A First Simulation on Optimizing EDS for Cabin Baggage Screening Regarding Throughput

Yanik Sterchi and Adrian Schwaninger

University of Applied Sciences and Arts Northwestern Switzerland (FHNW)

Olten, Switzerland

Center for Adaptive Security Research and Applications (CASRA)

Zurich, Switzerland

Abstract—Airport security screening is vital for secure air transportation. Screening of cabin baggage heavily relies on human operators reviewing X-ray images. Explosive detection systems (EDS) developed for cabin baggage screening can be a very valuable addition security-wise. Depending on the EDS machine and settings, false alarm rates increase, which could reduce throughput. A discrete event simulation was used to investigate how different machine settings of EDS, different groups of X-ray screeners, and different durations of alarm resolution with explosives trace detection (ETD) influence throughput of a specific cabin baggage screening process. For the modelling of screening behavior in the context of EDS and for the estimation of model parameters, data was borrowed from a human-machine interaction experiment and a work analysis. In a second step, certain adaptations were tested for their potential to reduce the impact of EDS on throughput. The results imply that moderate increases in the false alarm rate by EDS can be buffered by employing more experienced and trained X-ray screeners. Larger increases of the false alarm rate require a fast alarm resolution and additional resources for the manual search task.

Keywords—aviation security; explosive detection systems (EDS); human factors; discrete event simulation; throughput

I. INTRODUCTION

A secure air transportation system is essential for society and economy. Repeated terror attacks [1] have led to increased aviation security measures. All passengers and their belongings are screened at an airport security checkpoint (ASC) to ensure that they are not carrying any prohibited items (guns, knives, improvised explosives (IEDs), and other threat items). At a typical ASC passengers first have to divest themselves of their luggage and other items like pocket content, jackets, and headwear. These are then scanned by an X-ray machine and an airport security officer (ASO) searches the X-ray images for prohibited items. If a suspicious item is detected, the ASO reviewing the X-ray images (X-ray screener) declares the image as "not OK" (which is also referred to as an "alarm") and the corresponding bag or item is redirected for further inspection (secondary search). This is referred to as "alarm resolution" and typically includes a manual search of the bag and/or explosives trace detection (ETD), which consists of taking a swab at several locations of the bag and then having a machine analyze that swab for traces of explosive residue [2], [3]. If the suspicious item turns out to actually be a prohibited item, the X-ray screener's alarm (i.e. declaring the item as "not OK") is called a "hit". If the suspicious item turns out to be harmless, the X-ray screener's alarm is called a "false alarm". As described, both technology and humans are involved in the process. How well the whole system performs therefore depends on human factors, machine attributes, and the process defining their interaction [4]–[7].

In recent years, the detection of explosives has been increasingly in the focus of security in civil aviation [3]. Manufacturers of detection equipment have recognized the increasing threat by explosives and the difficulty to detect them without additional technological assistance [8]. Singh and Singh [9] or Wells and Bradley [3] provide a good overview of different technologies developed for the detection of explosives. As Sing and Singh [9] point out, X-ray technology is the most common method for luggage inspection at airports. In recent years, X-ray based EDS machines have been made available for cabin baggage screening. Such machines can use dual energy X-ray imaging to detect explosives via the estimation of the effective atomic number and material density [8], [9].

The introduction of EDS into cabin baggage screening is certainly an advantage security wise. But how does EDS affect throughput, i.e. the amount of items that can be screened within a certain time? Butler and Poole [2] argued that EDS can reduce throughput, but since then EDS machines have become faster and more reliable. It would therefore be interesting to examine effects of EDS on throughput taking into account up-to-date information on technology, humans and processes. In this study this was explored for one specific process using discrete event simulation. In addition, two measures to cope with potential negative effects on throughput were tested for their effectiveness.

II. PROCESS DESCRIPTION AND ASSUMPTIONS

System performance of an ASC depends on technology, humans, and the process defining their interaction. The difference between a conventional X-ray machine and an X-ray machine with EDS is that the latter analyzes the X-ray image information for potential explosives before the X-ray image is displayed to the X-ray screener. The analysis by the EDS can
be quite fast and does not necessarily delay the reviewing of the X-ray images by the X-ray screener. Once the EDS generates an alarm, there are at least two different approaches to resolve these alarms. One is on-screen alarm resolution: When the ASO reviews the X-ray image of the bag that triggered the alarm by the EDS, a frame is displayed around the area of the X-ray image which might contain explosives. The ASO then decides whether the bag needs further alarm resolution (e.g. using ETD and/or manual search). Another approach is to increase the level of automation (for an overview of levels of automation see [10]) and automatically redirect items that caused an alarm for alarm resolution by ETD and/or manual search. In this study we restricted ourselves to this second approach, which will be referred to as "automatic decision scenario". Further, the alarm resolution process has to be specified. It seems to be more adequate to resolve alarms by the EDS using ETD instead of manual search, as ETD is specialized for detecting explosives. But ETD is not suited for detecting other prohibited items. If an X-ray image triggers an alarm by the EDS, the X-ray screener would still need to review this image for prohibited items other than IEDs and explosives and pass it on for manual search, if required. If for example there was a knife in a bag that set off the alarm by the EDS and this alarm was only resolved with ETD, the knife would pass undetected. For this study the process is specified as illustrated in Fig. 1.

A. Relevant Variables of Machine and Human

With regard to ASC throughput the most important attributes of an EDS machine seem to be its ability to detect explosives, the amount of false alarms it generates, and how long it needs to process a bag. Comparable to an alarm by the X-ray screener, an alarm by the EDS can either be a hit (if an explosive is present) or a false alarm (if no explosive is present). If the EDS does not generate an alarm for a certain bag or item, this is called a "correct rejection" in the case that the decision was correct and no explosives were present and a "miss" in the other case. How likely these events are, depends on how many trace samples are taken [2]. To take into account that different durations for alarm resolution with ETD are possible, three different scenarios were tested in the current study: One with a low average duration of 30 s, a second taking 60 s, and a third taking 120 s on average.

Beside the duration of the ETD, a second aspect is important with regard to resolving the alarms by the EDS: The number of the alarms by the X-ray screener. The more alarms (both false alarms and detected prohibited items, e.g. liquids or scissors) the X-ray screener generates, the more busy the ASO responsible for alarm resolution by secondary search will be and the less capacity he or she will have to resolve alarms by the EDS. There are several factors known to produce substantial variance in X-ray screeners' hit and false alarm rates. Especially important is initial and recurrent individually adaptive computer-based training, which has been shown to be an effective and efficient tool to learn which items are prohibited and what they look like in X-ray images [7], [12]–[14]. Such training has also been shown to reduce false alarm rates [6]. Several studies did not find an effect for job experience alone, if not accompanied by training [5], [15]. Age was found to have a negative effect [5], but a rather small one compared to the effect of training [16]. In order to take into account potential influences of differences in training, age, and job experience of X-ray screeners, the simulation will be based on data from three different X-ray screener populations, which vary with regard to training hours, job experience, age, and airport.

So far, attributes of technology and humans which seem the most important for the effect of EDS on throughput have been...
discussed. A third aspect to consider is human-machine interaction. To test whether X-ray screeners reduce their false alarm rate when EDS is available, [17] conducted a laboratory experiment, where 150 certified X-ray screeners had to review X-ray images either with on-screen alarm resolution, automatic decision, or without any assistance by an EDS. There were no significant differences in X-ray screeners’ false alarm rates or evaluation times between the baseline condition without assistance and the condition with automatic decision. For the simulations of the present study we will therefore assume that the performance of X-ray screeners is not affected by EDS.

As explained above, the false alarm rate of the EDS machine, the false alarm rate of the X-ray screener, and also the duration for applying ETD are crucial for the effect of EDS on throughput in our automatic decision scenario, because these three factors affect the workload of the ASO responsible for resolving the alarms. In this study, two possible measures were examined on their effectiveness to reduce the workload of this ASO. The first and quite obvious measure is to assign a second ASO to the task of resolving alarms using manual search and/or ETD. This should double the rate at which alarms can be resolved (assuming there is sufficient room and equipment provided). Typically there is a limit to the number of items that can queue for alarm resolution. If this queue limit is reached, the X-ray screening process is interrupted until the responsible ASO has finished resolving one of the alarms and the queue is below its limit again. A second possible measure to expedite secondary search could be to instruct the X-ray screener to resolve one of the alarms as soon as the queue limit is reached and the screening process is interrupted. This allows the X-ray screener to use the time productively that he or she would otherwise spend waiting. A disadvantage of this measure is that the X-ray screener can still be busy resolving the alarm using manual search and/or ETD while the X-ray screening process is ready to be continued. Both these measures were tested in the simulation of the present study for their potential to reduce the impact of EDS on throughput.

### III. Method and Procedure

The simulation was implemented in FlexSim, an off the shelf 3D modelling and discrete event simulation software. Fig. 2 shows a screenshot of the model ASC lane. The basic layout, processes, and parameters of the model were set in accordance with a specific ASC design of a European airport, hereafter referred to as "reference ASC". To gain insight into the processes and parameters of the reference ASC, data from a previous work analysis were used. Further information was provided by ASOs working at the reference ASC and by experts in the field of aviation security.

In this section the model assumptions are described in their order within the baggage screening process. TABLE I gives an overview of the assumed model parameter values and distributions. At the beginning of the baggage screening process is the arrival of the passenger at the ASC. The model is set up to provide a constant flow of passengers to simulate capacity, i.e. the throughput that can be achieved if there constantly are passengers ready to be screened. In a second step, passengers place their baggage and other belongings on the conveyor belt with the help of an ASO. According to several interviewed ASOs working at the reference ASC, this should take only about 5 s per item, as most passengers prepare their belongings while waiting for their turn to place their items on the conveyor belt. In accordance with [18] divesting time was modeled to be gamma distributed. For the baggage screening process it is not directly relevant how many items each passenger carries; an average of 3 items per passenger with a minimum of 1 was assumed.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Distribution</th>
<th>Mean and Standard Deviation in Brackets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placing item on conveyor</td>
<td>Gamma</td>
<td>5 s (5 s)</td>
</tr>
<tr>
<td>Items per passenger</td>
<td>Poisson (translated)</td>
<td>3 (2)</td>
</tr>
<tr>
<td>Evaluation time X-ray screener</td>
<td>Empirical</td>
<td>CR: 3.90 s (1.32 s) Miss: 5.13 s (2.67 s)</td>
</tr>
<tr>
<td>Duration of alarm resolution with manual search</td>
<td>Lognormal</td>
<td>116 s (132 s)</td>
</tr>
<tr>
<td>Duration of alarm resolution with ETD</td>
<td>Gamma (translated, shape = 1)</td>
<td>Condition 1: 30 s (5 s) Condition 2: 60 s (10 s) Condition 3: 120 s (20 s)</td>
</tr>
</tbody>
</table>

CR: correct rejection; FA: false alarm

After the items have been placed on the conveyor belt, the items are screened by the X-ray machine. Then, the X-ray image is analyzed by the EDS and reviewed by the X-ray screener. For this component of the process the false alarm rate and evaluation time of both the X-ray screener and EDS machine have to be defined for the simulation. To model the alarm rate and evaluation time of the X-ray screener, empirical data from three groups of [17] was used. TABLE II shows how these reference groups differ with regard to their false alarm rate, training hours conducted with X-ray Tutor Version 3, work experience, age, and the airport they work at. Separate simulations were conducted for the false alarm rates of these three reference groups to explore how differences

1 The study was conducted using an X-ray screening simulator software and the EDS detection performance and settings were provided by authorities who are responsible for testing and certifying X-ray screening equipment with EDS functionality.

2 [17] tested four groups: from two different airports and with two different levels of work experience (less than one year or more than two years). Due to the low number of new ASOs at the first airport, the group of ASOs with less than one year work experience from the first airport was only tested for the control condition and is therefore not considered in this study.

3 Information on the X-ray Tutor computer-based training software can be found at www.casra.ch and for an earlier version of the software in [14]. Reference group 3 received initial training using another computer based training; the number of training hours per ASO could not be determined.
between X-ray screener groups affect the relationship between EDS and throughput.

<table>
<thead>
<tr>
<th>TABLE II. REFERENCE GROUPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Averages and Standard Deviations in Brackets</td>
</tr>
<tr>
<td>Airport</td>
</tr>
<tr>
<td>Reference group 1 Airport 1</td>
</tr>
<tr>
<td>Reference group 2 Airport 2</td>
</tr>
<tr>
<td>Reference group 3 Airport 2</td>
</tr>
</tbody>
</table>

In addition to false alarms there are some quite common prohibited items (e.g., liquids, gels, or scissors) that require alarm resolution. For the simulation, the empirical value from a work analysis at the reference ASC of 1.23% bags and trays containing detected prohibited items was taken.

At the reference airport, the X-ray screener has a minimum of 3.5 s to review an X-ray image before the next X-ray image appears on the screen. This minimum is defined by the belt speed, the required distance between the screened items, and the average size of these items. If an X-ray screener needs more than 3.5 s to evaluate an X-ray image, he or she can stop the belt temporarily. To model these durations, empirical evaluation time distributions of [17] were used and capped at the minimum of 3.5 s. These evaluation times did not differ much between the three reference groups, but depended on the decisions of X-ray screeners ("OK" or "not OK") and in case of "not OK" on whether there actually was a prohibited item present (hit) or not (miss). These differences were taken into account by using three different empirical distributions.

There are EDS machines available which have the same belt speed as the current X-ray machines at the reference ASC, meaning that an EDS would not necessarily affect the rate at which items are X-rayed. Hence the EDS was assumed not to require additional evaluation time in the simulation model. As explained in the previous section, EDS machines can greatly vary in their false alarm rates depending on technology, machine type, and targeted hit rate. Therefore, false alarm rates ranging from 1 to 15% were explored.

After the item leaves the X-ray machine, it can be picked up by the passenger if both the X-ray screener and the EDS have cleared the item (i.e., not generated an alarm). In case the X-ray screener, the EDS, or both produce an alarm, the item has to be redirected for alarm resolution. This can either be done manually or automatically. In the automatic decision scenario modeled in this study, the alarms were redirected automatically to not disrupt the X-ray screening process. If an item has been declared as "not OK" by the X-ray screener, it is assumed that the alarm resolution follows the same procedures as they are currently applied at the reference ASC. The time needed to resolve alarms of the X-ray screener was measured at the reference ASC and could be approximated well with a lognormal distribution with a mean of 116 s and a standard deviation of 132 s.

In case the EDS generates an alarm, this alarm is assumed to be resolved with ETD with an average duration of either 30 s, 60 s or 120 s, as explained in the previous section. It is assumed that the X-ray screener also reviews X-ray images which triggered the alarm by the EDS in order to detect prohibited items other than explosives. If that occurs, the X-ray screener sends the screened item to secondary search including a manual search (see also Fig. 1).

After the alarm has been resolved, the passenger can recollect his or her belongings. How long the passenger needs for the recollection should not affect baggage throughput, as long as there is enough space available. Recollection was therefore modeled not to influence the baggage screening process in terms of throughput. At the reference ASC there is a limit of three items that can queue for alarm resolution and the X-ray screening process is interrupted if this limit is reached. This was modeled accordingly.

Separate simulations were run for each combination of the three reference groups, 15 false alarm rate levels of the EDS (1-15%), one level without EDS, and the three durations for alarm resolution using ETD (30 s, 60 s, 100 s). To test the effectiveness of the two measures described in the previous section, they were also run for each of the 15 false alarm rate levels of the EDS plus one level without EDS. To keep the results manageable, the measures were only tested in combination with the first reference group (which was mutated alarm resolution by an ASO using ETD requires 25, 50 or 100 s respectively in most cases but that it can require longer in some cases. This was modeled with a gamma distribution with a shape factor of 1 and a mean of 5, 10 or 20 s respectively added to the mentioned minima.

In rare occasions passenger screening can produce a disruption in baggage screening, e.g., if a security officer has to wait for the passenger before manual search of a bag can be conducted. As the focus of this study is on the baggage screening process, passenger screening will not be considered in the model, also because it is not directly affected by the introduction of an EDS. But the baggage screening process can be affected by the passenger screening process and can therefore not be expected to always achieve its full potential throughput in operation.
Fig. 3. Mean and standard error (over 200 simulation runs) of throughput in items per hour, depending on false alarm rate of EDS (zero representing the baseline without EDS) and on reference group, green dashed: reference group 1 (airport 1, tenure > 2 years), blue dotted: reference group 2 (airport 2, tenure > 2 years), red solid: reference group 3 (airport 2, tenure < 1 year), and mean duration of alarm resolution using ETD of 30, 60 and 120s.

Fig. 4. Mean and standard error of throughput in items per hour, depending on false alarm rate of EDS, either with: red solid: single security officer resolving alarms, blue dotted: X-ray screener assisting with alarm resolution in case screening process is interrupted, green dashed: second security officer assigned to resolution of alarms.

As pointed out previously, a false alarm rate of 6% is feasible for certain currently available EDS equipment. For an average of 120 s for a secondary search with ETD, throughput was reduced by almost 40% compared to the baseline condition without EDS. If an alarm resolution with ETD only takes 60 s on average, then the reduction in throughput was much less, but also depended on the reference group of X-ray screeners: 19% for the first (i.e. the most experienced and best trained X-ray screeners), 23% for the second, and 25% for the third reference group (i.e. the least experienced and trained X-ray screeners). If the duration of alarm resolution with ETD was reduced to an average of 30 s, then throughput only decreased by 12% for the first, 15% for the second, and 17% for the third reference group. In this condition the well trained and experienced X-ray screeners of the first reference group still achieved a higher throughput than the third reference group with less than one year of job experience in the baseline condition without EDS (see Figure 2). Thus, the simulation results imply that a high throughput is still possible with EDS, if fast alarm resolution procedures using ETD can be implemented and if X-ray screeners are well trained.

Fig. 4 shows the relationship between capacity and the false alarm rate of the EDS for the standard security lane and the two measures that could be used to minimize negative effects on throughput as explained in the previous section. As could be expected, assigning a second ASO to the task of resolving alarms massively reduced the impact of the EDS’s false alarm rate on throughput (assuming the tools and space for parallel alarm resolution are available). Within the simulation, instructing the X-ray screener to resolve one alarm while the screening process is interrupted only started having a positive effect on throughput at higher false alarm rate levels. In practice however, the X-ray screener might coordinate with the ASO responsible for alarm resolution and support him or her with tasks short enough not to delay the screening process too much. Therefore, having the X-ray screener assist with alarm resolution could be more useful in practice than it was found to be the case in the simulation.
V. SUMMARY, CONCLUSIONS, AND LIMITATIONS

The results of the discrete event simulation indicate that the baggage throughput of an ASC can strongly be affected by EDS. This effect is mainly due to the time needed for alarm resolution using ETD, which highlights the importance of fast ETD alarm resolution procedures (e.g. efficient trace sampling) and a short analysis time of the equipment. Not only the false alarm rate of the EDS machine and alarm resolution time of ETD but also the false alarm rates of the X-ray screeners were found to be very important. Training has been shown to reduce false alarm rates [6]. Potential decreases in baggage throughput due to the introduction of an EDS could therefore be at least partially compensated by having well trained X-ray screeners.

Having a second ASO to expedite alarm resolution could effectively reduce the negative impact of an EDS on throughput, while help by the X-ray screener with alarm resolution seems not to be a useful option based on the simulation results. A field study or a further work analysis combined with simulation could clarify if more coordinated assistance with alarm resolution by the X-ray screener (i.e. by only performing tasks that do not prolong the interruption of the X-ray screening process) has the potential to increase capacity.

Another limitation of the present model is that the false alarm probabilities of the items were assumed to be independent from each other. This does not necessarily need to be the case in practice. Certain passenger groups are likely to have increased or reduced alarm probabilities. Especially at small airports or decentralized ASCs, where passenger groups are less mixed, there might be periods requiring more alarm resolutions than other periods. This might mitigate average capacity due to the non-linear relationship between alarm rate and capacity.

Common cause failure of X-ray screeners and EDS machines could be investigated empirically and considered in future models. Also, more research is needed on the effect of EDS on X-ray screeners’ performance under varying machine settings. The performance of the X-ray screeners working with EDS should also be examined over longer periods and in the field, as there is indication that the influence of automatic decision aids changes with experience [19] and training [20]. Another limitation concerns the 1.23% of detected prohibited items measured with a work analysis. The first reference group was a sample of the ASOs working at the reference ASC where the work analysis was conducted. It could however be expected that the other reference groups with less training and higher false alarm rates might also detect less prohibited items.

In sum, this first study on the effect of introducing EDS in cabin baggage screening provided important results, which could already have practical implications. However, more research including data collection from different ASCs as well as laboratory and field experiments are needed to validate and enhance these discrete event simulation results.

REFERENCES