北京大学学报(自然科学版) ,第 40 卷,第 1 期,2004 年 1 月
Acta Scientiarum Naturalium
Universitatis Pekinensis, Vol.40, No.1 (Jan, 2004)

四柱汉诺塔之初步探究

杨　楷　　徐　川
(北京大学计算机科学与技术系,北京,100871)

摘　要　1941 年,J. S. Frame 在《美国数学月刊》上提出了一种解决四柱汉诺塔问题的算法,但未给出最终公式的证明。本文按照这种算法总结出完成四柱汉诺塔游戏之最少步数的公式,并用数学归纳法证明了它。

关键词　四柱汉诺塔; 区; 剩余盘数 $R(n)$
中图分类号　O 224

0　引　言

汉诺塔问题是一个古老的数学问题。经典汉诺塔问题是三柱的。它起源于印度。意思是:第一个柱(A 柱)上有 n 个碟子,从底向上碟子大小依次减小,目标是通过第二个柱(B 柱)把所有碟子移到第三个柱(C 柱)上,但不能把大的碟子放到小的碟子上面。三柱汉诺塔经典算法是这样的:首先用三柱汉诺塔经典算法把 A 柱上面的 $n-1$ 个碟子通过 C 柱移到 B 柱上,然后把 A 柱剩下的一个碟子移到 C 柱上,最后用三柱汉诺塔经典算法把 B 柱上所有的碟子通过 A 柱移到 C 柱上[1]。

设完成 n 个碟子的三柱汉诺塔游戏需要移动的步数为 $T(n)$,则

$$T(n) = 2\times T(n-1)+1,\quad T(1) = 1。\tag{1}$$

因此,

$$T(n) = 2^n - 1。\tag{2}$$

如果 $n = 64$,则 $T(64) = 2^{64} - 1\approx 1.845\times 10^{19}$ 步。

四柱汉诺塔问题和三柱汉诺塔问题的惟一区别就是增加了一个柱(D 柱),如图 1 所示,目标是把 A 柱上的 n 个碟子通过 B 柱和 C 柱移到 D 柱上。虽然四柱汉诺塔只比三柱汉诺塔多一个柱,但是解决它的难度远大于三柱汉诺塔。

四柱汉诺塔的 Frame 算法是这样的:

对于碟数为 n 的四柱汉诺塔,假定碟数 i 小于 n 时的算法已经确定,i 个碟子的四柱汉诺塔需要移动的步数为 $F(i)$。可把 A 柱上的碟子分成上、下两部分[2]。下部分共有 r 个碟子,上部分 $n-r$ 个碟子,$1\le r\le n$。操作步骤如下:

①用四柱汉诺塔 Frame 算法把 A 柱上部分的 $n-r$ 个碟子通过 C 柱和 D 柱移到 B 柱上,需要 $F(n-r)$ 步;

收稿日期: 2003-02-27; 修回日期: 2003-06-10

②用三柱汉诺塔经典算法把 A 柱上剩余的 r 个碟子通过 C 柱移到 D 柱上,需要 $T(r) = 2^r - 1$ 步;

③用四柱汉诺塔 Frame 算法把 B 柱上的 $n - r$ 个碟子通过 A 柱和 C 柱移到 D 柱上,需要 $F(n - r)$ 步;

据此,计算出总步数为 $f(n, r)$,随后对所有的 $r(1 \le r \le n)$ 逐一进行尝试。选择一个 r 使得 $f(n, r)$ 取最小值,并定义此时的 r 为 $R(n)$。这样,就可以确定完成 n 个碟子的四柱汉诺塔游戏需要的最少步数[2]:

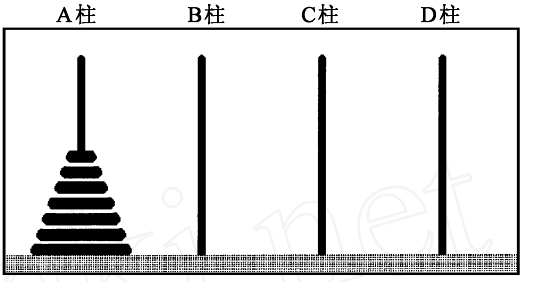

图 1　四柱汉诺塔的图示

Fig. 1　The illustration of 4-peg Hanoi Tower

$$F(n) = \min_r f(n, r) = \min_r [2F(n - r) + T(r)] = \min_r [2F(n - r) + 2^r - 1]。\quad (3)$$

1　有关数据及公式

根据上述算法,不难得到如下的数据(见表 1):

表 1　R(n) 与 F(n)

Table 1　$R(n)$ and $F(n)$

n	$R(n)$	$F(n)$
1	1	1
2	1	3
3	2	5
4	2	9
5	2	13
6	3	17
7	3	25
8	3	33
9	3	41
10	4	49
11	4	65
12	4	81
13	4	97
14	4	113
15	5	129
16	5	161
17	5	193
18	5	225
19	5	257
20	5	289
21	6	321

根据表 1 的数据,作者猜想:

$$F(n) = \left[n - \frac{R^2(n) - R(n) + 2}{2} \right] \cdot 2^{R(n)} + 1, \quad (4)$$

其中
$$R(n) = \left\lfloor \frac{\sqrt{8n + 1} - 1}{2} \right\rfloor。\quad (5)$$

本文的目的就是要证明(3),(5) ⇒(4)。

2　公式的证明

从上面的表格不难看出,虚线按照公式(5)将所有自然数 n 划分成区,每区中所有的 n 对应同一个 $R(n)$,该区简称为 R 区。如 $n = 3$, $n = 4$ 和 $n = 5$ 在同一个区,$R(n)$ 都等于 2。

证明(3),(5) ⇒(4)之前,先给出一个引理。

引理　若对 $n < \frac{(R_0 + 1)(R_0 + 2)}{2}$,有 $F(n) = \left[n - \frac{R(n)^2 - R(n) + 2}{2} \right] \cdot 2^{R(n)} + 1$,则对于落在 $R_0 + 1$ 区的任意一个 n,当 $n - r$ 不落在 $R_0 + 1$ 区时,$f(n, r)$ 是关于 r 的连续函数。

(引理的证明见附录)

下面用数学归纳法证明(3),(5) ⇒(4):

(Ⅰ) 显然,$n = 1$ 时,$R(n) = 1$, $F(n) = 1$。

$n=2$ 时, $R(n)=1$,　$F(n)=3$。

所以,对于 $n=1$ 和 $n=2$,公式(4)是正确的。

(Ⅱ) 假设对于 $n<\frac{(R_0+1)(R_0+2)}{2}$ 内的所有 n,

都有 $F(n)=\left[n-\frac{R^2(n)-R(n)+2}{2}\right]\cdot 2^{R(n)}+1$。($R_0$ 为自然数)

现考虑 $\frac{(R_0+1)(R_0+2)}{2}\leq n<\frac{(R_0+2)(R_0+3)}{2}$ 内所有 n。可将这些 n 分为以下 3 种情况讨论:

① $n=\frac{(R_0+1)(R_0+2)}{2}$;

② $\frac{(R_0+1)(R_0+2)}{2}<n\leq\frac{(R_0+2)(R_0+3)}{2}-2$;

③ $n=\frac{(R_0+2)(R_0+3)}{2}-1$。

先讨论 ①:

当 $n=\frac{(R_0+1)(R_0+2)}{2}$ 时, $n-r$ 必然落在 R_0 区或 R_1 区 ($R_1<R_0$)。

(所谓 $n-r$ 落在 R_1 区,即 $n-r<\frac{R_0(R_0+1)}{2}$)。

A: 若 $n-r$ 落在 R_0 区,

即 $\frac{R_0(R_0+1)}{2}\leq\frac{(R_0+1)(R_0+2)}{2}-r<\frac{(R_0+1)(R_0+2)}{2}$,有 $0<r\leq R_0+1$。

$$\therefore\ f(n,r)=2F(n-r)+2^r-1=2\left[\left(n-r-\frac{R_0^2-R_0+2}{2}\right)\cdot 2^{R_0}+1\right]+2^r-1。$$

$$\therefore\ \frac{\partial f}{\partial r}=-2^{R_0+1}+\ln 2\cdot 2^r,\quad 令\ \frac{\partial f}{\partial r}=0,\quad 则\ r=R_0+1-\frac{\ln\ln 2}{\ln 2}=R_0+1.529。$$

当 $r<R_0+1.529$ 时, $\frac{\partial f}{\partial r}<0$;　当 $r>R_0+1.529$ 时, $\frac{\partial f}{\partial r}>0$。

$$\therefore\ 当\ r\leq R_0+1\ 时,\ \frac{\partial f}{\partial r}<0。\tag{6}$$

B: 若 $n-r$ 落在 R_1 区,即 $\frac{(R_0+1)(R_0+2)}{2}-r<\frac{R_0(R_0+1)}{2}$,有 $r>R_0+1$。

$$f(n,r)=2F(n-r)+2^r-1=2\left[\left(n-r-\frac{R_1^2-R_1+2}{2}\right)\cdot 2^{R_1}+1\right]+2^r-1。$$

$\because$ $R_1<R_0$ 即 $R_1+1\leq R_0$,

$$\therefore\ \frac{\partial f}{\partial r}=-2^{R_1+1}+\ln 2\cdot 2^r\geq -2^{R_0}+\ln 2\cdot 2^r>-2^{R_0}+2^{r-1}。$$

$$\therefore\ 当\ r>R_0+1\ 时,\ \frac{\partial f}{\partial r}>0。\tag{7}$$

综合(6),(7)和引理,当且仅当 $r=R_0+1$ 时, $f(n,r)$ 达到最小值。

而且,$f(n, R_0+1) = \left(n - \frac{R_0^2 + R_0 + 2}{2}\right) \cdot 2^{R_0+1} + 1$,

这样,

$$F(n) = F\left[\frac{(R_0+1)(R_0+2)}{2}\right] = \left(n - \frac{R_0^2 + R_0 + 2}{2}\right) \cdot 2^{R_0+1} + 1 = R_0 \cdot 2^{R_0+1} + 1。\quad (8)$$

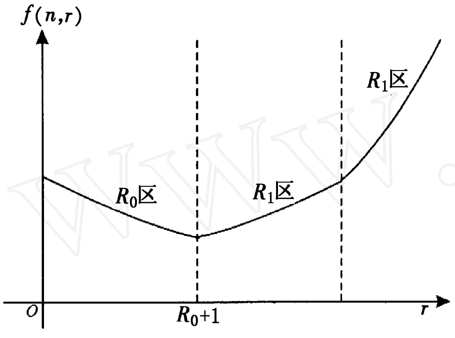

图 2 情况 ①中 $f(n,r)$ 与 r 的关系示意图

Fig. 2 The relationship between $f(n, r)$ and r in the first case

再讨论 ②:

当 $\frac{(R_0+1)(R_0+2)}{2} < n \leq \frac{(R_0+2)(R_0+3)}{2} - 2$ 时,$n - r$ 必然落在 R_0 区或 R_1 区或 R_0+1 区($R_1 < R_0$)。

A:若 $n - r$ 落在 R_0 区,即 $\frac{R_0(R_0+1)}{2} \leq n - r < \frac{(R_0+1)(R_0+2)}{2}$。

$\therefore$ $f(n, r) = 2F(n - r) + 2^r - 1 = 2 \cdot \left[\left(n - r - \frac{R_0^2 - R_0 + 2}{2}\right) \cdot 2^{R_0} + 1\right] + 2^r - 1$。

$\therefore$ $\frac{\partial f}{\partial r} = -2^{R_0+1} + \ln 2 \cdot 2^r$。

令 $\frac{\partial f}{\partial r} = 0$,则 $r = R_0 + 1 - \frac{\ln\ln 2}{\ln 2} = R_0 + 1.529$。

当 $n - \frac{(R_0+1)(R_0+2)}{2} < r < R_0 + 1.529$ 时,$\frac{\partial f}{\partial r} < 0$; (9)

当 $R_0 + 1.529 < r \leq n - \frac{R_0(R_0+1)}{2}$ 时,$\frac{\partial f}{\partial r} > 0$。 (10)

B:若 $n - r$ 落在 R_1 区,即 $n - r < \frac{R_0(R_0+1)}{2}$,有 $r > n - \frac{R_0(R_0+1)}{2}$。

$\because$ $n \geq \frac{(R_0+1)(R_0+2)}{2} + 1$,

$\therefore$ $n - \frac{R_0(R_0+1)}{2} \geq R_0 + 2$,$r > R_0 + 2$。

$$f(n, r) = 2F(n - r) + 2^r - 1 = 2\left[\left(n - r - \frac{R_1^2 - R_1 + 2}{2}\right) \cdot 2^{R_1} + 1\right] + 2^r - 1。$$

$\because$ $R_1 < R_0$ 即 $R_1 + 1 \leq R_0$,

$\therefore$ $\frac{\partial f}{\partial r} = -2^{R_1+1} + \ln 2 \cdot 2^r \geq -2^{R_0} + \ln 2 \cdot 2^r > -2^{R_0} + 2^{r-1}$。

当 $r > R_0 + 2$ 时,$\frac{\partial f}{\partial r} > 0$。

$\therefore$ 当 $r > n - \frac{R_0(R_0+1)}{2}$ 时,$\frac{\partial f}{\partial r} > 0$。 (11)

C:若 $n - r$ 落在 $R_0 + 1$ 区,即 $\frac{(R_0+1)(R_0+2)}{2} \leq n - r < \frac{(R_0+2)(R_0+3)}{2}$。

显然，$n \le \frac{(R_0+2)(R_0+3)}{2} - 2 = \frac{R_0^2+5R_0+2}{2}$，　有 $2R_0 \ge n - \frac{R_0^2+R_0+2}{2}$。

$$\therefore\ 2R_0 \cdot 2^{R_0+1} \ge \left(n - \frac{R_0^2+R_0+2}{2}\right)\cdot 2^{R_0+1},\quad 2R_0\cdot 2^{R_0+1}+2 > \left(n-\frac{R_0^2+R_0+2}{2}\right)\cdot 2^{R_0+1}+1,$$

$$\therefore\ 2\cdot(R_0\cdot 2^{R_0+1}+1) > \left(n-\frac{R_0^2+R_0+2}{2}\right)\cdot 2^{R_0+1}+1。\tag{12}$$

$$\therefore\ \overset{②(C)}{\min_r} f(n,r) = \min_r[2F(n-r)+2^r-1] > \min_r[2F(n-r)]$$
$$= 2F\left(\frac{(R_0+1)(R_0+2)}{2}\right) = 2\cdot(R_0\cdot 2^{R_0+1}+1)。\tag{13}$$

由(12)(13)可知，$\overset{②(C)}{\min_r} f(n,r) > \left(n-\frac{R_0^2+R_0+2}{2}\right)\cdot 2^{R_0+1}+1$。

这就是说，$n-r$ 落在 R_0+1 区时，即 $0<r\le n-\frac{(R_0+1)(R_0+2)}{2}$ 时，无论 r 取何值，都有

$$\overset{②(C)}{\min_r} f(n,r) > f(n,R_0+1)。\tag{14}$$

综合(9)，(10)，(11)，(14)和引理，只能在 $r=R_0+1$ 或 $r=R_0+2$ 上达到最小值。

而且，$f(n,R_0+1) = f(n,R_0+2) = \left(n-\frac{R_0^2+R_0+2}{2}\right)\cdot 2^{R_0+1}+1$，所以，不妨取 $r=R_0+1$。

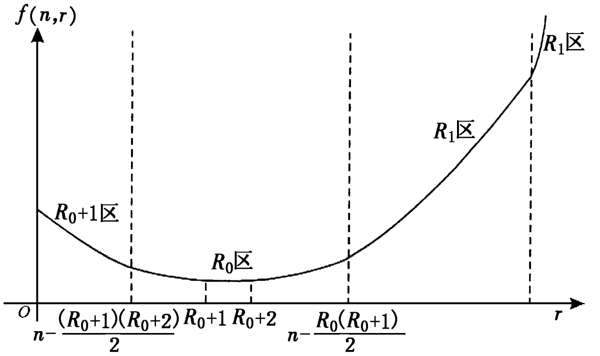

图 3　情况②中 $f(n,r)$ 与 r 的关系示意图

Fig. 3　The relationship between $f(n,r)$ and r in the second case

$$\therefore\ 当\ r=R_0+1\ 时，f(n,r)取最小值。这样，F(n) = \left(n-\frac{R_0^2+R_0+2}{2}\right)\cdot 2^{R_0+1}+1。\tag{15}$$

最后讨论 ③：

当 $n=\frac{(R_0+2)(R_0+3)}{2}-1$ 时，$n-r$ 必然落在 R_0 区或 R_1 区或 R_0+1 区 $(R_1<R_0)$。

A：若 $n-r$ 落在 R_0 区，即 $\frac{R_0(R_0+1)}{2} \le n-r < \frac{(R_0+1)(R_0+2)}{2}$，有 $R_0+1<r\le 2R_0+2$。

$$f(n,r) = 2F(n-r)+2^r-1 = 2\left[\left(n-r-\frac{R_0^2-R_0+2}{2}\right)\cdot 2^{R_0}+1\right]+2^r-1。\tag{16}$$

$\therefore\ \frac{\partial f}{\partial r} = -2^{R_0+1}+\ln 2\cdot 2^r$。令 $\frac{\partial f}{\partial r}=0$，则 $r=R_0+1-\frac{\ln\ln 2}{\ln 2}=R_0+1.529$。

当　$R_0+1<r<R_0+1.529$ 时，$\frac{\partial f}{\partial r}<0$；　(17)

当　$R_0+1.529<r\le 2R_0+2$ 时，$\frac{\partial f}{\partial r}>0$。　(18)

B：若 $n-r$ 落在 R_1 区，即 $n-r<\frac{R_0(R_0+1)}{2}$，有 $r>n-\frac{R_0(R_0+1)}{2}$，$r>2R_0+2$。

$$f(n,r)=2F(n-r)+2^r-1=2\left[\left(n-r-\frac{R_1^2-R_1+2}{2}\right)\cdot 2^{R_1}+1\right]+2^r-1。$$

$\therefore\ \frac{\partial f}{\partial r}=-2^{R_1+1}+\ln 2\cdot 2^r>-2^{R_0+1}+\ln 2\cdot 2^r$。

而 $r>2R_0+2$ 时,有 $r>R_0+2$。

$\therefore\ -2^{R_0+1}+\ln 2\cdot 2^r>0$,有$\frac{\partial f}{\partial r}>0$。

即 $r>2R_0+2$ 时,$\frac{\partial f}{\partial r}>0$。 (19)

C:若 $n-r$ 落在 R_0+1 区,即$\frac{(R_0+1)(R_0+2)}{2}\le n-r<\frac{(R_0+2)(R_0+3)}{2}$,有 $r\le R_0+1$。

$$f(n,r)=2F(n-r)+2^r-1。$$

可以利用②的结论:$F(n-r)=\left(n-r-\frac{R_0^2+R_0+2}{2}\right)\cdot 2^{R_0+1}+1=(2R_0+1-r)\cdot 2^{R_0+1}+1$。

这样,$f(n,r)=(2R_0+1-r)\cdot 2^{R_0+2}+2^r+1$。 (20)

$\therefore\ \frac{\partial f}{\partial r}=-2^{R_0+2}+\ln 2\cdot 2^r$。令$\frac{\partial f}{\partial r}=0$,则 $r=R_0+2-\frac{\ln\ln 2}{\ln 2}=R_0+2.529$。

当 $r\le R_0+1$ 时,有$\frac{\partial f}{\partial r}<0$。 (21)

显然,$r=R_0+1$ 时,$n-r$ 位于 R_0+1 区;$r\to R_0+1+0$ 时,$n-r$ 位于 R_0 区,

利用(16)和(20)可知,$f(n,R_0+1)=(2R_0+1)\cdot 2^{R_0+1}+1=\lim\limits_{r\to R_0+1+0}f(n,r)$。 (22)

(22)保证了 $f(n,r)$在 $r=R_0+1$ 处连续,引理保证了 $f(n,r)$在 $r=2R_0+2$ 处连续。

这样,综合(17),(18),(19)和(21),只能在 $r=R_0+1$ 或 $r=R_0+2$ 上达到最小值。

显然,$r=R_0+2$ 时,$n-r$ 位于 R_0 区,

$$f(n,R_0+2)=2\left[\left(n-r-\frac{R_0^2-R_0+2}{2}\right)\cdot 2^{R_0}+1\right]+2^r-1=(2R_0+1)\cdot 2^{R_0+1}+1。\tag{23}$$

比较(22)和(23)可知,$f(n,R_0+1)=f(n,R_0+2)$,不妨取 $r=R_0+1$。

这样,$F(n)=f(n,R_0+1)=\left(n-\frac{R_0^2+R_0+2}{2}\right)\cdot 2^{R_0+1}+1=(2R_0+1)\cdot 2^{R_0+1}+1$。

$\therefore$ 当 $r=R_0+1$ 时,$f(n,r)$取最小值。这样,

$$F(n)=\left(n-\frac{R_0^2+R_0+2}{2}\right)\cdot 2^{R_0+1}+1。\tag{24}$$

由①、②、③的(8),(15),(24)可知,对于$\frac{(R_0+1)(R_0+2)}{2}\le n<\frac{(R_0+2)(R_0+3)}{2}$内所有的 n,$r=R_0+1$能使 $f(n,r)$取最小值。

这样,$F(n)=\left(n-\frac{R_0^2+R_0+2}{2}\right)\cdot 2^{R_0+1}+1=\left[n-\frac{(R_0+1)^2-(R_0+1)+2}{2}\right]\cdot 2^{R_0+1}+1$。

综上所述,由(Ⅰ)和(Ⅱ)可知,任意给定一个自然数 n,可计算出 $R(n)=$

$\left\lfloor\frac{\sqrt{8n+1}-1}{2}\right\rfloor$，即 $\frac{R(n)[R(n)+1]}{2} \leq n < \frac{[R(n)+1][R(n)+2]}{2}$，都有 $F(n) = \left[n - \frac{R(n)^2 - R(n)+2}{2}\right] \cdot 2^{R(n)} + 1$。

这样,(3) ,(5) ⇒(4) 就得以证明。

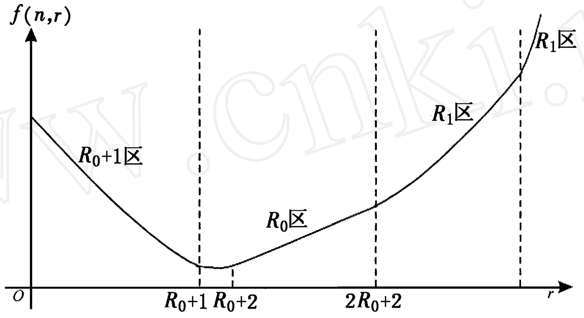

图 4　情况 ③中 f(n, r) 与 r 的关系示意图

Fig. 4　The relationship between $f(n, r)$ and r in the third case

3　结　论

完成 n 个碟子的四柱汉诺塔游戏,需要的最少步数为

$$F(n) = \left[n - \frac{R(n)^2 - R(n)+2}{2}\right] \cdot 2^{R(n)} + 1\text{，其中 } R(n) = \left\lfloor\frac{\sqrt{8n+1}-1}{2}\right\rfloor\text{。}$$

4　讨　论

$F(n)$ 是一个 $\sqrt{2n}\,2^{\sqrt{2n}}$ 量级的函数。而完成 n 个碟子的三柱汉诺塔游戏,需要的最少步数为 $T(n) = 2^n - 1$。因此,当 n 很大时, $F(n) \ll T(n)$。

附录 :引理的证明

证明:

当 $n - r$ 在某一个区(如 R 区, $R \leq R_0$)内变化时,

$$f(n,r) = 2F(n-r) + 2^r - 1 = 2\left[\left(n - r - \frac{R^2 - R+2}{2}\right) \cdot 2^R + 1\right] + 2^r - 1,$$

它当然是关于 r 连续的。

下面只需考虑 $n - r$ 由 R 区跳变至 $R - 1$ 区时, $f(n, r)$ 的行为。

当 $r = n - \frac{R(R+1)}{2}$ 时, $n - r = \frac{R(R+1)}{2}$,有 $n - r$ 落在 R 区。

$$\therefore \quad f(n,r)\Big|_{r=n-\frac{R(R+1)}{2}} = 2F(n-r) + 2^r - 1$$

$$= 2\left[\left(n - r - \frac{R^2 - R+2}{2}\right) \cdot 2^R + 1\right] + 2^r - 1$$

$$= (R-1)\cdot 2^{R+1} + 2^{n-\frac{R(R+1)}{2}} + 1。$$

当 $r \to n - \frac{R(R+1)}{2} + 0$ 时, $n - r \to \frac{R(R+1)}{2} - 0$,有 $n - r$ 落在 $R - 1$ 区。

$$\therefore \lim_{r \to n - \frac{R(R+1)}{2} + 0} f(n, r) = 2F(n-r) + 2^r - 1$$

$$= 2\left[\left(n - r - \frac{(R-1)^2 - (R-1) + 2}{2}\right)\cdot 2^{R-1} + 1\right] + 2^r - 1$$

$$= (R-1)\cdot 2^{R+1} + 2^{n-\frac{R(R+1)}{2}} + 1。$$

$$\therefore f(n, r)\Big|_{r = n - \frac{R(R+1)}{2}} = \lim_{r \to n - \frac{R(R+1)}{2} + 0} f(n, r)。$$

∴ $n - r$ 由 R 区跳变至 $R - 1$ 区时, $f(n, r)$ 也关于 r 连续。

作者的研究工作得到了北京大学计算机科学与技术系刘从义老师的支持和鼓励,在此表示诚挚的谢意。

参 考 文 献

1　谭浩强. C 程序设计. 第 2 版. 北京:清华大学出版社,1999. 161～163

2　Boardman J T, Garrett C, Robson G C. A recursive Algorithm for the Optimal Solution of a Complex Allocation Problem Using a Dynamic Programming Formulation. The Computer Journal, 1986, 29(2):182～186

The Preliminary Probe of 4-Peg Hanoi Tower

YANG Kai　XU Chuan

(Department of Computer Science and Technology, Peking University, Beijing, 100871)

Abstract　In 1941, J. S. Frame gave out an algorithm in *American Mathematical Monthly* to solve the problem of 4-peg Hanoi Tower, but he did not provide the proof for the final formulae. According to that algorithm, this article puts forward a formula to calculate the number of movements necessary for the 4-peg Hanoi Tower problem, and proves it using mathematical induction.

Key words　4-peg Hanoi Tower; zone; the number of remaining disks $R(n)$