The multi-sensor nuclear threat detection problem

Dorit S. Hochbaum

Abstract One way of reducing false-positive and false-negative errors in an alerting system, is by considering inputs from multiple sources. We address here the problem of detecting nuclear threats by using multiple detectors mounted on moving vehicles in an urban area. The likelihood of false alerts diminishes when reports from several independent sources are available. However, the detectors are in different positions and therefore the significance of their reporting varies with the distance from the unknown source position. An example scenario is that of multiple taxi cabs each carrying a detector. The real-time detectors' positions are known in real time as these are continuously reported from GPS data. The level of detected risk is then reported from each detector at each position. The problem is to delineate the presence of a potentially dangerous source and its approximate location by identifying a small area that has higher than threshold concentration of reported risk. This problem of using spatially varying detector networks to identify and locate risks is modeled and formulated here. The problem is then shown to be solvable in polynomial time and with a combinatorial network flow algorithm.
Keywords: Nuclear threat detection, network flow, parametric cut.

1 Introduction

We consider here a scenario in an urban environment facing potential nuclear threats such as "dirty bombs". With recent technology it has become operational and cost-effective for multiple detectors to be mounted on vehicles in public service. Sodium Iodine detectors are currently deployed on vehicles such as police cars, fire trucks, trains, buses or even taxi cabs. A scenario involving taxi cabs carrying detectors in Manhattan was recently proposed by Fred Roberts (at the ARI Washington DC conference) as a problem of interest. The information transmitted by the detectors

Dorit Hochbaum
UC Berkeley Ca 94720 e-mail: hochbaum@ieor.berkeley.edu

is to be used as input in a process which is to identify a "small region" with a "high concentration" of risk. We formalize and define this problem quantitatively and devise an efficient graph algorithm that solves the problem in polynomial time.

Detecting nuclear threats is a challenging problem under any circumstances. The detection task is more challenging when the relative positions of the detector and the source, if exists, are unknown. The sensitivity of the detectors is diminishing with distance from the source, thus their geographic position impacts the reliability of their reporting. Further, a detector may fail to detect correctly an existing threat (false-negative), or report an alert on the existence of a nuclear source when there is none (false-positive). The likelihood of false reports is diminished and their effect is mitigated when relying on reports from several independent sources. We are interested in reducing the likelihood of false-positive and false-negative reports on detecting nuclear threats in an urban environment. The idea is to mount detectors on every taxi cab in an environment such as New York City, or on police cars in areas where the density of taxi cabs is small. The position of each detector is known at any point in time from GPS information transmitted to a central control data processing facility.

The goal is to identify, at every period of time, a region within the area of interest, which is limited in size and with high concentration of alerts. The purpose is to delineate the presence of a potentially dangerous source and its approximate location. The detectors transmitted information, along with the geographical positioning of the collection of detectors is to be consolidated into reliable reporting on whether nuclear threat exists, and if so, its approximate position. In case this information is deemed to indicate a high enough likelihood of real danger, the detection operations shifts to a high alert state where higher sensitivity detectors and personnel with expertise will be deployed into the region of interest with the task of pinpointing, locating and disabling the source of the threat.

The alert concentration problem is quantified here as an optimization problem, combining two goals: One goal is to identify a small region; another goal is to have large number of alerts, or high concentration of alerts in the region. These two goals are potentially conflicting – focusing on a large number of alerts within an area is likely to result in the entire region; on the other hand focusing on concentration alone would result in a single block of the area containing the highest level of reported alert, thus disregarding information provided by other detectors in the adjacent area.

Overview of paper We first provide the formalism for describing the problem. We then formulate a mathematical programming model for the problem, parametrized by a weight, β, that balances the relative contribution of the two goals. We then show how the problem is solvable in polynomial time as a minimum s,t-cut problem. We further show how to solve the problem for all values of the parameter using a parametric cut procedure.

2 Notation and preliminaries

We introduce here graph theoretic notation to be used in formulating the problem. Without loss of generality we consider the region where the detectors are deployed to be a rectangular area subdivided in grid squares. These will be small enough to contain approximately one vehicle and up to two detectors (although this assumption plays no role in the formulation). Let V be the collection of positions (blocks or pixels of the grid) in the area considered.

We construct a directed graph G with the set of nodes V corresponding to the set of blocks. For each adjacent pair of blocks, if one is within the region and the other outside, the added length to the boundary is 1. The adjacency $[i,j]$ is represented by a pair of arcs in opposite directions each of capacity 1. These arcs are referred to as the "adjacency" arcs of G, and denoted by A_a.

Let $B_1, B_2 \subset V$ be two disjoint sets of nodes in a graph $G = (V, A)$ with arc capacity u_{ij} for each $(i,j) \in A$. The capacity of the *cut* separating the two sets is $C(B_1, B_2) = \sum_{i \in B_1, j \in B_2, (i,j) \in A} u_{ij}$. Note that this quantity is not symmetric as $C(B_1, B_2)$ is in general not equal to $C(B_2, B_1)$.

Let $S \subset V$ be the blocks of a selected sub-region. We measure the *size* of the area delineated by S, by the length of its boundary, counted as the number of block sides that separate S from $\bar{S}$. The length of the boundary of a subset of grid points S is then $\sum_{i \in S, j \in \bar{S}, (i,j) \in A_a} u_{ij}$. Since the set A_a contains arcs of capacity 1, this length is equal to $C(S, \bar{S}) = |\{[i,j] | i \in S, j \in \bar{S}\}|$. Note that there is no requirement that the set S is contiguous. Indeed it can be formed of several connected components. It will be shown that a ratio formulation of the problem can always obtain a solution forming a single connected component.

Let G_{st} be a graph (V_{st}, A_{st}), where $V_{st} = V \cup \{s,t\}$ and $A_{st} = A \cup A_s \cup A_t$ in which A_s and A_t are the source-adjacent and sink-adjacent arcs respectively. A flow vector $f = \{f_{ij}\}_{(i,j) \in A_{st}}$ is said to be *feasible* if it satisfies:
(i) Flow balance constraints: for each $j \in V$, $\sum_{(i,j) \in A_{st}} f_{ij} = \sum_{(j,k) \in A_{st}} f_{jk}$ (i.e., inflow(j) = outflow(j)), and
(ii) Capacity constraints: the flow value is between the lower bound and upper bound capacity of the arc, i.e., $0 \leq f_{ij} \leq u_{ij}$.

A *maximum flow* is a feasible flow f^* that maximizes the flow out of the source (or into the sink), called the *value of the flow*. The value of the maximum flow is $\sum_{(s,i) \in A_s} f^*_{si}$. An s,t cut in G_{st} (or *cut* for short) is a partition of V_{st} to $(S \cup \{s\}, T \cup \{t\})$. The capacity of the cut is $C(S \cup \{s\}, T \cup \{t\})$. The minimum s,t cut is the cut of minimum capacity, referred to here as *min-cut*. It is well known (Ford-Fulkerson [3]) that the maximum value of the flow is equal to the capacity of the min-cut. Every algorithm known to date that solves the min-cut problem, also solves the maximum flow problem.

2.1 The input

The information captured by a detector is a spectrum of gamma ray emissions recording the frequency at each energy level. As such this is not scalar-valued information. The analysis of the detected energies spectrum therefore presents a challenge. The analysis process is currently under development using advanced data mining techniques (by e.g. the DONUTS research group at UC Berkeley, [2]). The output of the analysis is an indication of whether the detected information indicates the presence of a nuclear threat or not.

At a given time instance, let D be the set of positions of taxis with $D^* \subseteq D$ representing the set of positions reporting alert. The set $\bar{D^*} = D \setminus D^*$ is the set of taxi positions reporting no alert.

We will consider an extension of the model to account for varying levels of alert and varying levels of no-alert. This is when the analysis of the detectors transmitted information delivers, for each alert reported by a detector at i, an alert confidence level of p_i. For each no-alert reported from position j, there is a no-alert confidence level of q_j.

3 Formulating the objective function

Since our objective involves multiple goals, we address these by setting a tradeoff between the importance of the short boundary versus the high concentration of alerts. One way of trading these off is by minimizing a ratio function – of the length of the boundary divided by the concentration of alerts in the region. Another, is to use a weighted combination of the goals.

To formalize the goal of "small area" we define an area to be of small size if it is enclosed by a "short" boundary. The boundary of an area is then the number of edges that separate in-region from out-region, or the rectilinear length of the boundary. In the graph $G = (V, A_a)$ defined in Section 2 the length of the boundary of a set $S \subset V$ is $C(S, \bar{S})$. Prior to proceeding, we note that this definition of length needs certain tweaking. Let the set of boundary blocks of the entire area considered be denoted by B. With the definition of the set of arcs A_a, the selection of any subset of B reduces the defined size of the region. For example, if the selected region is all of V then the length of the circumference $C(V, \emptyset)$ is equal to 0. To prevent that, we add a set of arcs A_B from the boundary nodes to an imaginary point in space. This will be quantified in a manner explained later. We let the corner blocks contribute 2 to the length of the boundary, if included in the set. The length of the boundary is thus $C(S, \bar{S}) + |B \cap S|$ where we count the corner block twice in B (instead of introducing additional notation.)

Next we formalized the goal of identifying high concentration of alerts. One indication of the level of alert in an area is the number of alerts at higher than threshold level within the area. Let for now, for the sake of simplicity, assume that the inputs are in the form of alert or no-alert. Let $D \subseteq V$ be the set of position occupied by vehicles. Let $D^* \subset D$ be the set of positions reporting alerts. Part of our objective is then to identify a subregion of positions containing S so that $|D^* \cap S|$ is maximized.

Maximizing the number of alerts within the selected region is an objective that has some pitfalls. For instance, if the region considered, S, contains, in addition to the alerts, also a relatively high number of no-alerts $\bar{D^*} \cap S$, for $\bar{D^*} = D \setminus D^*$, then this should diminish the significance of the alerts in the region. The extent to which the alert significance is diminished is not obvious at this point in time and will require simulation studies, which we plan to undertake. Therefore, we add yet another minimization objective, $\min |\bar{D^*} \cap S|$. This objective is then combined with the length of the boundary objective, as $\min C(S,\bar{S}) + |B \cap S| + \alpha|\bar{D^*} \cap S|$. Although, in terms of the model, we do not restrict the value of α, it is reasonable that α should be a small value compared to the contribution of alert positions, as discussed below. If we choose to disregard the number of no-alerts in the region, then this is captured by setting $\alpha = 0$.

3.1 Ratio function and weighted combination formulations

One way of combining a maximization objective $g(x)$ with a minimization objective $f(x)$ is to minimize the ratio of the two functions $\frac{f(x)}{g(x)}$. For the alert concentration problem the ratio objective function is:

$$\min_{S \subset V} \quad \frac{C(S,\bar{S}) + |B \cap S| + \alpha|\bar{D^*} \cap S|}{|D^* \cap S|}.$$

One advantage of using this ratio formulation is that it is guaranteed that an optimal solution will be a single connected component. This was proved in Hochbaum [5] for a general family of ratio problems. Formally, we define the concept of *additive functions*. For a set of connected components in the graph $A_1, \dots, A_k$, $A_i \subset V$, that are pairwise disjoint, the function $f()$ is said to be additive if $f(\cup_{j=1}^{k} A_j) = \sum_{j=1}^{k} f(A_j)$. For additive ratio functions there is an optimal solution consisting of a single connected component:

Theorem 1. *[Hochbaum 2008] For additive functions f and g, there exists an optimal solution to the problem* $\min \frac{f(x)}{g(x)}$ *consisting of a single connected component and its complement.*

It is easy to verify that our functions here are additive, and hence the existence of a single connected component optimal solution follows.

An alternative to the ratio presentation is to minimize a function which is a linear combination of the two objectives. Using β as a weight for the relative importance of the weights, the objective function is:

$$\min_{S \subseteq V} \;\; C(S,\bar{S}) + |B \cap S| + \alpha|\bar{D^*} \cap S| - \beta|D^* \cap S|.$$

It is in comparison to β that the value of α should be small. This will reflect the perceived relative diminishing of the threat in the presence of no-alerts in the region.

The solution procedure for this linear combination problem can be used as a routine for solving the respective ratio problem. A standard technique for solving a ratio problem is to "linearize" it. The β-question for the problem min $\frac{f(x)}{g(x)}$ is:
Is $\min \frac{f(x)}{g(x)} \leq \beta$?
This is equivalent to solving the linear version
Is $\min f(x) - \beta g(x) \leq 0$?

Therefore if we can solve the linear version of the problem for each β and the logarithm of the number of possible values of β is small enough (of polynomial size) then the ratio problem is solved by a polynomial number of calls to the linear version. Note that the reverse is not necessarily true and the ratio version and the linear version might be of different complexities (for details on these issues that reader is referred to [5].)

Here we devise an algorithm that solves the linearized problem, and for all values of β, in strongly polynomial time.

The linearized objective of the concentrated alert (CA) problem is then modified:

$$\text{(CA)} \quad \min_{S\subseteq V} \ C(S,\bar{S}) + |B\cap S| - \beta|D^*\cap S| + \alpha|\bar{D}^*\cap S|.$$

The problem (CA) has two parameters, β and α. We show solve the problem for all values of β provided that α is fixed, and vice versa. Each of these algorithms will be shown to be running in strongly polynomial time for all values of the parameter.

3.2 Constructing the graph

Let the region be represented by a collection of nodes V of a directed graph where each block is represented by a node. The set of nodes is appended by a *source* dummy node s and *sink* dummy node t.

An edge represents two adjacent blocks, where the adjacency can be selected to be any form of adjacency. Here we use either the *4-neighbors* adjacency or *8-neighbors* adjacency. The weight of each edge is set to 1, and each edge $[i,j]$ is replaced by two directed arcs, (i,j) and (j,i) both of capacity 1. These arcs form the set A_a.

We connect the set of arcs A_B to the sink t with capacities of 1 except for "corner" blocks that contribute 2 to the length of the boundary. Each position that contains an alert taxi in D^* is set to be adjacent to s with a directed arc of capacity β. Each position that contains a no-alert taxi, in $\bar{D}^*$, is set to be adjacent to t with a directed arc of capacity α. We denote these sets of arcs by A_β and A_α respectively.

We therefore constructed a directed s,t graph $G_{st} = (V_{st}, A)$, where $V_{st} = V \cup \{s,t\}$ and $A = A_a \cup A_B \cup A_\beta \cup A_\alpha$. The construction of the graph is illustrated in Figure 1 where an alert position is indicated by a solid circle, and a no-alert position by a crossed circle.

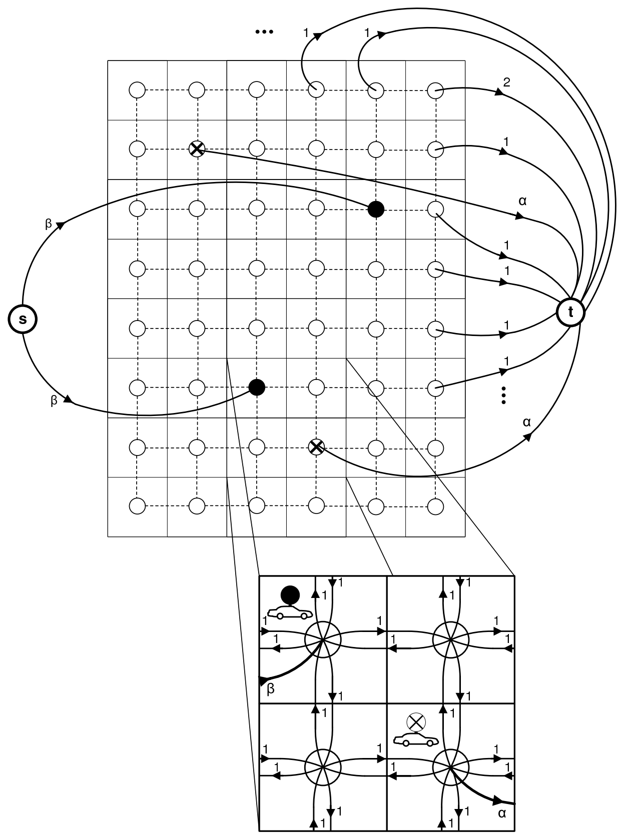

Fig. 1 The graph G_{st}.

We now have a graph $G = (V \cup \{s,t\}, A)$ with arc capacities u_{ij} for arc $(i, j) \in A$, on which we define the minimum s,t-cut problem and show that solving it provides the optimal solution to our CA problem.

4 Main theorem

Let a cut be a partition (S,T) of V of capacity $C(S,T) = \sum_{(i,j)\in A, i\in S j\in T} u_{ij}$.

Theorem 2. *The source set of the minimum cut in the graph G_{st} is the optimal region for problem CA.*

Proof. Let $(S \cup \{s\}, T \cup \{t\})$ be a partition of $V \cup \{s,t\}$ and thus an s,t-cut in G. We compute this cut's capacity:

$$\begin{aligned} C(S\cup\{s\},T\cup\{t\}) &= |B\cap S| + |\bar{D}^*\cap S|\alpha + |D^*\cap T|\beta + \textstyle\sum_{(i,j)\in A, i\in S j\in T} 1 \\ &= |B\cap S| + |\bar{D}^*\cap S|\alpha + (|D^*| - |D^*\cap S|)\beta + C(S,T) \\ &= |D^*|\beta + |B\cap S| + C(S,T) + |\bar{D}^*\cap S|\alpha - |D^*\cap S|\beta. \end{aligned}$$

Now the first term is a constant $|D^*|\beta$. Thus minimizing $C(S\cup\{s\},T\cup\{t\})$ is equivalent to minimizing $|B\cap S| + C(S,T) + |\bar{D}^*\cap S|\alpha - |D^*\cap S|\beta$, which is the objective of the CA problem.

We conclude that solving the concentrated alert problem reduces to finding the minimum s,t cut in the graph G_{st}. The region we are seeking will then correspond to the source set S of the minimum cut $(S\cup\{s\},T\cup\{t\})$.

5 The weighted version of the alert concentration problem

The information provided by the detector may be too ambiguous to translate to a simple binary statement of the form of alert or no-alert. Instead, one defines a threshold level, and within the above-threshold alert category, one creates a function that maps the alert profile transmitted from location i to a weight value p_i that is monotone increasing with the increased confidence in the significance of the alert information.

Similarly, the below-threshold category of no-alert maps into a weight value q_i that is monotone increasing with the increased confidence in the significance of the no-alert information. The modified *weighted concentrated alert* problem is then to find a sub-region S, optimizing the function

$$\text{(WCA)} \ \min_{S\subseteq V} \ C(S,\bar{S}) + |B\cap S| + \alpha \textstyle\sum_{i\in\bar{D}^*\cap S} q_i - \beta \sum_{i\in D^*\cap S} p_i.$$

In order to solve this weighted problem we modify the assignments of capacities to the arcs in the graph G_{st} as follows:
For each position i in D^* we let the capacity of the arc from the source to i be, $u_{si} = \beta p_i$, and for each position i in $\bar{D}^*$ we let the capacity of the arc from i to the sink be, $u_{it} = \alpha q_i$. We call the graph with these modified arc capacities G^W_{st}. We claim that a weighted version of Theorem 2 applies:

Theorem 3. *The source set of the minimum cut in the graph G^W_{st} is the optimal region for problem WCA.*

Proof. The proof is a simple generalization of Theorem 2. We include it here for the sake of completeness.

Let $(S\cup\{s\},T\cup\{t\})$ be, as before, an s,t-cut in G, of capacity:

$$\begin{aligned} C(S\cup\{s\},T\cup\{t\}) &= |B\cap S| + \alpha \textstyle\sum_{i\in\bar{D}^*\cap S} q_i + \beta\alpha \sum_{j\in D^*\cap T} p_j + \sum_{(i,j)\in A, i\in S j\in T} 1 \\ &= |B\cap S| + \alpha \textstyle\sum_{j\in\bar{D}^*\cap S} q_i + \beta(\sum_{i\in V} p_i - \sum_{j\in D^*\cap S} p_j) + C(S,T) \\ &= \beta \textstyle\sum_{i\in V} p_i + |B\cap S| + C(S,T) + \alpha \sum_{i\in\bar{D}^*\cap S} q_i - \beta \sum_{j\in\bar{D}^*\cap S} p_j. \end{aligned}$$

Since $\beta \sum_{i \in V} p_i$ is a constant, the source set of the minimum cut is also minimizing $|B \cap S| + C(S,T) + \alpha \sum_{i \in \bar{D}^* \cap S} q_i - \beta \sum_{j \in \bar{D}^* \cap S} p_j$.

6 Solving for all values of the parameters

As the value of β is changing the solution changes as well. Instead of solving for each value of β we note that the source adjacent arcs are monotone increasing in β and the sink adjacent arcs's capacities are unaffected. Therefore this is a scenario of the parametric maximum flow minimum cut problem. The complexity of solving such a problem is the same as the complexity of solving for a single minimum s,t cut [4] with the push-relabel algorithm. We plan to use the parametric pseudoflow algorithm, [6], that also solves the problem in the same complexity as a single cut. The source code of the solver we use is available at [1].

Since we can find the solution for all values of β, this leads to finding the optimal solution to the respective ratio problem which corresponds to the largest value of β where the solution value is still ≤ 0.

It is possible to conduct the sensitivity on the parameter α independently from that on β. In other words, we keep β fixed and then study the effect on the solution of solving for all possible values of α.

7 Numerical examples

Several instances of the problem were devised on a grid. In the figures below a full circle represents a detector position reporting alert and a crossed circle represents a detector position reporting no-alert. The length of the boundary was taken to be its rectilinear length. That is, the 4-neighbor adjacency was selected. The problem instances were run for $\beta = 3.99$ and $\alpha = 0.5\beta$. The reason why the value of β is just under 4 is to prevent the generation of regions consisting of singletons of alert positions.

In Figure 7 the set V is a 7×10 grid. The set of three alert positions forms the optimal solution. The optimal region is indicated by darker shade. Notice that in this example there are, on row 10, two alert positions separated by an empty position. Although these might indicate an elevated alert status for that area, the vacant grid position rules out selecting this set. The presence of vacant positions therefore should not necessarily be interpreted as diminishing the alert level. These are only the result of a random distribution of the positions.

To allow for regions to be generated even if they contain alert positions separated by a small number of empty grid points, we assign a small value of β, denoted by γ, to each vacant grid point. That means that every vacant position is interpreted as a “minor” alert position and the objective function has an extra term $-\gamma|V \setminus D|$. The modification in the graph of Figure 1 is to add arcs from source s to every

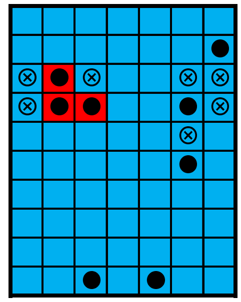

Fig. 2 The solution for a 7×10 grid.

unoccupied square in the grid with capacity γ each. Theorem 2 is easily extended for this case. In the next set of examples we set $\gamma = 0.021$.

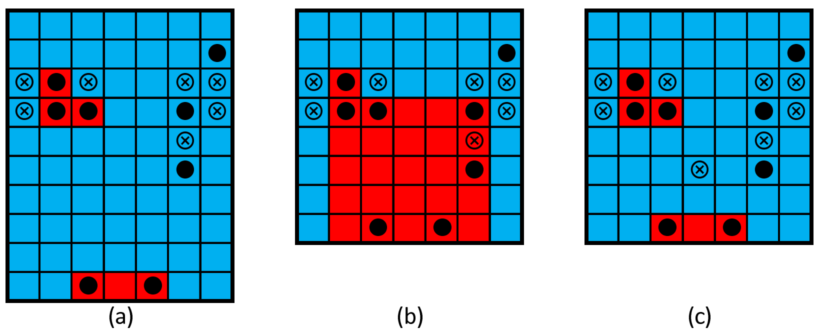

Fig. 3 The effect of introducing γ values to vacant grid positions.

As we see in Figure 3(a), the addition of the γ coefficient indeed changes the optimal solution, and now we have two alert regions, one of which contains an empty position. However, if the two regions are "close", and in the given configuration they are 3 rows apart, as shown in Figure 3(b), then the two regions merge into one. In Figure 3(c) one sees that adding one no-alert position within the region has the effect of separating the two regions. The determination of which values to set and when regions should be consolidated is to be determined by nuclear detection experts and the geographical parameters of the region, as well as the density of the detectors' distribution in the region.

8 Conclusions

We present here a formulation and an efficient algorithm solving the alert concentration problem. The approach presented allows to focus resources on real threats and reduce the likelihood of false-positive and false-negative alerts. We believe that each practical setting should be characterized by the density of the the vehicles carrying detectors, by the sensitivity of the detectors – in terms of distance from a source – and by the finesse of the grid. Each setting requires different values of the parameters β, γ and α. The plan for follow up research is to have these values fine-tuned by simulating the application of the procedure on simulated data. This will be done by considering the resulting size of the region generated and how it corresponds to the detectors' range.

Acknowledgements The author's research has been supported in part by CBET-0736232 and NSF award No. DMI-0620677.

References

1. B. Chandran and D. S. Hochbaum. Pseudoflow solver, accessed January 2007. http://riot.ieor.berkeley.edu/riot/Applications/Pseudoflow/maxflow.html.
2. Domestic Nuclear Threat Security initiative. http://donuts.berkeley.edu/.
3. L. R. Ford and D. R. Fulkerson. (1956). Maximal flow through a network. *Canadian Journal of Math.*, 8, 339–404.
4. Gallo G., M. D. Grigoriadis, R. E. Tarjan. (1989). A fast parametric maximum flow algorithm and applications. *SIAM J. Comput.* 18(1), 30–55.
5. D. S. Hochbaum. Polynomial time algorithms for the normalized cut problem and bi-criteria, multi-objective and ratio problems in clustering, imaging and vision grouping: Part II. 2008.
6. D. S. Hochbaum. The Pseudoflow algorithm: A new algorithm for the maximum flow problem. *Operations Research*, to appear. Extended abstract in: The pseudoflow algorithm and the pseudoflow-based simplex for the maximum flow problem. (1998). Procdeedings of IPCO98. *Lecture Notes in Computer Science*, Bixby, Boyd and Rios-Mercado (Eds.) 1412, Springer, 325–337.