The Communication Complexity of Approximate Set Packing and Covering

Noam Nisan*

Institute of Computer Science,
The Hebrew University of Jerusalem
Israel
noam@cs.huji.ac.il

Abstract. We consider a setting where k players are each holding some collection of subsets of $\{1..n\}$. We consider the communication complexity of approximately solving two problems: The *cover number*: the minimal number of sets (in the union of their collections) whose union is $\{1...n\}$ and the *packing number*: the maximum number of sets (in the union of their collections) that are pair-wise disjoint.
We prove that while computing a $(\ln n)$-approximation for the cover number and an $min(k, O(\sqrt{n}))$-approximation for the packing number can be done with polynomial (in n) amount of communication, getting a $(1/2 - \epsilon) \log n$ approximation for the cover number or a better than $min(k, n^{1/2-\epsilon})$-approximation for the packing number requires exponential communication complexity.

1 Introduction

The problems studied here were motivated by communication bottlenecks in combinatorial auctions. In a combinatorial auction n items are sold concurrently, and k bidders are bidding for them. The combinatorial nature of the auction comes from the fact that players value *bundles* of goods as opposed to individual goods. I.e. each player holds a valuation function v_i that assigns a non-negative value for each set of items $S \subset \{1...n\}$. The valuation function v is usually assumed to be monotone (items can be freely disposed). The problem that combinatorial auctions aim to solve is to find the optimal allocation between the different bidders, i.e. a partition $(S_1...S_k)$ of the items that maximizes $\sum_i v_i(S_i)$.

Even when each player is only interested in a single set of items (i.e. v_i is totally specified by a single set S_i and a price p_i), the allocation problem is computationally intractable even to approximate within a factor of $n^{1/2-\epsilon}$ (a simple reduction from approximate Clique [4]). However, it seems that in practice problems with hundreds of items can be solved [11, 10], and those with thousands of items rather well approximated [13]. This paper explores another concern, that of actually transferring the required information about v_i – information that

* Supported by a grant from the Israeli Academy of Sciences.

is exponentially larger than n. When each v_i takes only $0, 1$ values, each v_i is completely specified by a collection of "desired sets". An allocation in this case is exactly a set packing and we need to analyze the communication requirements of finding or approximating the optimum. (It turns out that the original, general, "weighted case" has essentially identical communication requirements.) From now on this paper deals with this non-weighted set-packing problem. More details regarding the applications to combinatorial auctions themselves may be obtained from a companion paper [8].

The set packing communication problem is as follows: k players are each holding a collection A_i of subsets of $\{1...n\}$, and the players are looking for the largest packing (i.e. collection of pair-wise disjoint sets) in the union of their collections. One may naturally also consider the related problem of set covering: each player holds a collection A_i of subsets of $\{1...n\}$, and the players are looking for the smallest covering (i.e. collection of sets whose union is $\{1...n\}$) in the union of their collections[1].

It is quite easy to prove an exponential lower bound on the communication needed to compute either the cover number or the packing number. What we are concerned here is approximating these numbers. Somewhat surprisingly, there has been very little work to date on approximation in the communication complexity model. We feel that studying approximation in a communication complexity setting is interesting for at least two reasons. First, it abstracts the communication requirements of approximation problems in exactly the same way as the exact version does for exact problems. Moreover, in many applications of communication complexity the input size is exponentially larger than the permissible amount of communication. In such cases one can rarely hope for exact solutions and thus approximation is the goal. Second, as the communication complexity model is mostly used for its lower bounds, relaxing the problem strengthens the lower bound, and hence may find further applications. E.g. the natural "number in the hand" model of k-player communication (where each player knows a single piece of the input) is only rarely considered (rather than the more powerful "number on the forehead" model) since increasing the number of players reduces the power of the model and hence is less interesting for lower bounds. However, when approximation enters the picture this may change since approximation factors may increase with the number of players. E.g. for the set packing problem, getting a k-approximation with k players is strictly harder than getting a $(k + 1)$-approximation with $k + 1$ players. Hence lower bounds for large numbers of players in the "number in the hand" model are of interest.

We are able to show tight bounds on the approximation factors obtainable with polynomial communication:

[1] The weighted version of the set covering problem also has a natural interpretation as "combinatorial procurement auctions" – the n items are now procured in an auction from bidders who are each bidding to supply subsets S for the price $v_i(S)$, and the auctioneers' goal is to minimize his total payment.

Theorem: The set cover problem may be approximated in polynomial communication to within a factor of $\ln n$. Any approximation to within a factor of $(1/2-\epsilon)\log n$ requires exponential communication. The lower bound holds also for randomized or nondeterministic protocols.

Theorem: The set packing problem may be approximated in polynomial communication to within a factor of $min(k, \sqrt{n})$. Any approximation to within a factor of better than $min(k, n^{1/2-\epsilon})$ requires exponential communication. The lower bound holds also for randomized or nondeterministic protocols.

Both proofs rely on a single simple technique: probabilistic construction of a reduction. One is immediately struck with the similarity of these gaps to those in the computational setting. One may compute a $\ln n$-approximation for covering and a $O(\sqrt{n})$-approximation for packing in polynomial time, but it is NP-hard to approximate covering to within $(1/2-\epsilon)\log n$ [7] and packing to within $n^{1/2-\epsilon}$ [4]. It seems that this similarity is rooted in the gap between the fractional solution and the integral one: The gap between the fractional cover number and the integral cover number is $\Theta(\log n)$ and the gap between the fractional packing number and the integral one is $\Theta(\sqrt{n})$ [6, 1]. Moreover, for the case of the covering problem, the underlying combinatorial construction used here is indeed similar to that used in the NP-completeness result of [7].

It is interesting to observe that the fractional cover number and fractional packing number may be both computed exactly in a polynomial amount of communication. The linear programs describing the fractional cover number and the fractional packing number, each have a variable for each subset of the items, and a constraint for each item. The dual LP may thus be solved by a separation-based LP algorithm in polynomial time. Since each constraint of the dual LP corresponds to a subset that is held by one of the players, a separating hyperplane can be found by a single player, once the variable values are given. Thus each step of the algorithm may be implemented by the players using a polynomial amount of communication by broadcasting the variable values to all players, and one player responding with a separating hyperplane.

2 Set Packing

2.1 Upper Bound

We are in a situation that each of k players holds a collection $A_i \subseteq P(\{1...n\})$, and they want to approximately find a maximum packing.

Getting an approximation factor of k is easy by just picking the single player with largest packing in his collection. If $k > \sqrt{n}$ we can do better by using the following simple greedy protocol: at each stage each player announces the smallest set $S_i \in A_i$ that is disjoint from all previously chose sets (requiring n bits of communication from each of k players). The smallest such set is chosen to be in the packing. This is repeated until no more disjoint sets exist – after at most n rounds. It is easy to verify that this packing is at most a factor of

$\sqrt{n}$ smaller than the optimal packing. The total communication is bounded by $O(kn^2)$.

Note: The weighted case (where each player has a valuation function $v_i : P(\{1...n\}) \to R^+$ and the aim is to maximize $\sum_i v_i(S_i)$ over all packings $S_1...S_k$) can also be approximated this way to within the same factor, by choosing, at each stage the set S_i that maximizes $v_i(S_i)/\sqrt{|S_i|}$ [5].

2.2 Approximate Disjointness Lower Bound

Our lower bound for the set packing problem will use a reduction from the following "approximate-disjoitness" problem which is of independent interest, and which was studied in [2]:

The Approximate Disjointness Problem

k players are each holding an N-bit string. The string for player i specifies a subset $A_i \subseteq \{1...N\}$. They are required to distinguish between the following two extreme cases:

- Negative: $\cap_{i=1}^{k} A_i \neq \emptyset$
- Positive: for every $i \neq j$, $A_i \cap A_j = \emptyset$

Theorem 1. *[2] The approximate disjointness problem requires communication complexity of $\Omega(N/k^4)$. This lower bound applies also to randomized protocols with bounded 2-sided error, and also to nondeterministic protocols.*

The proof in [2] generalizes the probabilistic lower bound for disjointness in the two-player case of [9]. They only point out the randomized lower bound, but their proof gives a nondeterministic bound as well. We provide here a simple self-contained proof due to Jaikumar Radhakrishnan and Venkatesh Srinivasan of a somewhat stronger bound for the deterministic (and in fact also the nondeterministic) case.

Theorem 2. *(Radhakrishnan and Srinivasan) The approximate disjointness problem requires communication complexity of $\Omega(N/k)$.*

Proof. First note that any c-bit protocol for the approximate disjointness problem partitions the space of inputs into at most 2^c "boxes", where a box is a Cartesian product $S_1 \times S_2 \times ... \times S_k$, where for each i, $S_i \subseteq \{0,1\}^N$. Each box must be labeled with an answer, and thus the boxes must be "monochromatic" in the following sense: No box can contain both a positive instance and a negative instance. (There is no restriction on instances that are neither positive nor negative.)

We will show that there are exactly $(k+1)^N$ positive instances, but that any box that does not contain a negative instance can contain at most k^N positive instances. It follows that there must be at least $(1+1/k)^N$ boxes to cover all the positive instances and thus the number of bits communicated is at least the logarithm of this number, and the theorem follows.

To count the number of positive instances, note that any partition of the N items between the k players and "unallocated" provides a positive instance. The number of such partitions is exactly $(k+1)^N$.

Now consider a box $S_1 \times S_2 \times ... \times S_k$ that does not contain any negative instance. Note that for each item $i \in \{1...N\}$ there must exist a player $j = j_i$ such that for all $A \in S_j$, $i \notin A$. (Since otherwise we choose for all j, $A_j \in S_j$ with $i \in A_j$ and we have that $\cap_j A_j \neq \emptyset$ – a negative instance in the box.) We can obtain an upper bound to the number of positive instances in this box by noting that each such positive instance corresponds to a partition of the N items among the k players and "unallocated", but now with the added restriction that each item i can not be in the partition of j_i. Thus each item has only k possible locations for it and the number of such partitions is at most k^N.

2.3 Set Packing Lower Bound

We have k players, each holding a collection $A_i \subseteq P(\{1...n\})$. We will prove a lower bound on the communication complexity needed in order to distinguish between the case where the packing number is 1 and the case that it is k. I.e. to distinguish the case where there exists k disjoint sets $S_i \in A_i$ (a k-packing), and the case where any two sets $S_i \in A_i$ and $S_j \in A_j$ intersect (packing number is 1). We will reduce this problem from the approximate-intersection problem on vectors of size $t = e^{n/(2k^2)}/k$. The reduction uses a set of partitions $F = \{P^s | s = 1...t\}$, where each P^s is a partition $(P_1^s, ..., P_k^s)$ of $\{1...n\}$ into k subsets. This set of partitions will have the following property:

Definition 1. *A set of partitions $F = \{P^s | s = 1...t\}$ has the pair wise-intersection property if for every choice of $1 \leq i \neq j \leq k$ and every $1 \leq s_i \neq s_j \leq t$ we have that $P_i^{s_i} \cap P_j^{s_j} \neq \emptyset$. I.e. that any 2 parts from different partitions intersect.*

Lemma 1. *There exists a set F of partitions with the pair wise-intersection property of size $|F| = t = e^{n/(2k^2)}/k$.*

Proof. We will use the probabilistic method where each partition in the set will be chosen independently at random, and for each partition, each element is placed independently at random in one of the parts of the partition. Now fix $1 \leq i \neq j \leq k$ and two indices of partitions $1 \leq s_i \neq s_j \leq t$. The probability that they do not intersect can be calculated as:

$$Pr[P_i^{s_i} \cap P_j^{s_j} = \emptyset] = (1 - 1/k^2)^n \leq e^{-n/k^2}$$

Since there are at most k^2t^2 such choices of indices, we get that as long as $k^2t^2 < e^{n/k^2}$, the required set of partitions exists.

We can now specify the reduction of approximate disjointness on vectors of size t to the approximate set packing problem. Player i who gets as input the set $B_i \subset \{1...t\}$ will construct the collection $A_i = \{P_i^s | s \in B_i\}$. Now, if there exists $s \in \cap_i B_i$ then a k-packing exists: $P_1^s \in A_1 ... P_k^s \in A_k$. If, one the other hand, for all $i \neq j$, $B_i \cap B_j = \emptyset$, then for any two sets $P_i^{s_i} \in A_i$ and $P_j^{s_j} \in A_j$ we have $s_i \neq s_j$ and thus $P_i^{s_i} \cap P_j^{s_j} \neq \emptyset$.

From the $\Omega(N/k^4)$ lower bound for the approximate disjointness problem, it follows that the communication complexity of the approximate set packing problem is at least $\Omega((e^{n/(2k^2)}/k)/k^4) = \Omega(e^{n/(2k^2)-5\log k})$. We thus get:

Theorem 3. *Any k-player protocol for approximating the packing number to within a factor less than k requires $e^{n/(2k^2)-5\log k}$ communication. The lower bound holds also for randomized or nondeterministic protocols.*

In particular, as long as $k < n^{1/2-\epsilon}$, the communication complexity is exponential.

3 Set Cover

3.1 Upper Bound

We are in a situation that each of k players holds a collection $A_i \subseteq P(\{1...n\})$, and they want to approximately find a minimal cover. We use the well known greedy method: At each stage each player i announces a set $S_i \in A_i$ that covers the largest number of remaining elements (this takes n bits for each of the k players for each round), and the largest set among all players is chosen. The elements that are contained in this set are removed, and a new stage starts, now considering only the remaining elements. The protocol ends when all items have been covered – after at most n stages. Thus the total amount of communication is $O(kn^2)$. It is well known [6] that this procedure produces a $(\ln n + 1)$-approximation to the minimum cover.

Note: The weighted case (where each player has a valuation function $v_i : P(\{1...n\}) \to R^+$, where for all $S, T \subset \{1...n\}$ we have that $v(S \cup T) \leq v(S) + v(T)$, and the aim is to *minimize* $\sum_i v_i(S_i)$ over all covers $\{S_1...S_k\}$) can also be approximated this way to within the same factor, by choosing, at each stage the set S_i that minimizes $v_i(S_i)/|S_i|$ [3].

3.2 Lower Bound

Let Alice hold a collection $A \subseteq P(\{1...n\})$, and Bob hold a collection $B \subseteq P(\{1...n\})$. We will show that it requires exponential communication to separate between the case that for some $S \in A$ we also have $S^c \in B$ (in which case the cover number is 2), and the case where no r sets cover $\{1...n\}$ (for some r to be chosen).

Let us restrict both A to include only sets from a collection $C \subseteq P(\{1...n\})$ to be constructed below, and B to include only sets whose complement is in C. The class C will have the following property:

Definition 2. *A class C has the r-covering property if every collection of at most r sets who either are from C or whose complement is from C, does not cover all of $\{1...n\}$ as long as a set and its complement are not taken together. I.e. for all pair-wise different $S_1...S_r \in C$, and all $0 \leq j \leq r$ we have that $S_1 \cup ... \cup S_j \cup S^c_{j+1} \cup ... \cup S^c_r \neq \{1...n\}$.*

This property was also used in [7], but we use different parameters.

Lemma 2. *For any given $r \leq \log n - O(\log \log n)$, a set C of size t satisfying the r-covering property exists with $t = e^{n/(r2^r)}$.*

Proof. We will use the probabilistic method choosing sets $S_1...S_t$ to put in C at random, where each element $1 \leq i \leq n$ appears in each set S_j with probability $1/2$. For a random collection of r sets, and a single element i, we have that i is in their union with probability $1 - 2^{-r}$. The probability that the union of a given collection of r sets is all of $\{1...n\}$ is thus $(1 - 2^{-r})^n \leq e^{-n/2^r}$. The probability that some collection of r sets from among the t sets in C and the t set whose complement is in C is thus bounded from above by $\binom{2t}{r}e^{-n/2^r} \leq t^r e^{-n/2^r}$. Thus as long as $t < |C| = e^{n/r2^r}$ such a set exists.

At this point we can reduce the r-approximate covering problem from the disjointness problem on a vector of size t. Alice who is given a vector $x \in \{0,1\}^t$ will construct the collection $A = \{S_i | x_i = 1\}$. Bob who is given $y \in \{0,1\}^t$ will construct the collection $B = \{S^c_i | y_i = 1\}$. Now notice that if for some i, $x_i = 1 = y_i$ then the cover number is exactly 2 (given by S_i and S^c_i). Otherwise, the cover number will be at least $r+1$ since C has the r-covering property. Since the disjointness problem on vectors of length t has communication complexity $\Omega(t)$ [12], even for nondeterministic or randomized protocols [9] we get:

Theorem 4. *Any protocol that approximates the cover number to within a factor less than $r/2$ requires communication $\Omega(e^{n/(r2^r)})$. The lower bound holds also for randomized or nondeterministic protocols.*

In particular for any $r \leq (1/2 - \epsilon) \log n$, the communication complexity of r-approximation must be exponential.

Acknowledgements

The problems discussed here stem from my joint work with Ilya Segal on the communication compelxity of combinatorial auctions [8]. I thank Ziv Bar-Yosef for pointing my attention to [2], Jaikumar Radhakrishnan and Venkatesh Srinivasan for allowing me to include their proof of theorem 2, and Amir Ronen for comments on an earlier draft.

References

1. R. Aharoni, P. Erods, and N. Linial. Optima of dual integer linear programs. *Combinatorica*, 8, 1988.
2. N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *Journal of Computer and System Sciences*, 58(1):137–137, 1999.
3. V. Chvatal. A greedy heuristic for the set covering problem. *Math. Operations Reserach*, 4:233–235, 1979.
4. Johan Hastad. Clique is hard to approximate to within $n^{1-\epsilon}$. *Acta Mathematica*, 182, 1999.
5. Daniel Lehmann, Liadan Ita O'Callaghan, and Yoav Shoham. Truth revelation in rapid, approximately efficient combinatorial auctions. In *1st ACM conference on electronic commerce*, 1999.
6. L. Lov'asz. The ratio of optimal integral and fractional covers. *Discrete Mathematics*, 13, 1975.
7. Carsten Lund and Mihalis Yannakakis. On the hardness of approximating minimization problems. In *Proceedings of the Twenty-fifth Annual ACM Symposium on Theory of Computing*, pages 286–293, 1993.
8. Noam Nisan and Ilya Segal. The communication complexity of efficient allocation problems, 2001. Preliminary version available from http://www.cs.huji.ac.il/˜noam/mkts.html.
9. A. A. Razborov. On the distributional complexity of disjointness. In *ICALP*, 1990.
10. T. Sandholm, S. Suri, A. Gilpin, and Levine D. Cabob: A fast optimal algorithm for combinatorial auctions. In *IJCAI*, 2001.
11. Rakesh Vohra and Sven de Vries. Combinatorial auctions: A survey, 2000. Availailabe from http://www.kellogg.nwu.edu/faculty/vohra/htm/res.htm.
12. Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *ACM Symposium on Theory of Computing*, pages 209–213, 1979.
13. Edo Zurel and Noam Nisan. An efficient approximate allocation algorithm for combinatorial auctions. In *ACM conference on electronic commerce*, 2001. Available from http://www.cs.huji.ac.il/˜noam/mkts.html.