.

Proposition 9. *Except for the first approximately* 211 *terms from the sequence* $a(i)$*,*

$$G_{\mathbb{Z}}(N) > e \ln N + \alpha - \frac{1}{10^{180}}.$$

5 Concluding remarks

In this paper we have looked at the problem of jumping from 1 to N while minimizing costs. When jumping along the reals we saw that each jump was bounded between $\sqrt{2}$ and $19/4$. If we relate this back to the problem of key tree management this seems to indicate that if v is a parent of u then the number of descendants of v is between $\sqrt{2}$ and $19/4$ times greater than the number of descendants of u (except possibly the root). This agrees well with the known fact that every internal node of an internal key tree (except possibly the root) has degree 2, 3 or 4 (see [2]).

The jumping sequence gives a rough approximation for the optimal tree with the depth of the tree being the number of jumps. We saw that the number of jumps for N fixed and q approaching 1 is of order $\ln N$ while for q fixed and N large is of order $-\ln \ln \frac{1}{q}$.

The constants given in Theorem 1 are not optimal. One interesting problem would be to know what the best constants are, in particular what are the best upper and lower bounds for the jumps.

Our analysis has focused on two different cost functions involved with jumping from a to b, i.e., $(1-q^b)/a$ and b/a. There are many different possible cost functions and it would be interesting to see what combinatorial properties different cost functions have. The authors will be looking at this problem in a future work.

References

[1] R. L. Graham, M. Li and F. F. Yao, *Optimal tree structures for group key management with batch updates*, SIAM J. on Disc. Math **21** (2007), 532–547.

[2] M. Li, Z. Feng, N. Zang, R. Graham and F.F. Yao, *Approximately optimal trees for group key management with batch updates*, to appear.

[3] X. S. Li, Y. R. Yang, M. G. Gouda, and S .S. Lam, *Batch Re-keying for Secure Group Communications*, WWW10, May 2-5, 2001, Hong Kong.

[4] N. J. A. Sloane, **On-line Enyclopedia of Integer Sequences**, `http://www.research.att.com/~njas/sequences/`

[5] F. Zhu, A. Chan, and G. Noubir, *Optimal Tree Structure for Key Management of Simultaneous Join/Leave in Secure Multicast*, Proceedings of MILCOM, 2003.