Designing optimal aviation baggage screening strategies using Memetic-Algorithms

Anuar Aguirre, Jose F. Espiritu and Heidi A. Taboada

Industrial, Manufacturing & Systems Engineering Department. The University of Texas at El Paso. 500 W. University Avenue, El Paso, Texas 79968-0521.

Abstract

In the present research, a new memetic algorithm is developed to determine the optimal baggage screening strategy which minimizes the expected annual cost. The cost function includes the costs of false alarms, false clears, purchase costs and operating costs. A baggage screening strategy consists of various hierarchical levels of security screening devices that a checked bag may pass through. In the problem solved, only one type of security device is permitted per level and the number of security devices required at each level is based on the number of bags arriving at each particular level. The solution to the aviation baggage screening problem indicates the number and type of devices to be installed at each level. Different computational experiments using the proposed methodology are presented.

Keywords: Aviation security; Memetic Algorithms; Cost analysis

1. Introduction

Since July of 1996 security on airports was improved by the use of Explosive detection systems, Passenger-Baggage Matching and the Automated Passenger Profiling. This improvement was made by the establishment of the Commission on Aviation Safety and Security by the Vice-President Al Gore. Those improvements were possible because the Federal Aviation Administration (FAA) received funding from the Congress in the Omnibus Consolidated Act of 1997. The FAA received up to $1 billion for security from 1996-2000, and roughly one-third of this funding was for the purchase and deployment of security equipment at airports (Coughlin, et al; 2002). The baggage screening efforts had increased since the terrorist attacks of 2001.

In terms of what should be reported by a screening system and its performance characteristics, certain guidelines have been set by aviation authorities (Singh, et al; 2003). Most of the Baggage Screening Devices that are used today are Explosive Detection Systems (EDS). One of the main problems with the EDS machines is that they cannot discern the difference between common products and known threat items. It is projected that the average rate per year of 1.5 billion of checked bags, with an estimated rejection rate of 30 percent, that makes a 450 million bags per year- more than 1.2 million per day. All of those rejected bags need to either go to further screening by another technology or by hand search. For those cases there is a need of additional machines, time, or labor requirements for more-intensive additional screening of more-intensive additional screening of more than a million bags per day are very onerous (Butler, et al; 2002).

The Checked Baggage Screening (CBS) Model, is a software model developed by the FAA to help in the determination of which baggage screening devices to buy, which airports need more devices, how many devices are located in each airport and where the devices are located in the airport. Checked Baggage Screening Model is an operational cost model developed for exclusive use of the FAA personnel, airline analysts and some economists. CBS helps to forecast the cost of implementing some different baggage screening strategies. CBS forecast different strategies each one having different levels (up to ten), where each level has a specific type of device.
One advantage of the CBS model is that it can be used to determine different strategies for different airports. The CBS model gives the expected annual cost for purchase and operating the devices, where the strategy is based on the number of flights arriving or departing from a specific airport. Also, CBS projects the expected throughput of each baggage-screening device. The expected direct cost per expected prevented attack to the expected cost of an aviation terrorist incident provides one measure for the cost effectiveness of 100% checked bag screening (Jacobson and Karnani, 2005). A bag is alarmed when after screening the bag the security device finds something unfamiliar on it. A bag is cleared when after screening the bag the security device does not find anything unfamiliar on it. Each bag has the possibility of containing an explosive (called threaten bags), and also each device has the probability to fail detecting a threaten bag.

A true clear is when a security device successfully checks a bag and finds that it is safe from any threat; where a false clear is the opposite, a false clear is when a security device finds that a bag does not have any threat when actually it contains a threat. A true alarm is when a security device successfully found a threaten bag, and a false alarm is when a security device found a threat on a bag when actually the bag does not contain any threat. Each outcome has an associated cost, for example when a bag is said to be alarmed, there is an extra effort on it (time and resources) to inspect it and know if it really contains a threat or if it is clear bag. There is no perfect baggage-screening device; each one has the probability of false clear or false alarm. At each level of the strategy chosen, there is the probability to a bag be alarmed and be sent to the next level to be checked again or be cleared to go directly to the aircraft.

In this research the CBS model cost function is used, also 39 different baggage-screening devices are considered. There is set that the maximum number of levels is 10, giving a total of $8.15 \times 10^{48}$ possible strategies. Since it is almost impossible for any computer to solve such number of strategies in a reasonable amount of time, the use of metaheuristics is proposed. Metaheuristics are a class of approximate methods designed to solve hard combinatorial optimization problems arising within various different areas (Crispim, et al; 2005). There are several Metaheuristics methods such as genetic algorithms, ant colony, monkey algorithm or memetic algorithms. In the present research we will develop a new memetic algorithm to solve the aviation baggage screening problem. The main objective of this research is to find the best strategy of baggage screening devices that minimizes the total operational cost.

2. Data and Definitions

Table 1. Shows the data associated with the baggage screening security devices, each device has a specific false alarm probability, false clear probability, purchase cost and throughput (bags per hour). The data used in this study is the same used by Candalino, et al., 2004.

<table>
<thead>
<tr>
<th>Device #</th>
<th>Cost</th>
<th>($P_{fa}$)</th>
<th>($P_{fc}$)</th>
<th>Throughput</th>
<th>Device #</th>
<th>Cost</th>
<th>($P_{fa}$)</th>
<th>($P_{fc}$)</th>
<th>Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>900000</td>
<td>0.4</td>
<td>0.075</td>
<td>1200</td>
<td>21</td>
<td>50</td>
<td>0.2</td>
<td>0.095</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>330000</td>
<td>0.25</td>
<td>0.095</td>
<td>600</td>
<td>22</td>
<td>83000</td>
<td>0.2</td>
<td>0.085</td>
<td>180</td>
</tr>
<tr>
<td>3</td>
<td>250000</td>
<td>0.2</td>
<td>0.085</td>
<td>100</td>
<td>23</td>
<td>16600</td>
<td>0.3</td>
<td>0.075</td>
<td>360</td>
</tr>
<tr>
<td>4</td>
<td>850000</td>
<td>0.15</td>
<td>0.075</td>
<td>100</td>
<td>24</td>
<td>150000</td>
<td>0</td>
<td>0.065</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>850000</td>
<td>0.3</td>
<td>0.065</td>
<td>254</td>
<td>25</td>
<td>965000</td>
<td>0.15</td>
<td>0.0755</td>
<td>120</td>
</tr>
<tr>
<td>6</td>
<td>3000</td>
<td>0.5</td>
<td>0.055</td>
<td>1000</td>
<td>26</td>
<td>850000</td>
<td>0.15</td>
<td>0.0945</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>2000</td>
<td>0.1</td>
<td>0.0545</td>
<td>2</td>
<td>27</td>
<td>330000</td>
<td>0.18</td>
<td>0.0435</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>500</td>
<td>0.5</td>
<td>0.0535</td>
<td>150</td>
<td>28</td>
<td>450000</td>
<td>0.25</td>
<td>0.0825</td>
<td>100</td>
</tr>
</tbody>
</table>
Also for each bag there is a probability that it contains a threat, in this case the probability of a threat \( r_T \) is set as \( 5 \times 10^{-10} \). The annual operating cost per device \( (C_o) \) is $125,000, the cost of a false alarm \( (C_{FA}) \) is set to be $30, and finally the cost of a false clear \( (C_{FC}) \) is $1,400,000,000. The expected number of bags screened per year \( (S) \) is shown in equation 1, where \( S \) is equal to the maximum number of bags arrived per hour over the ten years that are going to be analyzed times the 24 hours of a day times the 365 days of a year.

\[
S = \left( \frac{\text{Maximum bags per hour}}{10} \right) \times 24 \times 365
\]  
\( (1) \)

The expected annual cost of false clears is shown in equation 2

\[
\text{False Clear Cost} = C_{FC} \times S \times r_T \times \left[ P_{FC}(1) + \sum_{i=2}^{n} P_{FC}(i) \prod_{j=1}^{i-1}(1 - P_{FC}(j)) \right]
\]  
\( (2) \)

Where the False clear cost is the multiplication of cost of a false clear times the expected number of bags screened per year times the probability of a threat, all of this times the all the final probability of a false clear after the bag has passed through all the levels. The expected annual cost of a false alarm is the multiplication of the cost of a false alarm time the expected number of bags screened per year times one minus the probability of a bag containing a threat times the probability of a false alarm after passing all the levels as is shown in equation 3.

\[
\text{False Alarm Cost} = P_{FA} \times S \times (1 - r_T) \times \prod_{i=1}^{n} P_{FA}(i)
\]  
\( (3) \)

In order to determine the purchase cost it is needed to know how many devices are needed at each level. In order to find how many devices are needed per level first the number of bags that passes through each level needs to be known. The number of bags per level \( i \) is found by multiplying the maximum number of bags per hour times the probability of a false alarm up to level \( i - 1 \) plus the probability of a true alarm up to level \( i - 1 \), as shown in equation 4.

\[
Bags(i) = \text{Maximum number of bags per hour} \times \left\{ \left( (1 - r_T) \times \prod_{j=1}^{i-1} P_{FA} \right) + \left( r_T \times \prod_{j=1}^{i-1}(1 - P_{FC}(j)) \right) \right\}
\]  
\( (4) \)

Then the number of devices per level \( i \) is calculated as follows:
Having the number of devices per level the next step is to find the cost per level. The cost at level \( i \) is the multiplication of the number of devices at level \( i \) times the purchase cost of device used at level \( i \) as is shown in equation 6.

\[
\text{Level Cost } (i) = \text{Devices}(i) \times \text{PurchaseCost}(device_i) \tag{6}
\]

The total purchase cost will be the sum of all the level cost as is shown in equation 7.

\[
\text{Total Purchase Cost} = \sum_{i=1,2,...,n} \text{Level Cost}(i) \tag{7}
\]

Since the strategy found will be based on a ten-year period ten to obtain the expected annual purchase cost will divide the total purchase cost. Then, the total annual cost will be the summation of the annual purchase cost plus the operating cost times the total number of devices plus the false clear cost plus the cost of a false alarm, the annual total cost can be obtained using equation 8.

\[
\text{Total Annual Cost} = \frac{\text{TotalPurchaseCost}}{10} + C_0 \sum_{i=1}^{n} \text{Devices}(i) + \text{FalseClearCost} + \text{FalseAlarmCost} \tag{8}
\]

### 3. Methodology

A new Memetic algorithm is developed to solve this baggage security-screening problem. Basically, a Memetic Algorithm” (MA) is combination of evolutionary algorithms with local search (Moscato, et al; 1989). Evolutionary Algorithms (EAs) are a class of a search and optimization techniques that work on a principle inspired by nature: Darwinian Evolution. The concept of natural selection is captured in EAs. Specifically, solutions to a given problem are codified in so–called chromosomes (Krasnogor, et al; 2005). One of the most widely used evolutionary algorithms are the Genetic Algorithms (GAs), where two evolution operators are used, the crossover operator and the mutation operator. GAs work with a random population of solutions (chromosomes), where the fitness of each chromosome is determined by evaluating it against an objective function (Kamepalli; 2001). After getting the value of each solution the best chromosomes exchange information to produce new chromosomes. Either the crossover operator or the mutation operator could do this information exchange. The new solutions are then evaluated and used to evolve the population if they provide better solutions than the solution obtained at the initial population. This process is repeated for a large number of generations until a stopping criterion is reached.

Some of the characteristic of the Memetic Algorithms are that: MA’s work with a coding of the parameter set, not the parameters themselves, MA’s search from a population of points, not a single point, MA’s use payoff (objective function) information. MA’s use probabilistic transition rules, not deterministic rules, and MA’s search local solutions to improve the fitness of the current population. A Memetic Algorithm is a hybrid between a local search and a Genetic Algorithm. The Memetic Algorithm uses the same evolutionary operators than the Genetic Algorithm (crossover and genetic operator).

The crossover operator combines the information of two chromosomes (called parent chromosomes) in order to obtain a new chromosome that contains the information of both parent chromosomes. The mutation operator uses the information of a chromosome and modifies it by changing the information of one cell of a chromosome for another completely different. Mutation is a genetic operator that alters one
or more gene values in a chromosome from its initial state. This can result in entirely new gene values being added to the gene pool. With these new gene values, the memetic algorithm may be able find a better solution.

The Memetic Algorithm used in this paper first finds a random population of $N$ chromosomes that will be evaluated. Then, the fitness value of the initial random population is obtained, and the chromosomes are sorted using their fitness value. After the chromosomes are sorted the best solutions are set apart to give an initial population to the local search. In this paper we take the idea of a Tabu Search of switching the values of the genes in the chromosome to generate new solutions. The criterion used is to switch all the pair genes from the best solutions to generate new neighbor solutions. Then fitness value of the new set of chromosomes is found and sorted again, and the best solutions are kept to be used for the genetic operators.

The crossover operator used in this paper is set to choose one gene from parent 1, then one gene from parent 2, and then a gene from parent 1 and so on until the length of the gene is met. The reason to doing this type of crossover operator is because it gives more information from both parents and also because it gives a son chromosome that has a variety of information from both parents. Also another child is born with the information of the same pair of parents both in this case the gene 1 will be the information of the gene 1 at parent 2 instead of the information of gene 1 in parent one, then the gene 2 of the son will be the information of the gene 2 of the parent one and so on until the length of the chromosome is meet. After getting all the new chromosomes all of those are set apart to be joined with the chromosomes generated by the mutation operator. Figure 1 shows the steps of the developed memetic algorithm.

The mutation operator used is set to be 1% and since the length of the chromosome used in this paper is set to be 10 (the maximum number of levels), the maximum information changed is set to be one gene. That gene that is going to be changed is selected at random and also the information of that gene is changed at random. All the new chromosomes generated by doing the mutation operator were joining with the chromosomes found using the crossover operator and for all of these new chromosomes the fitness value is obtained. Finally these values are compared with the values obtained after the local search and the best $N$ solutions are then set to be our new initial population, this process is repeated 1000 times.

4. Computational Results

A Matlab Program was developed and using all the information previously presented the problem was solved. The software developed was run in Matlab R2010a by Matworks® on a 2.7 GHz Intel Core i7,
4GB 1333MHz DDR3 computer taking almost 3 minutes to solve the problem. The solutions obtained are summarized in the table 1.

Table 1. Final Solutions

<table>
<thead>
<tr>
<th>Solution</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>31 35</td>
<td>7.5960e+007</td>
</tr>
<tr>
<td>10 2 11</td>
<td>5.1868e+007</td>
</tr>
<tr>
<td>7 24 1 21</td>
<td>3.3223e+007</td>
</tr>
<tr>
<td>34 6 14 37 10</td>
<td>2.2121e+007</td>
</tr>
<tr>
<td>8 29 20 24 32 36</td>
<td>3.3871e+007</td>
</tr>
<tr>
<td>32 14 31 12 13 34 25</td>
<td>3.8645e+007</td>
</tr>
<tr>
<td>1 16 22 38 34 10 37 31</td>
<td>2.9438e+007</td>
</tr>
<tr>
<td>24 10 8 17 25 30 20 28 24</td>
<td>3.6221e+007</td>
</tr>
<tr>
<td>9 20 26 14 25 36 16 28 34 32</td>
<td>2.9029e+007</td>
</tr>
</tbody>
</table>

5. Conclusions

A Memetic Algorithm was developed and used to select a strategy for a baggage-screening problem, where the cost is the selection parameter. The cost is a representation of a generalized cost function that consists of: Purchase, False Alarm, False Clear and Operating Costs. The parameters used for the Memetic Algorithm were an initial population of 100 chromosomes and the Memetic Algorithm was run 1000 times. The Memetic algorithm used take a local search similar to a Tabu Search for the best 40 initial solutions. Then, a crossover and mutation operators were performed to the best solutions. The problem was solved by creating a program using Matlab and solved in an approximate time of 3 minutes. The best solution found is [34-6-14-37-10], where the in the first level a machine number 34 is used, then at level 2 the machine 6 selected and so on until at level 5 the machine 10 is selected. This solution gives a cost of 2.2121e+007 dollars. This approach has many possible modifications and improvement that can be done. For example changing the cost of false alarm or false clear will modify the solution set found. Also, this problem can be solved as a multi-objective problem, giving a set of Pareto-optimal solutions.

References