An optimal competitive on-line algorithm for the minimal clique cover problem in interval and circular-arc graphs

Jerzy W. Jaromczyk
Department of Computer Science
University of Kentucky
Lexington, KY 40506, USA
jurek@cs.uky.edu

Andrzej Pezarski
Institute of Computer Science
Jagiellonian University
Krakow, Poland
pezarski@ii.uj.edu.pl

Maciej Ślusarek
Institute of Computer Science
Jagiellonian University
Krakow, Poland
slusarek@ii.uj.edu.pl

ABSTRACT

We define and study an on-line version of the *shooter problem O-LSP*. In the standard off-line version the problem is: given a finite set S of line segments in R^2 and a point p find the smallest number of half-lines that start at p and intersect all the segments in S. In our *O-LSP* on-line version the segments $S = \{p_1, ..., p_n\}$ are given one by one and the selection of the suitable intersecting half-line for p_i must be done, and cannot be changed, immediately after seeing p_i. In this process we can use a half-line that already intersects some of the segments S_0 in $\{p_1, ..., p_{i-1}\}$ by possibly rotating this half-line in such a way that it keeps intersecting S_0. Via the well-know relation between the shooter problem and circular-arc graphs we reduce it to the on-line Minimal Clique Cover problem (*MCC*) for the classes of interval and circular-arc graphs. Specifically, we are interested in the competitiveness of on-line algorithms for *MCC*. An on-line algorithm A is c_F-competitive for the family F of graphs if for all $G \in F$, $A(G) \leq c_F OPT(G) + b_F$, c_F and b_F constants, where $A(G)$ is the solution found by algorithm A for graph G and $OPT(G)$ is the optimal (off-line) solution. We analyze on-line algorithms and two simple (and seemingly similar) greedy strategies (called LGR and EGR) for MCC (and *O-LSP*) and show both upper and lower bounds on their competitiveness ratio c_F. We demonstrate that:

- $c_F \geq 2$, for any on-line algorithm;
- c_F is unbounded for LGR; hence LGR is not competitive.
- $c_F = 2$ for EGR; hence EGR is optimal.

Keywords
Shooting Problem, On-line algorithms, Interval Graphs, Circular-Arc Graphs

1. DEFINITIONS AND BACKGROUND

An interval graph G is the intersection graph of a family of closed intervals in the real line. Similarly, a circular-arc graph is the intersection graph of a family of closed arcs in a circle. Each vertex v of G corresponds to an interval (an arc, respectively) and the collection of intervals (arcs, respectively) is the representation of G. Interval graphs are often defined in terms of posets as each of them is the co-comparability graph of a set of closed intervals; see [F85]. Note that interval graphs form a proper subclass of circular-arc graphs. The class of interval and circular-arc graphs have been intensively studied, in particular because of their practical applications (e.g., in memory allocation and in organizing records in databases [BL76]).

The clique cover for graph G is defined as the family of subgraphs of G such that each subgraph is a clique and their union is G. Note that subgraphs in the clique cover do not need to be vertex-disjoint. The clique cover of the smallest cardinality $\theta(G)$ is called a minimal clique cover and the algorithmic problem of finding it will be denoted in this paper by *MCC*.

Specifically, we study on-line algorithms for MCC and our motivations stem from the *shooting problem* studied in computational geometry; see [ChN99,JK02]. The problem can be stated as follows: given a finite set S of line segments in R^2 and a point p find the smallest cardinality set of half-lines that start at p and intersect (stab) all the segments in S. After projecting the segments onto a disc centered at p, the problem corresponds to stabbing arcs; see Figure 1. The on-line version of MCC means that the segments are given one-by-one is some order and the shooter must decide immediately after seeing this segment which of the current shooting directions or a new one will be used; after the decision has been made the shots cannot be reassigned. Clearly, segments intersected by the same shot form a clique in the corresponding circular-arc graph and the smallest number of such cliques determines the smallest number of shots. An important theorem, due to Hsu (Theorem 3.2, [HT91]), relates the size of the MCC in circular-arc graphs G with the maximal independent set size $\alpha(G)$.

THEOREM 1.1. *[HT91] If G is not a clique then $\theta(G) = \alpha(G)$ or $\theta(G) = \alpha(G) + 1$.*

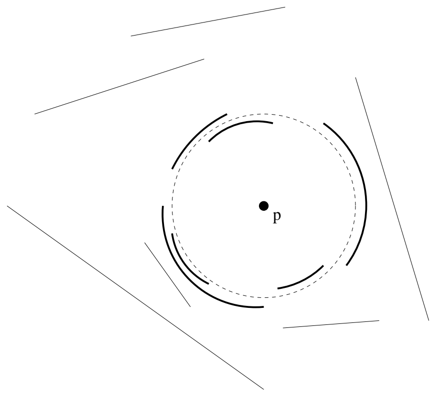

Figure 1: An instance of the shooting problem.

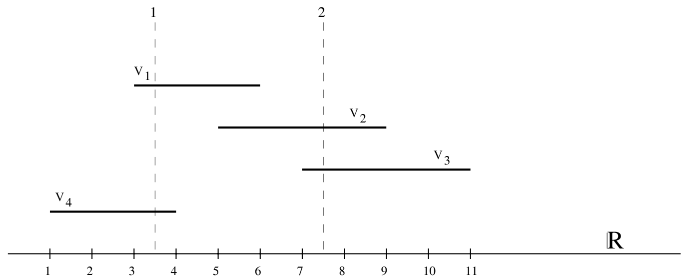

Figure 2: Shots and their ranges: $A(v_1) = A(v_4) = 1$**,** $r_4[1] = [3, 4]$**,** $A(v_2) = A(v_3) = 2$**,** $r_4[2] = r_3[2] = [7, 9]$**.**

MCC problem for graphs G is clearly related to the graph-coloring problem where the objective is to find the minimum integer k, called the chromatic number $\chi(G)$, and a function $f : V(G) \to \{1, 2, \ldots k\}$ such that no edge $e = (u, v)$ has $f(u) = f(v)$. For interval graphs G, we have $\theta(G) = \chi(\overline{G})$, where $(\overline{G})$ is the graph complement of G. There is a number of strong results related to on-line coloring algorithms for interval and circular-arc graphs; see e.g. [K98,KQ95,S89,S95]. Since the complements of interval graphs are not, in general, interval graphs, the above relation does not help to solve our *MCC* problem via coloring. Similarly, results for graph-coloring do not help with the circular-arc graphs.

2. ON-LINE *MCC* PROBLEM

An on-line presentation $G^{\ll}$ of a graph G is a linear order $\ll$ of vertices V of $G(V, E)$. $G_i^{\ll}$ is the on-line graph induced by the first i elements $\{v_1, \ldots, v_i\}$ of V in $\ll$ order. The on-line minimum clique cover (on-line *MCC* in short) is specified as follows:

An algorithm A is an on-line algorithm for the minimum clique cover of an on-line graph G, if given a presentation $G^{\ll}$ with the order $V = \{v_1, \ldots, v_n\}$ it computes a sequence of positive integers $A(v_i), i = 1, \ldots, n$, where $A(v_i)$ is the name of a clique that covers v_i, in such a way, that for each i, $A(v_i)$ depends exclusively on $G_i^{\ll}$.

In other words, the vertices are input one by one, and the number of the clique that covers v_i is established irrevocably after reading $v_1, \ldots, v_i$, together with their adjacency structure.

The quality of on-line algorithms is measured by the *competitive ratio.* An on-line algorithm A is c_F-competitive for the family F of graphs if for all $G \in F$, $A(G) \leq c_F OPT(G) + b_F$, c_F and b_F constants, where $A(G)$ is the solution found by algorithm A for graph G and $OPT(G)$ is the optimal (off-line) solution. In case of the Minimum Clique Cover, we have $OPT(G) = \theta(G)$.

We are looking for competitive algorithms for the on-line *MCC* problem on interval and circular arc-graphs. For the sake of simplicity we describe all results for the class of interval graphs, mentioning the differences for circular-arc graphs whenever necessary.

The main assumption for our considerations is that the input graph is given already in its interval (respectively: circular-arc) representation, i.e., $v_i = [p_i, q_i]$ – a closed interval on the real line or a closed arc on a circle. This is exactly the input to *O-LSP*. The letter s, possibly with subscripts, will be used to denote shots (i.e., the ordinal numbers for the cliques). If A is an on-line clique cover algorithm then by $A(v)$ we denote the shot that covers v, the one that is assigned immediately after reading v. Observe that in general when the algorithm ends there can be also other shots that stab v.

For an interval graph a shot (i.e., a clique) can be depicted as a vertical line and a set of segments stabbed with this line. Usually the choice of such a line for a fixed clique in an interval graph is not unique. To make our reasoning precise we introduce the notion of the shot's range; see Figure 2.

DEFINITION 2.1. *Assume that s is a shot number that is already in use after processing $G_m^{\ll}$. The range $r_m[s]$ of s at phase m is defined as follows: $r_m[s] = \bigcap\{v_j \mid 1 \leq j \leq m, A(v_j) = s\}$.*

Observe that for each shot s in use we have $r_m[s] \neq \emptyset$. The ranges of shots potentially decrease in course of the computation, $r_{m+1}[s] \subseteq r_m[s], m = 1, \ldots, n-1$.

3. GREEDY ALGORITHMS

DEFINITION 3.1. *Algorithm A for the on-line* MCC *problem is greedy if for each vertex v_i in $G^{\ll}$, upon assigning $A(v_i)$, if there is already a shot range that intersects v_i then $A(v_i)$ is not set to a new shot number.*

In other words, a greedy strategy always tries to assign shots already in use, if at all possible. Note that this rule alone may lead to ambiguous decisions. In the following example: $v_1 = [1, 3]$, $v_2 = [4, 6]$, $v_3 = [2, 5]$ a greedy algorithm yields $A(v_1) = 1, A(v_2) = 2$, and $A(v_3)$ can be either 1 or 2. Despite this ambiguity we can formulate an easy yet very important property that turns out to be useful for the subsequent considerations.

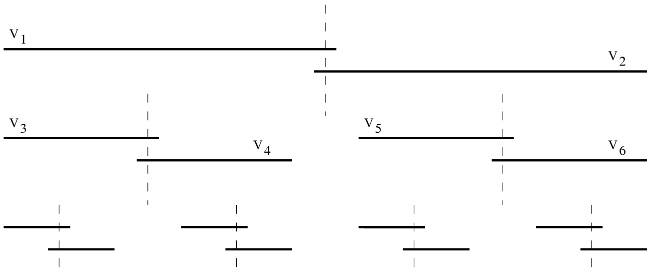

Figure 3: Tree of intervals

LEMMA 3.1. *If $s_1 \neq s_2$ are two shots already in use by a greedy algorithm A at the moment m then $r_m[s_1] \cap r_m[s_2] = \emptyset$.*

PROOF: The lemma follows easily from the monotonicity of shots' ranges. □

We are able to formulate and prove the first of our main results.

THEOREM 3.1. *For any $\epsilon > 0$ there does not exist a $(2-\epsilon)$-competitive algorithm (of any kind) that solves the on-line* MCC *problem.*

PROOF: We are going to build a strategy for an adversary who plays against a *MCC* on-line algorithm A. In order to clear away the influence of the additive constant b_F that stands in the definition of the competitiveness we have to prove that for any $\epsilon > 0$ and $m > 0$ the adversary can impose $A(G)/\theta(G) > 2 - \epsilon$ for some interval graph G such that $\theta(G) \geq m$.

For any natural number k we construct an appropriate sequence of $n = 2^{k+1} - 2$ intervals. The construction for $k = 3$ is depicted in Figure 3. The adversary uses this sequence, perhaps several times, according to the following rules:

1. If all A's clique assignments adhere to the greedy principle (as shown in Figure 3) this part of the game ends after exhausting the whole sequence. We obtain $A(G) = n/2 = 2^k - 1$ and $\theta(G) = \alpha(G) = 2^{k-1}$. Hence $A(G)/\theta(G)$ can be made arbitrarily close to 2 by a suitable choice of k.
2. If at some moment algorithm A defines two new cliques for two consecutive intervals v_{2i-1}, v_{2i}, $i = 1, 2, \ldots$ (that is, A does not adhere to the greedy principle) complete the current level of the tree and finish this part of the game. Let $j - 2$ be the index of the last vertex on this level, i.e. j be the closest power of two greater than $2i$. Then the results are: $\theta(G) = j/4$, and, summing up the last level separately with the remaining ones, $A(G) \geq (j/4 + 1) + (j/4 - 1) = j/2$, which makes the ratio $A(G)/\theta(G)$ greater or equal 2.

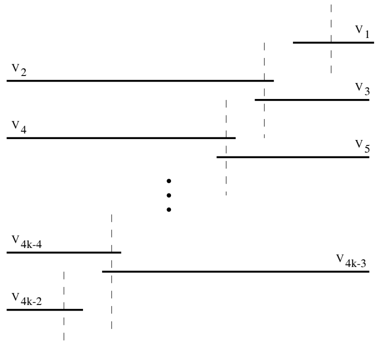

Figure 4: Intervals for *LGR*.

The whole game is constructed by repeating the above strategy in separate areas (intervals) of the real line sufficiently many times in order to obtain the required size of the graph $\theta(G) \geq m$. □

The ambiguity of the clique assignment by the greedy property can be resolved in various ways. We analyze two most intuitive strategies, *leftmost greedy LGR* and *earliest greedy EGR*.

The *LGR* strategy has a clear geometrical flavor, and it works according to the following greedy principle: From the available shots ranges that intersect the current interval choose the leftmost one. Since the ranges of any pair of shots are disjoint (see Lemma 3.1) this assignment is well defined.

The *EGR* works similarly: it selects the earliest available shot instead of the leftmost one. In this regard it resembles classical First-Fit packing algorithms.

Despite some similarity of the two variants of the greedy approach there exists a broad gap between their efficiency. The former turns out to be non-competitive while the latter is optimal.

THEOREM 3.2. *LGR is not competitive.*

PROOF: For any $k > 0$ we construct a sequence of $4k - 2$ intervals, as depicted in Figure 4. It is easy to see that for the graph G generated by these intervals we have $LGR(G) = 2k$ and $\theta(G) = 2$ – just a shot along left ends of odd-numbered intervals and another one through the right ends of the even numbered intervals suffice. Hence the ratio $LGR(G)/\theta(G) = k$ which is arbitrarily large. □

Below we present two lemmas that describe some interesting properties of EGR strategy. Let S be an independent set of maximum cardinality $|S| = \alpha(G)$. We may assume that no interval from V is properly contained in any of the intervals from S; such a set S exists and can be effectively constructed

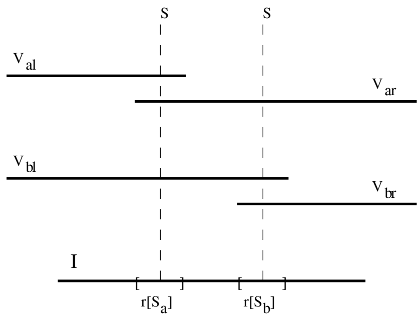

Figure 5: Ranges included in intervals.

for a given interval graph [HT91]. Denote the last property of S by Π. Recall that the size of S states a lower bound to the efficiency of MCC algorithms.

LEMMA 3.2. *At any stage j of computation, for any $v \in S$ there exists at most one shot s such that $r_j[s]$ is properly included in v.*

PROOF: Assume the contrary: at some moment j there are shots s_a, s_b whose ranges are properly included in some interval $v \in S$. From Lemma 3.1 it follows that without loss of generality the situation can be depicted as in Figure 5. We have $EGR(v_{a_l}) = A(v_{a_r}) = s_a$ and $EGR(v_{b_l}) = EGR(v_{b_l}) = s_b$. Observe that property Π implies that all the four intervals considered extend beyond the respective endpoints of v.

Let s_a be the shot that has been introduced earlier of the two. Then s_b could not be assigned to v_{b_l} since at the time v_{b_l} was considered by the algorithm, shot s_a had already been in use. A symmetric argument applies to the case that s_b is earlier than s_a and interval v_{a_r}, showing that the initial assumption is false. This completes the proof of the lemma. □

By replacing property Π with the maximality of the independent set S, a similar argument can be used to prove:

LEMMA 3.3. *Let $v_l = [p_l, q_l]$ and $v_r = [p_r, q_r]$ be two adjacent intervals from the set S such that $q_l < p_r$, and let $I = [q_l, p_r]$ be the gap between them. At any stage j of computation there exists at most one shot s such that $r_j[s]$ is properly included in I.*

From the above two lemmas we derive that the number of shots whose ranges are fully included either in some interval from set S or in some gap between intervals equals $2\alpha(G) + 1$, which is sufficient to prove that EGR is 4-competitive. However, we can show a stronger result.

THEOREM 3.3. $EGR(G) \leq 2\theta(G) + 1$.

PROOF: Assume that the output of algorithm EGR on an interval graph presentation $G^{\ll}$ is given. Order all of the shots from left to right and assign to them new ordinal numbers 1 through $m = EGR(G)$. Since the ranges of the shots are disjoint this assignment is well defined.

Fix $i \in \{1, \ldots, \lfloor m/2 \rfloor\}$ and let $r[2i-1] = [a_{2i-1}, b_{2i-1}]$, $r[2i] = [a_{2i}, b_{2i}]$ be the ranges of two adjacent shots, $a_{2i-1} < b_{2i-1} < a_{2i} < b_{2i}$. Without loss of generality assume that shot $2i-1$ was used for the first time before shot $2i$ was introduced. Then there exists an interval $v_{j_i} = [p_{j_i}, q_{j_i}]$ such that $EGR(v_{j_i}) = 2i$, $q_{j_i} = b_{2i}$ (i.e. v_{j_i} defines the right boundary of the range of v_{2i}), and $p_{j_i} > b_{2i-1}$ (otherwise v_{j_i} would be stabbed by the earlier shot $2i-1$).

Such an interval v_{j_i} exists for each pair of shots $2i-1$, $2i$, $i = 1, \ldots, \lfloor m/2 \rfloor$, and v_{j_i}, $i = 1, \ldots, \lfloor m/2 \rfloor$ are pairwise disjoint. They form an independent set of G of size $\lfloor m/2 \rfloor$. Therefore $\theta(G) \geq \lfloor m/2 \rfloor$, hence $EGR(G) \leq 2\theta(G) + 1$. □
Observe that the whole argument in the proof does not change if we replace intervals by arcs. Therefore we obtain the final result:

THEOREM 3.4. *There is an optimal on-line algorithm for* MCC *and* O-LSP *problems with the competitive ratio* 2.

Acknowledgement Support of the Kentucky Biomedical Research Infrastructure Grant is acknowledged.

Bibliography

[BL76] S. Booth, S. and S. Lueker. Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms, J. Comput. Syst. Sci. 13 (1976), 335-379.

[ChN99] Chaudhuri, J. and S. C. Nandy. Generalized shooter location problem. Lecture Notes in Computer Science 1627 (1999), 389-399.

[F85] Fishburn, P. C. Interval Orders and Interval Graphs: A Study of Partially Ordered Sets. New York: Wiley, 1985.

[HT91] Hsu, W.-L., and K.-H. Tsai, Linear time algorithms on circular-arc graphs, Information Processing Letters 40 (1991), 123-129.

[JK02] Jaromczyk, J. W. and M. Kowaluk, A kinetic view of the shooter problem, Proceedings of the 18th European Workshop on Computational Geometry, 2002.

[K98] Kierstead, H. A., Recursive and On-Line Graph Coloring, in Handbook of Recursive Mathematics, vol. 2, Recursive Algebra, Analysis and Combinatorics, Elsevier (1998), 1233-1269.

[KQ95] Kierstead, H. A. and J.Qin, Coloring interval graphs with First-Fit, Discrete Math. 144 (1995) 47-57.

[S89] Ślusarek, M., A coloring algorithm for interval graphs, in: Mathematical Foundations of Computer Science 1989, Lecture Notes in Computer Science 379 (1989) 471-480.

[S95] Ślusarek, M., Optimal on-line coloring of circular arc graphs, RAIRO Inform. Theor. Appl. 20 (1995) 423-429.