Discrete Mathematics 68 (1988) 265–272
North-Holland

THE ADDITION GAME: AN ABSTRACTION OF A COMMUNICATION PROBLEM

Dezsö MIKLÓS*
Mathematical Institute of the Hungarian Academy of Sciences, Budapest, Hungary

Mark J. NEWMAN
Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

Ákos SERESS*
Mathematical Institute of the Hungarian Academy of Sciences Budapest, Hungary

Douglas B. WEST**
University of Illinois, Urbana, IL 61801, U.S.A.

Received 11 August 1986
Revised 25 February 1987

Consider a communication paradigm in which n players each have value 1 initially. When two players with values a and b communicate on a move their values both become $a+b$. For n even, let $f(n)$ and $h(n)$ be the minimum number and maximum number of moves required to reach the position where all n values equal n (which is reachable when $n \neq 6$). Then $h(n) \sim \frac{1}{2} n \lg n$ and $f(n) \geq n - 2 + \lg n$. Furthermore, $f(n) \leq n + \lg n + \lg \lg n + \cdots$ for an infinite sequence of special values of n, and for arbitrary even n, $f(n) \in n + O((\lg n)^3)$. This question is motivated by the problem of gossiping without duplicate transmission.

1. Introduction

We consider several extremal problems associated with a communication paradigm we call the "addition game". Begin with n players each having value 1. On each move, two players communicate, and at the end of the move each has value equal to the sum of their previous values. The object of the game is to reach the position where every player has value n, and to do so with the minimum (or maximum) number of moves. Note that players reach value n in pairs, so the game is infeasible when n is odd. A short search of possibilities shows it is also infeasible when $n = 6$, but we shall see shortly that it is feasible for all other even n.

For even n, let $f(n)$ be the minimum number of moves needed to reach the final position in the addition game, and let $h(n)$ be the maximum realizable number of moves to reach it. Denote $\log_2 n$ by $\lg n$. We will show that $h(n)$ is

* Research partially supported by Hungarian National Foundation for Scientific Research Grant no. 1960.
** Research supported by ONR Grant N00014-85K0570 and by NSF grant DMS 8504322.

asymptotic to $\frac{1}{2}n \lg n$, that $f(n) \geq n - 2 + \lg n$, that $f(n) \leq n + \lg n + \lg\lg n + \cdots$ for an infinite sequence of special values of n, and that in general $f(n) \in n + O((\lg n)^3)$. This can be improved to $n + O((\lg n)^2)$ by a difficult and lengthy proof, which we omit.

We originally investigated the addition game as an abstraction of the problem of gossiping without duplicate transmissions. In that problem the players start with distinguishable units (of "gossip"), and when two communicate they tell each other everything they know. In the final position each player must have received a copy of each unit exactly once. The addition game removes the distinctness restriction. Any feasible sequence of calls for the gossip problem is a feasible sequence of moves for the addition game. Letting $g(n)$ (resp. $k(n)$) be the minimum (resp. maximum) number of calls in a gossiping scheme without duplicate transmissions, we thus have $f(n) \leq g(n) \leq k(n) \leq h(n)$.

The feasibility question for the gossip problem was non-trivial. Like the addition game, it is infeasible when n is odd or equals 6, and not much effort is needed to show it is also infeasible when $n \in \{10, 14, 18\}$. Its feasibility for all multiples of 4 was noted originally in [1], and for the remaining even values of n in [2]. This shows that the addition game is feasible for even n when $n \notin \{6, 10, 14, 18\}$, and schemes are easily exhibited for $n \in \{10, 14, 18\}$.

The inductive gossiping scheme given in [1] (see [5]) uses $\frac{1}{2}n \lg n + O(n)$ calls, and the scheme in [2] can also be modified easily to use this many calls. As mentioned above, this gives a lower bound for $h(n)$, which we show optimal. For the minimization problem, [5] showed $g(n) \leq \frac{9}{4}n - 6$ when $n \equiv 0 \pmod 4$, which [3] proved optimal. [3] also showed $g(n) = \frac{9}{4}n - 4 \pm \frac{1}{2}$ when $n \equiv 2 \pmod 4$. Hence $f(n)$ is linear, but as noted above the correct constant is 1. ([4] generalized the linear scheme to k-party gossiping without duplication.)

The addition game and the gossip problem as described above are special cases of a more general addition game. Consider parameters n, m, d, with n divisible by d. Start each of the n players with a canonical basis vector in d dimensions, with n/d players getting each vector. Moves consist of replacing the vectors held by each of two players by the sum of those vectors. In the desired final position, each player has the vector $m\mathbf{1}_d$. The original addition game is the case $d = 1$ and $m = n$, while the problem of gossiping without duplication is the case $d = n$ and $m = 1$. Again we say the game is *feasible* if the final position is reachable, and that any sequence of moves doing so *solves* the game. We call this the (n, m, d) addition game, or simply the (n, m) addition game if $d = 1$.

As usual, feasibility requires n even. If n/d and d are both even (and sufficiently large), the $(n, n/d, d)$ addition game is solved by the following scheme. First apply the addition game to the n/d players having each basis vector. Re-group the players into n/d sets of size d using one from each group, and use the gossip scheme n/d times. The total number of calls used is $d \cdot f(n/d) + (n/d) \cdot g(d)$.

The feasibility question is even more interesting when $d = 1$ and $m \neq n$. For example, when $n=4$ and $m \neq 7, 11$, it appears that the game always has solutions. The truth of this conjecture would yield $f(n) \in n + \mathrm{O}((\lg n)^2)$, as discussed in Section 4, which partially explains the omission of the lengthy proof of $f(n) \in n + \mathrm{O}((\lg n)^2)$ mentioned above. The feasibility question for $d=1$ and general m, n, i.e., moving n players from value 1 to value m, was first posed by Richards.

In Section 2 we prove the constraints claimed for $f(n)$ and $h(n)$, and in Sections 3 and 4 we provide near-optimal constructions for $f(n)$.

2. Constraints on f and h

First, the maximization problem.

Theorem 1. *$h(n) \leqslant \frac{1}{2}n \lg n$, which is asymptotically optimal. More generally, for the (n, m, d) addition game with $d=1$, any feasible sequence takes at most $\frac{1}{2}n \lg m$ moves.*

Proof. As noted in the introduction, schemes taking asymptotically this many moves already exist, and we need only show we cannot use more. The proof we present here, shorter than the pigeonhole argument originally found by the authors, is due to Ziegler for the case $m=n$.

Let the product of the players' values after the tth move be α_t. Suppose the values of two players are a and b before they move together. By the arithmetic-geometric mean inequality, $(a+b)(a+b) = 2ab + a^2 + b^2 \geqslant 4ab$. Therefore, regardless of the choice of move, $\alpha_t \geqslant 4\alpha_{t-1}$. Since the initial and final values of α are 1 and m^n, the number of moves is at most $\log_4 m^n = \frac{1}{2}n \lg m$. □

Now consider the least number of moves to reach the final position.

Theorem 2. $f(n) \geqslant n - 2 + \lg n$.

Proof. Let α_t be the sum of the players' values after move t. We have $\alpha_0 = n$, and the final value of α is n^2. On any move, the gain in α equals the value each participant in the move has at the end of the move. Hence any move that finishes two players has a gain of n, and no move has gain more than n. Also, the greatest gain is always achieved by moving the two farthest players. After $\lfloor \lg \frac{1}{2}n \rfloor$ moves, the total value is at most $n - 2 + 2(2^{\lfloor \lg \frac{1}{2}n \rfloor}) \leqslant 2n - 2$. Hence there must be at least $(n^2 - 2n + 2)/n$ moves remaining after that. □

This bound is surprisingly close to optimal and surprisingly hard to improve. The closeness to optimality is surprising because the argument is so easy and because consistently moving to obtain the greatest possible immediate gain leads to a sequence of $\frac{1}{2}n \lg n$ moves. On the other hand, since the scheme suggested by the argument clearly cannot be attained, it should be possible to improve this bound, perhaps to show that the construction of the next section is optimal.

3. Construction for special values of n

We consider a special sequence of values, recursively defined.

Theorem 3. *Define a_k by $a_1 = 4$, $a_k = 2^{a_{k-1}} + a_{k-1}$ for $k \geq 2$. Then $f(a_k) \leq a_k + a_{k-1} + \cdots + a_1$.* I.e., *$f(n) \leq n + \lg n + \lg\lg n + \cdots$ for $n = a_k$.*

Proof. Note that $a_{k-1} = \lg(a_k - a_{k-1}) \leq \lg a_k$, and that for $n = 4$ it is easy to complete the game in 4 moves. Let $M = a_k + \cdots + a_1$. For $k > 1$, we specify an explicit sequence of M moves for $n = a_k$. To describe the state of the game after any move, we use the format $({}^{v_1}_{m_1}| \cdots |{}^{v_k}_{m_k}; t)$, where $v_1, \ldots, v_k$ is a list of values (also called "positions") currently held by players, $m_1, \ldots, m_k$ are the corresponding multiplicities of players holding those values or positions, and t is the total number of moves made so far. The initial state is $({}^{1}_{n}; 0)$, and the final state in a minimal sequence is $({}^{n}_{n}; f(n))$. It is natural to combine several such descriptions in a single table; each line of the table give a population distribution over the values indicated by the columns.

The general strategy is to develop two players rapidly by moving them with each other to a large value, then moving them with unstarted players (players at position 1, also called *new* players) till reaching the end. The leaders move one unit on each wave of these latter moves, but the number of leaders doubles on each wave. If the initial leaders start by doubling their position together k times, this would solve the problem if $n - 2^k = \lg n - 1$, and approximately achieve the bound of Theorem 2. Unfortunately, no choice of n, k satisfies this equation. Nevertheless, for $n = a_k$ as defined above, we will combine several phases of this type. In each phase, we leave two players along the way that we gather up in the next phase, to help the next phase to work.

Phase 1. Make a_{k-1} moves on two new players, reaching the position indicated by line 1 of Table 1. Matching each of these with new players twice yields the next two lines; note that we are leaving two players behind. Phase 1 ends with $a_{k-1} - 2$ steps of matching all the leading players with new players. The number of leading players doubles each time till value n is reached. The game position is now that of the last line of the table. With $n = a_k = 2^{a_{k-1}} + a_{k-1}$, this equals $({}^{1}_{a_{k-1}-2}|{}^{n-a_{k-1}+1}_{2}|{}^{n}_{n-a_{k-1}}; a_k)$.

Phase 2. This is actually a succession of phases like Phase 1. We claim we can reach $({}^{1}_{a_{k-j}-2}|{}^{n-a_{k-j}+1}_{2}|{}^{n}_{n-a_{k-j}}; \sum_{i=0}^{j-1} a_{k-i})$, for $j = 1, \ldots, k-2$. Phase 1 provides

Table 1. Phase 1 of the scheme for special n

1	$2^{a_{k-1}}$	$2^{a_{k-1}}+1 = n-a_{k-1}+1$	$2^{a_{k-1}}+2$	$2^{a_{k-1}}+a_{k-1} = n$	Total moves
a_k-2	2	0	0	0	a_{k-1}
a_k-4	0	4	0	0	$a_{k-1}+2$
a_k-6	0	2	4	0	$a_{k-1}+4$
$a_k-2^{a_{k-1}}-2$	0	2	0	$2^{a_{k-1}}$	$a_{k-1}+2^{a_{k-1}}$

Table 2. Phase 2 of the scheme for special n

1	$2^{a_{k-j}}$	$n-a_{k-j+1}+1$	$n-a_{k-j}+1$	$n-a_{k-j}+2$	n	Total moves
$a_{k-j+1}-2$	0	2	0	0	$n-a_{k-j+1}$	M'
$a_{k-j+1}-4$	2	2	0	0	$n-a_{k-j+1}$	$M'+a_{k-j}$
$a_{k-j+1}-4$	0	0	4	0	$n-a_{k-j+1}$	$M'+a_{k-j}+2$
$a_{k-j+1}-6$	0	0	2	4	$n-a_{k-j+1}$	$M'+a_{k-j}+4$
$a_{k-j+1}-2^{a_{k-j}}-2$	0	0	2	0	$n-a_{k-j+1}+2^{a_{k-j}}$	$M'+a_{k-j}+2^{a_{k-j}}$

the basis for induction on j. For the induction step, consider Table 2. The induction hypothesis guarantees reaching the position of line 1, where $M' = \sum_{i=0}^{j-2} a_{k-i}$, the previous number of moves. Make a_{k-j} moves on two new players together and match them with the two leftover players from the previous phase to get the positions of lines 2 and 3. Match two of the resulting players with new players to reach line 4. Finally, as in Phase 1, complete this with $a_{k-j}-2$ steps of matching all leading unfinished players with new players, reaching line 5. This position is the same as $({}^{1}_{a_{k-j}-2}|{}^{n-a_{k-j}+1}_{2}|{}^{n}_{n-a_{k-j}}; \sum_{i=0}^{j-1} a_{k-i})$, as desired.

Phase 3. At the end of Phase 2, we have $j=k-2$. With $a_2=20$, the position is $({}^{1}_{18}|{}^{n-19}_{2}|{}^{n}_{n-20}; M-24)$; note that as j increased the position of the two leftover players moved cloer to n. Next we make $4=a_1$ moves with two new players and then combine them with the previously leftover players, as in early phases, yielding $({}^{1}_{16}|{}^{n-3}_{4}|{}^{n}_{n-20}; M-18)$. This phase is different because we must leave leftover players in two positions. The remaining moves can be read from the

Table 3. Phase 3 of the scheme for special n

1	2	$n-3$	$n-2$	$n-1$	n	Total moves
14	0	2	4	0	$n-20$	$M-16$
12	0	2	2	4	$n-20$	$M-14$
8	0	2	2	0	$n-12$	$M-10$
4	4	2	2	0	$n-12$	$M-8$
4	2	2	0	0	$n-8$	$M-6$
4	0	0	0	4	$n-8$	$M-4$
0	0	0	0	0	n	M

positions indicated by successive lines in Table 3, completing the proof that $f(n) \leqslant M$ when $n = a_k$. □

4. The general construction

As mentioned in the introduction, if we could solve the general addition game for $(n, m, d) = (4, m, 1)$ with $m \neq 7, 11$, we could prove $f(n) \in n + O((\lg n)^2)$. We discuss this here to motivate the general construction.

Suppose n is divisible by 4 and $n = \sum 2^{c_i}$, with all $c_i \geqslant 2$. Suppose also that $n - c_i + 2 \neq 7, 11$ for all i, which holds for $n \geqslant 14$. For each i, move 4 players by themselves to $n - c_i + 2$, and follow this with $c_i - 2$ steps of matching all the leaders to new players, as in the proof of Theorem 3. This moves 2^{c_i} players to value n. By Theorem 1, the moves in the initial phase total at most $2\sum \lg(n - c_i + 2)$. The latter phase has a move for each unstarted player after the first phase, i.e., $n - 4k$ moves, where n was expressed as a sum of k powers. We get $f(n) \leqslant n + k(2\lg n - 4) \leqslant n + 2(\lg n)^2 - 8\lg n + 8$, since each such n is expressible with at most $\lg n - 2$ powers of 2.

On the other hand, if $n \equiv 2 \pmod 4$, we must be more careful. Values through 18 can be considered explicitly. Beyond that, we express $n - r$ as a sum of powers of 2 as above, where r is the smallest odd multiple of 2 satisfying $r \geqslant \max\{22, 2\lg n\}$. We move $n - r$ players to n as above. For the remaining players, express n as the sum of exactly $\frac{1}{2}r$ powers of 2 (as detailed later), and move the r players in pairs so the sum of all their values will be exactly n in this way. By the result of [3] mentioned earlier, there is a sequence of $\frac{9}{4}r - 3.5$ moves on these r players that puts them all at position n. Again we are using at most $\lg n$ powers of 2 in this added step, but since we place only pairs rather than quadruples, we get $(\lg n)^2$ additional steps, not $2(\lg n)^2$, so $f(n) \leqslant n + 3(\lg n)^2 + O(\lg n)$.

The ideas used above are all present in our general upper bound proof. In lieu of the missing lemma about moving 4 players to m, we begin by proving a weaker lemma of a similar nature, which itself requires a numerical lemma.

Lemma 1. *If p is an integer with $2^l \leqslant p \leqslant 2^{2^l}$, then p can be expressed as the sum of 2^l non-negative powers of 2. Furthermore, the c_i in $p = \sum_{i=1}^{2^l} 2^{c_i}$, can be chosen so that $\sum c_i \leqslant \binom{1+\lfloor \lg p \rfloor}{2}$.*

Proof. Since $p \leqslant 2^{2^l}$, the binary expansion of p has at most 2^l terms; i.e., there is a unique partition of p into distinct powers of 2 with largest exponent $\lfloor \lg p \rfloor$. Starting with this partition, increase the number of terms to 2^l by successively splitting positive powers, replacing 2^c by two copies of 2^{c-1}. Consider such a partition minimizing $\sum c_i$; note that still $\max c_i \leqslant \lfloor \lg p \rfloor$. Only the smallest positive power can be repeated, and if repeated it appears at most twice, since $2^j + 2^j + 2^k$ with $j \geqslant k > 0$ can be replaced by $2^{j+1} + 2^{k-1} + 2^{k-1}$. If no positive power is repeated, the sum is at most

$$\sum_{i=1}^{\lfloor \lg p \rfloor} i = \binom{1 + \lfloor \lg p \rfloor}{2}.$$

If a positive power 2^j is repeated, consider the expression obtained by replacing one copy of 2^j by $2 + \sum_{i=1}^{j-1} 2^i$. Now 2 is the only repeated positive power and $\sum c_i$ has not decreased, so $\sum c_i$ is bounded by $\binom{1+\lfloor \lg p \rfloor}{2} + 1$. However, equality requires $p \geq 2^{1+\lfloor \lg p \rfloor}$, which is impossible. □

Lemma 2. *Suppose a and m satisfy* $k \leq a \leq m + k - 2^k$, *where* $k = \lceil 1 + \lg \lg m \rceil + \varepsilon$ *with* ε *chosen as* 0 *or* 1 *so that* $m - (a - k)$ *is even. Then we can solve the* $(2^a, m)$ *addition game, using at most* $2^a + \frac{1}{2}(\lg m)^2 + \frac{19}{2} \lg m$ *moves.*

Proof. First write $m - (a - k)$ as the sum of exactly 2^{k-1} positive powers of 2. Since $m \leq 2^{2^{k-1}}$, $2^k \leq m - (a - k)$, and $m - (a - k)$ is even, we can apply Lemma 1 to $\frac{1}{2}(m - (a - k))$ with $l = k - 1$ to write $m - (a - k) = \sum_{i=1}^{2^{k-1}} 2^{c_i}$ with c_i positive and $\sum (c_i - 1) \leq \binom{\lfloor \lg m \rfloor}{2}$. Now, for each i we move two players to position $2^{c_i - 1}$ with $c_i - 1$ moves. This yields a set of 2^k players with values summing to $m - (a - k)$. This phase uses at most $\binom{\lfloor \lg m \rfloor}{2}$ moves.

Next we use a sequence of moves on these players that corresponds to a minimal sequence solving the gossip problem without duplicate transmission for 2^k gossips. The value held by each player in this set reaches each other player exactly once, so at the end of these $\frac{9}{4}2^k - 6$ moves we have 2^k players at position $m - (a - k)$.

Finally, $a - k$ rounds of matching all the leaders with new players puts all 2^a players at position m. This step uses $2^a - 2^k$ moves, since each move here starts a new player. Since $2^{k-1} \leq 4 \lg m$, the total number of moves over all three phases is less than

$$2^a + \binom{\lfloor \lg m \rfloor}{2} + \tfrac{5}{4}2^k \leq 2^a + \tfrac{1}{2}(\lg m)^2 + \tfrac{19}{2} \lg m. \quad \square$$

Now, we can attack $f(n)$ itself. The idea is to move players to n in bunches of size 2^{a_i} by using Lemma 2, and then handle the leftover players separately.

Theorem 4. $f(n) \leq n + (\lg n)^3 + 22(\lg n)^2 + \frac{27}{2} \lg n + O(1)$.

Proof. Besides the bunches of size 2^a, we will have r leftover players. More precisely, we want to write n in the form $n = r + \sum_{i=1}^{t} 2^{a_i}$ with restrictions that will make the steps in the scheme work. The restrictions are $a_i \geq k$ and $6 \lg n > r \geq \max\{22, 2 \lg n\}$, where $k = \lceil 1 + \lg \lg n \rceil$. The reason for the restriction $a_i \geq k$ is to employ Lemma 2 to solve the $(2^{a_i}, n)$ addition game. We do not need the ε of Lemma 2 because n is even. I.e., if $a_i = k$, then $n - (a_i - k)$ is even, and if $a_i > k$, then $a_i \geq k + 1$. Note that $2 \lg n \leq 2^k < 4 \lg n$.

To choose the a_i's, begin with the binary expression for n, in which 2^c is the smallest term with $c \geq k$. If the sum of the smaller terms is less than $2\lg n$, we must include more, but putting 2^c into r may be too much. The solution is to replace 2^c by $2^{c-1}+\cdots+2^{k+1}+2^k+2^k$. Now the large terms, including one copy of 2^k, are the 2^{a_i}'s, and the small terms, including the other copy of 2^k, sum to r. Since $2^k \geq 2\lg n$, we have $r \geq 2\lg n$. For large enough n, this also guarantees $r \geq 22$, which is the reason for O(1) in the statement of the theorem.

For the upper bound on r, we have $r < 2^k < 4\lg n$ if the replacement was not needed, because r is then the sum of distinct powers of 2 less than 2^k. On the other hand, if the replacement was needed, then $r < 2\lg n + 2^k < 6\lg n$. We also have the upper bound $t < 2\lg n$, since the binary expansion starts with at most $\lg n$ terms and breaking down 2^c introduces less than $\lg n$ additional terms.

Having obtained the desired bounds, we can present and analyze the scheme. First, use Lemma 2 for each a_i to move set of 2^{a_i} players to position n; this is the reason for $a_i \geq k$. For the remaining r players, express n as the sum of $r/2$ positive powers of 2; the fact that $r \geq 2\lg n$ ensures that this is possible. Given the expression $n = \sum_{i=1}^{r/2} 2^{c_i}$, move a pair to each 2^{c_i-1} ($r/2$ instances of at most $\lg n$ moves). Finally, apply $g(r)$ calls to gossip without duplication on those r players. Altogether, the number of moves is at most

$$\sum 2^{a_i} + t \cdot \tfrac{1}{2} \cdot (\lg n)^2 + \tfrac{1}{2}(19t + r)\lg n + \tfrac{9}{4}r.$$

The upper bounds on t and r complete the proof. □

References

[1] H.W. Lenstra, letter to R.L. Graham.
[2] Á. Seress, Gossiping old ladies, Discrete Math. 46 (1983) 75–81.
[3] Á. Seress, Quick gossiping without duplicate transmissions, Graphs and Combin. 2 (1986) 265–283.
[4] Á. Seress, Quick gossiping by conference calls, SIAM J. Alg. Disc. Meth., submitted.
[5] D.B. West, Gossiping without duplicate transmissions, SIAM J. Alg. Disc. Meth. 3 (1982) 418–419.