Optimising multi-layered security screening

David Anderson¹

Received: 30 June 2021 / Accepted: 25 October 2021 / Published online: 22 November 2021
© The Author(s) 2021

Abstract

Screening for prohibited items at airports is an example of a multi-layered screening process. Multiple layers of screening – often comprising different technologies with complementary strengths and weaknesses – are combined to create a single screening process. The detection performance of the overall system depends on multiple factors, including the performance of individual layers, the complementarity of different layers, and the decision rule(s) for determining how outputs from individual layers are combined. The aim of this work is to understand and optimise the overall system performance of a multi-layered screening process. Novel aspects include the use of realistic profiles of alarm distributions based on experimental observations and a focus on the influence of correlation/orthogonality amongst the layers of screening. The results show that a cumulative screening architecture can outperform a cascading one, yielding a significant increase in system-level true positive rate for only a modest increase in false positive rate. A cumulative screening process is also more resilient to weaknesses in the individual layers. The performance of a multi-layered screening process using a cascading approach is maximised when the false positives are orthogonal across the different layers and the true positives are correlated. The system-level performance of a cumulative screening process, on the other hand, is maximised when both false positives and true positives are as orthogonal as possible. The cost of ignoring orthogonality between screening layers is explored with some numerical examples. The underlying software model is provided in a Jupyter Notebook as supplementary material.

Keywords Aviation security · Systems thinking · Multi-layered screening · Orthogonality · Detection

✉ David Anderson
david.anderson@ec.europa.eu

¹ Joint Research Centre, European Commission, Retieseweg 111, 2440 Geel, Belgium

Introduction

Screening for prohibited items at airports is an example of a multi-layered screening process. Multiple layers of screening – often comprising different technologies with complementary strengths and weaknesses – are combined to create a single screening process. For example, passengers may be screened with (i) a metal detector, (ii) pat down, and (iii) trace detection equipment. Checked baggage might be screened by (i) X-ray explosive detection systems, (ii) human analysis of X-ray images, and (iii) hand search of luggage. The detection performance of the overall system depends on multiple factors, including the performance of individual layers, the complementarity of different layers, and the decision rule(s) for determining how outputs from individual layers are combined. The aim of this work is to understand and optimise the overall system performance of a multi-layered screening process using operations research.

Operations research can employ a variety of approaches (research, simulation, system analysis) to understand and optimise the performance of components and whole systems, often with a view to optimising overall outcomes given real-world resource constraints. Wright et al. (2006) published a survey of operations research in support of homeland security. They identified over 170 papers and categorised them according to either the threat vector (chemical, biological, radiological, nuclear or explosives) or the thematic portfolio (border & transport security, critical infrastructure protection, cyber, emergency preparedness & response, threat analysis). In the area of aviation security, they identified operations research concerning passenger profiling, passenger screening and design of access control systems, false alarm minimisation, costs of bag-matching, cost analysis of different screening policies, and modelling of passenger throughput. Lee et al. (2008) reviewed operations research as applied specifically to aviation security, grouped according to five categories: i) passenger and carry-on baggage screening, ii) checked baggage screening, iii) performance indicators, iv) analysis of success rates and v) design of effective screening systems. Within these categories, they surveyed over 50 articles dealing with, for example, the effectiveness of systems to estimate passenger risk level, cost–benefit analyses, deterrence, system design, lowering false alarms and associated costs, threat assessment, human factors and airport logistics.

Jackson & LaTourrette (2015) examined layered aviation security in a holistic sense, encompassing not only checkpoint screening but also intelligence, pre-screening, crew vetting, air marshals, and so on. Their analysis is qualitative, not quantitative, but the article provides an excellent and descriptive insight into the complex myriad of inter-dependencies that may exist between different layers of security. The authors point out that an assessment of the effectiveness of security layers must also take into account the behaviour of the attacker – who will actively seek the path of least resistance and modify their attack in response to security measures. The authors argue convincingly that “*layers in a multi-layer security system will not always combine as straightforwardly as intuition would suggest*”.

Johnston (2010) identified potential pitfalls in multi-layered security for diverse applications, noting that “just because each layer ... is intended to provide

some sort of security does not mean that each layer will compensate for the weaknesses of all the other layers". He stressed the importance of critical thinking and avoiding a wishful-thinking mentality, and pointed to cultural and psychological habits that are sometimes correlated with a sub-optimal multi-layered security process.

Stewart & Mueller (2018) also examined layered aviation security in a holistic sense. Pointing to the apparent mismatch between the significant expenditures in aviation security and the limited cost–benefit or risk analysis, the authors calculated the benefit-to-cost ratio for many of the different layers of security, identifying those which they deem cost-effective and those not. Their analysis assumes a single probability of detection for each layer and does not take into account dependency between each of the layers.

An earlier paper by Stewart & Mueller (2013) examined the cost and effectiveness of three specific layers of aviation security in the US, namely secondary physical barriers, federal air marshals and armed flight deck officers. The analysis was based on a model where an alarm from any layer caused a system-level alarm. The authors conclude that air marshals fail the cost–benefit assessment, and state that the same benefit can be achieved at lower cost using secondary physical barriers and armed flight deck officers.

While some papers consider detection equipment to be one layer of a holistic screening process, others focus on technology-based detection as a multi-layered screening process in itself. Kobza & Jacobson (1996) published one of the first papers that looked specifically at the detection performance of a multi-layered system taking into account dependence between layers. They developed a probability model to quantify the effect of layer dependence on system-level true positive and false positive rates. Focussing on a two-layer system, they explore the importance of decision rules for combining the layers, namely case 1 where any device gives a system-level alarm and case 2 where any device gives a system-level clear. The authors signposted that technology development should be directed towards a situation of favourable dependence amongst the different layers of a combined system.

In Europe, the screening of passengers and their luggage is not influenced by any pre-screening or profiling. In the US, however, there is a history of pre-screening programmes (CAPPS, CAPPS II, Secure Flight, TSA Pre-Check), and hence there are many papers in the literature that examine the effectiveness of pre-screening. Since the outcome of pre-screening influences the subsequent screening at the airport, pre-screening can – to a certain extent – be considered as simply an additional layer of detection on the front end of multi-layered screening processes. Martonosi & Barnett (2006) used a simple mathematical model to examine arguments for and against pre-screening of passengers. Arguments for pre-screening are based on the assumption that terrorists are segregated with a high degree of effectiveness, and that the enhanced screening they undergo is far superior to that of regular screening. Arguments against pre-screening include the potential for terrorists to reverse-engineer the system and thereby thwart it. The authors conclude that neither point of view is fully convincing and suggest that improving the baseline security might be more beneficial than improving the pre-screening or enhanced screening steps.

Jacobson et al. (2001) estimate the joint conditional probability density functions for the different layers of a multi-layered screening process. They use heuristic algorithms to minimise the false alarm rate for a given detection rate. Babu et al. (2006) also developed a methodology for minimising false alarm rate for a given detection rate by randomly assigning passengers to different groups, where each group corresponds to a different combination of screening equipment. This work was extended by Nie et al. (2009) to include a pre-selection step to categorise passengers by perceived risk, but the approach does not consider dependence between layers nor different rules for combining outputs from each layer.

Nie, et al. (2012) examined how to best utilise selectee lanes intended for higher risk passenger at screening checkpoints. Assuming a passenger pre-screening system that effectively classifies passengers into several risk classes, they proposed a simulation-based queueing design framework that assigns passengers from different risk classes to the selectee lane based on how many passengers are already in the lane. They concluded that passenger checkpoint screening can be improved by incorporating a simulation-based queueing design model to assign passengers to the selectee lane.

Nie (2011) proposed a risk-based cost-effectiveness model where checked bags are classified into several risk classes. For a multiple-device screening system, the optimal sequence of the screening devices is determined, with the objective of minimising the expected cost per bag. Unlike many other works that only consider two devices, the model considers a four-layered screening process, whereby an alarm from any device will lead to a system-level alarm. In the conclusions, Nie signposts that other decision thresholds between one and n (where n is the number of layers) could be studied in future research, as could the dependency between layers.

Nie (2019) developed a cost model for a two-device screening system that addresses conditional dependence, with the objective of minimising the expected cost per bag. The model incorporates a joint probability density function, which is assumed to follow a bivariate normal probability density function. The authors also address the cost of ignoring conditional dependence and conclude that a consideration of correlation between devices can sometimes lead to a different choice of equipment as the optimal combination. The main conclusion is that the optimal expected cost per bag is higher when the correlation between devices is higher.

To summarise the relevant literature to date, one could say that many articles on multi-layered security screening are quite theoretical and/or focussed on cost-effectiveness, with little numerical values or experimental data concerning the actual detection performance. Many articles consider multi-layered screening in a qualitative sense, and those that consider it quantitatively often do not take into account the dependency, or correlation, between different layers of screening. Many of the earlier papers consider a two-layer screening arrangement. However, most multi-layered screening processes around the world today have more than two layers. For example, if an alarm is not resolved by multiple-layers of screening technology, the final decision is often taken by a human screener – who is essentially just another layer of detection with certain strengths and weaknesses.

In this work, a numerical model of a multi-layered screening process is created, and a computer simulation is performed to predict and compare the behaviour of

different systems and designs. These technology-agnostic simulations are intended to provide insight into the key design parameters of multi-layered screening systems and guide the optimisation of future systems. Novel aspects of this work include the use of realistic profiles of alarm distributions based on experimental observations and a focus on the influence of correlation/orthogonality amongst the layers of screening.

The computer simulation helps elucidate a screening strategy that – compared to how screening is currently performed in most airports around the world – would significantly increase the system-level true positive rate with only a modest increase in system-level false positive rate. The strategy involves screening items multiple times regardless of whether they initially alarm or not, and then applying a threshold (e.g., two alarms out of three) to determine the system-level outcome (alarm or clear). The more orthogonal the screening layers, the better the performance of this strategy.

The intended application is aviation security, where different layers of detection equipment are required to detect many different kinds of threats on persons or in baggage, and the equipment exhibits a range of detection probabilities depending on how challenging each threat item is. However, the findings of this work are relevant for optimising any multi-layered screening process intended to detect multiple and diverse targets.

Problem description

Regulatory & operational considerations

Aviation security is a well-known situation where multi-layered screening processes are implemented. In Europe, the European Commission has established common rules in the field of civil aviation security aimed at protecting persons and goods from unlawful interference with civil aircraft. Regulation (EC) N°300/2008 of the European Parliament and of the Council lays down common rules and basic standards on aviation security and procedures to monitor the implementation of the common rules and standards. The legislation specifies performance requirements for the individual means of screening, and there are some provisions relating to how pieces of screening equipment may be combined. However, there are no quantitative requirements for the system-level performance of multi-layered screening processes. Similar regulatory approaches exist in other regions of the world.

On the operational side, aviation security measures at airports are subject to logistical boundary conditions that require a compromise to be found between security effectiveness, cost and passenger facilitation. For the screening of hold baggage, for example, a large airport might have to process in the order of 100,000 items of hold baggage each day, with an average screening time in the order of five seconds per bag. In terms of currently available commercial solutions, only X-ray technology has a high enough throughput to screen all bags. Other screening techniques, such as explosive trace detection (ETD) and hand search, are more time-consuming and are therefore reserved for items or persons that have generated alarms at previous screening stages.

Cascading screening (Case T3)

In most airports around the world, a system of cascading screening has been established in response to regulatory and operational requirements described above. It is noted that there are some situations in aviation security where a limited proportion of items/persons are screened on a continuous, random basis with additional layers even if an alarm is not generated at the previous layer. However, in the majority of cases, the cascading system described below is utilised.

The cascading screening process is that all items/persons are screened at the first layer (level one) and items/persons generating an alarm are subject to follow-up screening at level two. Items/persons generating alarms at level two are screened at level three, and so on. The process of cascading alarms to subsequent levels continues until the alarm is resolved, i.e., it is identified to the satisfaction of the screener as a true positive or a false positive (see Fig. 1).

A system-level alarm is generated if, and only if, all layers produce an alarm. On the other hand, if a layer generates a ‘clear’ output (i.e., no alarm), then previous alarms are deemed to have been overruled. This screening architecture has been referred to as “*any device gives a clear*” (Kobza & Jacobson 1996). Contrary to popular misconception, this screening architecture is *not* an implementation of the Swiss-cheese model (Wikipedia 2021). The Swiss-cheese model is based on re-screening of items even if they do not alarm, whereas the cascading model is based on re-screening *only* those items that do alarm. In this paper, we will refer to this cascading approach as Case T3, where T3 indicates a threshold of three individual alarms required (out of three) to generate a system-level alarm.

Cumulative screening (Case T2)

The cascading approach to multi-layered screening is not the only possibility. The idea of cumulative screening is to screen all items with multiple layers of detection, regardless of whether the initial result is alarm or clear, and then to apply a threshold for the cumulative number of alarms (e.g. two out of three) that defines when an item is considered suspicious or not. This approach is illustrated in Fig. 2.

Fig. 1 Schematic of a multi-layered screening process commonly used in aviation security. A cascading design is used, and a system-level alarm is generated if, and only if, all layers produce an alarm. We refer to this approach as Case T3

Fig. 2 Schematic of a multi-layered screening process based on cumulative screening. All items are screened multiple times regardless of whether the initial result is alarm or clear, and a threshold for the cumulative number of alarms defines when an item is considered suspicious or not. In this paper, we define Case T2 as a cumulate screening strategy where a system-level alarm is raised if two or more alarms (out of three) are generated

In this paper, we define Case T2 as a cumulate strategy where a system-level alarm is raised if two or more alarms (out of three) are generated. The idea of cumulative screening in aviation security is not new; it was mentioned as one possible strategy in a 2013 report on reducing false alarms in baggage screening [National Research Council 2013]. In this work, we delve deeper into cumulative screening by modelling realistic distributions of alarm probabilities over test items and by studying the effect of correlation/orthogonality between different screening layers.

In terms solely of security outcome, the cascading approach described in the previous section is actually equivalent to that obtained with the cumulative approach when the decision threshold for the cumulative approach is set to three alarms out of three. That is because in both approaches, a system-level alarm will be raised if, and only if, all three individual layers generate an alarm. The only difference is that fewer screening instances are performed in the cascading approach. We will see later, however, that a cumulative screening approach with a decision threshold of two alarms or more (out of three) is arguably a better system design.

Admittedly, a cumulative screening process is currently not feasible in civil aviation in situations of high passenger volume, due to the lack of high-throughput screening technologies. However, it is expected that data streams from detection equipment will become richer, and high-throughput technologies will be developed and made commercially available (e.g., machine-learning-based analysis of images, or X-ray cabinets with in-built trace detection). International regulators and airport representatives have recently advocated for the adoption of open architecture in aviation security (ACI-Europe 2020), which would also facilitate cumulative screening processes through enhanced data fusion and interoperability. When multiple, high-throughput screening options become available, a cumulative screening architecture will be logistically possible, and the approach in this paper is intended to support the design of such systems in the future.

Aim of this paper

We have outlined the regulatory and operational constraints on a multi-layered security screening process for aviation security, and we have described two possible system designs that could be adopted for a multi-layered security screening process (i.e., Case T3 and Case T2). Another important aspect is that security equipment is tasked with detecting a vast range of different possible threats,

and different kinds of threats are detected with varying degrees of efficacy. This means that simply considering the average alarm rate is not sufficient for a proper understanding of the system; we also have to consider the complementarity of each layers' strengths and weaknesses.

With all these issues in mind, this paper attempts to answer the following questions:

- For a given set of input parameters for the individual layers, what is the system-level performance of a multi-layered security screening process, based on the cascading architecture that is currently implemented in most airports around the world?
- How does the degree of correlation between the different layers quantitatively influence the system-level performance?
- How do cumulative screening approaches compare with the cascading model, and can this insight be leveraged to design better multi-layered screening processes in the future?

Supplementary material

In this work, the performance of multi-layered screening processes is modelled using the Python programming language. The software code is provided in the form of a Jupyter Notebook (Project Jupyter 2021) as supplementary material for this article. Jupyter Notebook is an application commonly used for creating and sharing documents that contain live code, equations, visualizations, and explanatory text. At the time of writing, the default environment after installation of Jupyter using Anaconda (Anaconda, Inc. 2021) contained all the necessary libraries to run the notebook, with the exception of Jupyter widgets, which provide interactivity. Instructions for installing the *ipywidgets* package can be found in the notebook itself.

The accompanying Jupyter Notebook contains an interactive dashboard where users can specify average alarm rates of each layer of a multi-layered system for both true positives and false positives. The degree of correlation between the layers can be tuned, as can the way the desired degree of correlation is applied across the layers (e.g., left-to-right, right-to-left or randomly). Users can also experiment with different distributions of alarm rates (e.g., simple, binary, J-shaped, U-shaped). There is also an option to randomly assign any of the four J-shaped distributions (see Sect. 3) and performs many iterations to determine the average behaviour (a sort of Monte Carlo simulation). The resulting performance of the system is displayed as a receiver-operator-characteristic (ROC) plot that indicates the system-level performance for various thresholds of cumulative alarms.

Modelling individual layers

The performance of security detection equipment is regulated and tested primarily in terms of the average detection rate (i.e., proportion of true positives) and average false alarm rate (proportion of false positives) observed over a range of different test items. The average alarm rate tells us nothing, however, about the probability of alarm on any *single* item. For any single item, the probability of alarm could be anything between 0 and 100%. Figure 3 shows four examples of different distributions of alarm probabilities over a range of hypothetical test items.

There is limited data available on the distribution of alarm probabilities of commercial security screening equipment because this information is not required during type testing. It is also very resource-intensive to collect this information, as it requires many repeats (e.g., 20 or more) for each test item, to have reasonably precise alarm rates per test item. We obtained three datasets in which numerous repeats were made of each individual test item (threat items and benign items) on a variety of commercial detection equipment employing various sensing technologies. These datasets are described in Table 1, and more information about the equipment is available in the references provided.

The dataset DS-1 comprises false positive rates for five different models of liquid explosive detection systems (LEDS). The dataset DS-2 comprises true positive rates for four different models of LEDS. The commercial LEDS equipment used in datasets DS-1 and DS-2 includes a variety of sensing technologies, including transmission X-ray, Raman spectrometry, wideband radio frequency and infrared. The dataset, DS-3, comprises true positive rates for 14 different commercially available models of explosives detection systems (EDS), which is the term used in aviation security for X-ray equipment that can automatically detect explosives in baggage. Since there were only 20 runs per test (2,800 total screenings overall), the results in DS-3 were pooled into a single distribution, giving a total of 10 separate distributions of alarm rates (half for true positive rates and half for false positive rates).

To create realistic models of the performance of individual layers of screening equipment, a family of four empirical J-shaped distributions was developed to describe the behaviour of the experimental data over a range of average alarm rates. The distributions are called J-shaped because when looking at

Fig. 3 Hypothetical distributions of alarm rates over a range of test items, each with the same average alarm rate of 80%

Table 1 Overview of three datasets used to develop empirical models of alarm rate distributions

Dataset	DS-1	DS-2	DS-3
Equipment	5 different models of liquid explosives detection systems (LEDS)	4 different models of liquid explosive detection systems (LEDS)	14 different models of explosive detection systems (EDS)
Type of alarm rates	false positive rates	true positive rates	true positive rates
Original purpose of dataset	investigate influence of container filling level on detection	investigate detection of acids by LEDS	verify detection of explosive simulants
Reference	(Dzhongova et al. 2017)	[not published]	(Vahčić et al. 2019)
Unique test items	39	64	10
Repeats per test item	100	50	20
Total number of screenings	over 65,000	12,800	2,800

the probability density function (PDF), the distribution looks like a square-bend, J-hook (see Fig. 4). The equations are given below (Eq. 1 to Eq. 4), where r is the alarm rate from 0 to 100% and $P(r)$ is the probability of each alarm rate. These equations are valid for weighted average alarm rates, $\bar{r}$, between 0 and 50%. For weighted average alarm rates between 50 and 100%, the function is calculated for $(1-\bar{r})$ and then inverted along the r -axis.

J-shaped_1:

$$P(r) = \begin{cases} 1, & r = 0.00 \\ \alpha + \frac{e^{-30r}}{9}, & 0.01 \leq r \leq 0.20 \\ \alpha, & 0.21 \leq r \leq 1.00 \end{cases} \quad (1)$$

J-shaped_2:

$$P(r) = \begin{cases} 1, & r = 0.00 \\ 0.25 \cdot e^{-70r}, & 0.01 \leq r \leq 0.25 \\ 0, & 0.26 \leq r \leq 0.99 \\ \alpha, & r = 1.00 \end{cases} \quad (2)$$

J-shaped_3:

$$P(r) = \begin{cases} 7 \cdot P(0.01), & r = 0.00 \\ \alpha \cdot r^{-1.3}, & 0.01 \leq r < 0.79 \\ P(1.00)/60, & 0.80 \leq r < 0.99 \\ 560, & r = 1.00 \end{cases} \quad (3)$$

J-shaped_4:

$$P(r) = \begin{cases} 1.00, & r = 0.00 \\ 0.2, & r = 0.01 \\ \alpha, & 0.02 \leq r < 0.99 \\ 28 \cdot \alpha, & r = 1.00 \end{cases} \quad (4)$$

For each equation, the coefficient, α , is determined iteratively so that the weighted average alarm rate, $\bar{r}$, equals the desired value:

$$\bar{r} = \sum_{r=0.00}^{1.00} P(r) \cdot r \quad (5)$$

The four J-shaped distributions of alarm rates are illustrated in Fig. 4, both in terms of probability density function (top row) and the cumulative distribution function (bottom row). The cumulative distribution function is more helpful to perceive the differences between the distributions.

To check the four empirical equations are a good match with the ten experimentally observed profiles, J-shaped distributions (solid red lines) are superimposed on the experimental data (dashed blue lines) in Fig. 5. The Pearson correlation coefficients (ranging from 0.89 to 1.00) are provided above each plot.

Modelling multi-layered systems

In the previous section, individual layers of detection were modelled using distributions of alarm rates (true positives and false positives) that closely match experimental observations of the alarm rates of commercial threat-detection equipment. In this section, we describe how these layers are combined to build a model of a multi-layered detection system.

Two pieces of equipment are *correlated* if they tend to detect the same items and miss the same items. A lack of correlation (either positive or negative) is referred to as *uncorrelated*. Negative correlation is sometimes referred to as *orthogonality*. In the case of multi-layered screening, orthogonal layers of screening can be understood to refer to layers with complementary strengths and weaknesses that are staggered, or offset, as much as possible across a range of test items. Two kinds of equipment with different sensing technology, such as millimetre-wave body scanners and ion mass spectrometry, will be more orthogonal than, say, two X-ray baggage scanners produced by different manufacturers.

In this work, a multi-layered screening process is represented as a two-dimensional numerical matrix; one dimension represents a representative sample of different items to be screened and the other the different detection layers of the system. Each element in the matrix describes the alarm probability for a given item and layer of detection. The system-level alarm rate, R_{sys} , for Case T3 is the average over N test items of the probability that each item alarms on levels L_1 , L_2 and L_3 . For a three-level system that is:

Fig. 4 Examples of the four J-shaped distributions of alarm rates, for an average alarm rate of 80%. The top row is the probability density function (PDF), and the bottom row is the cumulate distribution function (CDF)

Fig. 5 Comparison of experimental data (dashed blue lines) and empirical models (solid red lines) for ten distributions of alarm rates. The alarm rates are plotted as cumulative distribution functions (CDFs), and the Pearson correlation coefficient between experimental data and model is indicated above each subplot

$$R_{SYS} = \frac{1}{N} \sum_{i=1}^N (r_{L1i} \times r_{L2i} \times r_{L3i}) \quad (6)$$

The correlation amongst different layers of a multi-layered screening system is maximised when the alarm rates for each layer increase monotonically. This can be achieved programmatically simply by sorting the alarm distribution of each layer in ascending order.

The alarm rates in different layers of screening are uncorrelated when the alarm rates are distributed randomly across the test items, for each layer. This can be achieved programmatically using pseudo-random number generators.

The situation of maximum orthogonality of a multi-layer screening system is, by definition, the arrangement such that the average alarm rate, R_{SYS} , is minimised. For a system comprising i layers, with N test items considered for each layer, there are up to $(N!)^{i-1}$ possible unique values of R_{SYS} . We use a simple heuristic algorithm to maximise the orthogonality of the layers' alarm rates as follows:

1. the N alarm probabilities in each layer are sorted in ascending order;

- the mean alarm rate, $\bar{r}_i$, is calculated for each layer, i ;
- for each layer, the sorted alarm rates are shifted by $\left(N \cdot \frac{\bar{r}_i}{\sum_i \bar{r}_i}\right)$, where N is the number of test items.

The effect of this algorithm is illustrated in Fig. 6 for a three-layered screening process.

A function was developed to tune the degree of orthogonality between different screening layers. First, a *correlation factor*, ϕ , which ranges from -1 to 1, is defined. A correlation factor of -1 corresponds to the situation of maximum orthogonality, a factor of 0 corresponds to randomly arranged layers, and a factor of +1 corresponds to the maximum correlation. The correlation factor is linearly scaled to the system-level alarm rates at these three points, resulting in the following piecewise function (Eq. 7), for the system-level alarm rate, R_{SYS} , as a function of the correlation factor, ϕ :

$$R_{SYS}(\Phi) = \begin{cases} \Phi \times (R_{SYS}(0) - R_{SYS}(-1)), & \Phi < 0 \\ \Phi \times (R_{SYS}(+1) - R_{SYS}(0)), & \Phi > 0 \end{cases} \quad (7)$$

where:

$R_{SYS}(-1)$ = system-level alarm rate, for situation of maximum orthogonality;

$R_{SYS}(0)$ = system-level alarm rate, when alarm rates in each layer are uncorrelated;

$R_{SYS}(+1)$ = system-level alarm rate, for situation of maximum correlation.

In the accompanying software, the tuning of the correlation/orthogonality of a multi-layered screening process is achieved with the following steps:

Fig. 6 Illustration of how the algorithm to maximise orthogonality of a multi-layer system works; (a) alarm rates for each layer are sorted in ascending order, then (b) shifted across the test items in proportion to their average alarm rate

1. determine the target value of the system-level alarm rates, R_{SYS} , for a given value of the correlation factor, ϕ , using Eq. 7;
2. if ϕ is less than zero then apply the function to maximise the orthogonality, otherwise for ϕ greater than zero apply the function to maximise the correlation;
3. a loop is executed to randomly swap the order of pairs of values in the probability distribution. This will cause R_{SYS} to shift towards $R_{SYS}(0)$, i.e. the situation of randomly distributed alarm rates. The execution of the loop is stopped as soon as the target value of is R_{SYS} met.

The influence of the correlation factor, ϕ , on the system-level alarm rate, R_{SYS} , for different mean alarm rates and different shapes of alarm distributions is illustrated in Fig. 7, for a Case T3 situation. We see that if the alarm rates are closely distributed around a central value (i.e., the binomial distribution in Fig. 7a), then correlation has little to no influence on the system-level alarm rate. On the other hand, for distributions with extreme values (J-shaped and binary in Fig. 7b and c respectively), R_{SYS} is significantly impacted by the degree of correlation between the layers. We also note that although a three-layer system with alarm rates of [80%, 80%, 80%] and [70%, 80%, 90%] have the same simple average, the performance of the latter will be lower (and the difference increases with increasing correlation factor).

Results and discussion

We started by modelling individual layers of detection based on distributions of alarms that closely match experimental observations. We then combined these layers, and developed functions to describe and adjust the correlation between the different layers. In this section, we explore the system-level performance of different multi-layered screening processes. The numerical examples given below are for illustrative purposes only and should not be taken as estimates of real-world

Fig. 7 Influence of the correlation factor, ϕ , on the system-level alarm rate, R_{SYS} , for different mean alarm rates and different shapes of alarm distributions for Case T3, i.e., a system alarm is raised when alarms are raised on all three individual layers

Fig. 8 System-level true positive rates for a three-layer screening process for three different situations of correlation between layers: orthogonal (left), uncorrelated (middle) and correlated (right). The average true positive rate of each layer is 80%. The distribution of alarm rates over the test items on the top row (a) is the 'binary' distribution, i.e., only 0% or 100% values, and on the bottom row (b) the more realistic 'J-shaped' distribution

screening processes. Users can experiment with different starting values using the interactive Jupyter Notebook provided as supplementary material to this article.

Cascading screening (Case T3)

We first consider a three-layer screening process where each layer has an average true positive rate (TPR) of 80%. An item is considered suspicious if, and only if, it alarms on all three layers. The system-level TPR for such a screening process is shown in Fig. 8, for three different cases of correlation: orthogonal (left), uncorrelated (middle) and correlated (right). In Fig. 8a, the 'binary' distribution of alarm rates is modelled, i.e., only values of exactly 0% or 100% are permitted. In Fig. 8b, the more realistic 'J-shaped' distribution of alarm rates is used (i.e., the distribution based on experimental observations). We see in Fig. 8b that the system-level TPR varies from 43% to 69%, depending on the degree of correlation/orthogonality,

which is quite a very significant variation in performance. Orthogonal layers have a lower system-level TPR than correlated layers, so from this point of view, orthogonality is detrimental.

We now turn our attention to the situation for false positive rates (FPR). We consider a three-layer screening process where each layer has an average FPR of 20%. Again, an item is considered suspicious if, and only if, it alarms on all three layers. The system-level false alarm rate for such a screening process is shown in Fig. 9, for three different cases of correlation/orthogonality, namely orthogonal (left), uncorrelated (middle) and correlated (right). In Fig. 9a, the ‘binary’ distribution of alarm rates is modelled, i.e. only values of exactly 0% or 100% are permitted. In Fig. 9b, the more realistic ‘J-shaped’ distribution of alarm rates is modelled. We see in Fig. 9b that the system-level FPR varies from 0% to 10%, depending on the degree

Fig. 9 System-level false positive rates for a three-layer screening process for three different situations of correlation between layers: fully orthogonal (left), fully correlated (right) and random/uncorrelated (middle). In each scenario, the average false alarm rate of each layer is 20%. The distribution of alarm rates over the test items on the top row (a) is the ‘binary’ distribution, i.e. only 0% or 100% values, and on the bottom row (b) the ‘J-shaped’ distribution

of correlation/orthogonality. Orthogonal layers have a lower system-level FPR than correlated layers, so from this point of view orthogonality is beneficial.

It is clear that the degree of correlation/orthogonality between each layer of screening can have a dramatic effect on the system-level performance, both for true positives and false positives. For the Case T3 design, the ideal system would be comprised of screening layers that are orthogonal for true negatives but correlated for true positives.

Little is known about the correlation amongst different commercial security screening equipment, because system-level correlation it is neither regulated nor tested at the present time. We can, however, make some inferences from airport operations. In the case of hold baggage screening, 15% to 30% would seem to be a reasonable estimate of operational false alarm rates. Personal communication with subject-matter experts from aviation security authorities in Europe indicates that the false alarm rate after three layers of hold baggage screening in airports is less than 1%. Therefore, we conclude that current layers of security screening are not correlated for false alarms. At best, they are uncorrelated, and they may even be somewhat orthogonal. This is not surprising, as the screening processes must clear billions of false alarms that occur year after year all over the world.

What about the degree of correlation for true positives? This is something of a known unknown. From a science and engineering point of view, one would expect the degree of correlation between layers to be similar for true positives and false positives. This is because the correlation is influenced primarily by i) the list of targets to be detected by each layer, and ii) the sensing technology employed by each layer of detection, both of which are the same for both true positives and false positives. For a Case T3 design, the idealised situation of correlated true positives and orthogonal false positives has, to say the least, not been demonstrated. In our example, if the true positives have a similar degree of correlation to that of the false positives, then a three-layered system in which each layer has a detection rate of 80% would deliver a system-level detection rate between 43% and 51%, for a correlation factor between -1.0 and 0.0, respectively.

Cumulative screening (Case T2)

We turn now to cumulative screening and Case T2, i.e., two or more alarms out of three generate a system-level alarm. We consider again a three-layered screening process where each screening layer has an average TPR of 80% and an average FPR of 20%. Similar to the previous section, we consider three scenarios where TPR and FPR are (a) orthogonal, (b) uncorrelated, and (c) correlated. Unlike the previous section, where we only considered the system-level performance when all three layers generated an alarm, now we also determine the overall performance for varying alarm thresholds, T , of zero, one, two and three alarms out of three. The results are shown in Fig. 10 in the form of receive-operator-characteristic (ROC) plots, where each point represents a different threshold for the number of individual alarms required to generate a system-level alarm.

Fig. 10 ROC plots for three-layered screening systems for different thresholds, T , of alarms required to raise a system-level alarm, and for (a) orthogonal, (b) uncorrelated and (c) correlated systems. Values in parentheses are the true positive and false positive rates for each threshold, and N is the average number of screenings per bag (described in Sect. 5.4)

We can see from Fig. 10 that the more orthogonal the screening layers are, the better the overall performance of Case T2, i.e., a threshold of $T \geq 2$. For a perfectly orthogonal system, the ‘two alarms out of three’ strategy would achieve a system-level TPR of around 97% with a FPR of around 3%, in this example. An uncorrelated system would yield a TPR of around 91%, with a FPR of around 9%. Compared to the current approach of requiring all layers to alarm in order to be considered a threat (i.e., $T = 3$), cumulative screening with $T \geq 2$ with orthogonal layers yields a dramatic increase in TPR (approximately double) for a modest increase in FPR of a few percentage points.

When screening with the $T \geq 2$ approach (i.e., Case T2), the more orthogonal a system is, the higher the system-level TPR and the lower the FPR. We note that orthogonality was detrimental to TPR using the cascading strategy (effectively $T = 3$) but is beneficial using cumulative strategy with $T \geq 2$. The reason is clear when looking at the ROC plots in Fig. 10. The ROC data points form an envelope that “expands” towards the top left corner [TPR = 100%, FPR = 0%] with increasing orthogonality, but the individual point corresponding to the $T = 3$ strategy (i.e., all layers have to alarm) moves closer to the bottom-left corner [TPR = 0%, FPR = 0%].

Resilience of cumulative screening

In this section, we highlight how the cumulative screening strategy (Case T2) is much more resilient to a decrease in detection rate in any individual layer in the system, compared to the cascading screening approach (Case T3).

We model a three-layered screening process with average TPRs of [95%, 80%, 80%] and average FPRs of [15%, 15%, 15%], both with a correlation factor, ϕ , of -0.75, i.e. fairly orthogonal. The results are shown in Fig. 11a. Next, we lower the TPR of the first layer from 95% to 50%, while all other parameters are the same. The results are shown in Fig. 11b. We see that the system-level TPR for Case T3 drops from 57% to 21%, while the system-level TPR for Case T2 drops from 98% to 89%. In other words, the Case T2 is much more resilient to fluctuations

Fig. 11 Comparison of two multi-layered systems with detection rates of (a) 95%, 80%, and 80%, and (b) 50%, 80% and 80%. The corresponding ROC plots in the right-most column show a reduction in overall detection rate from 57% to 21% for a decision threshold, $T=3$, and 98% to 89% for $T \geq 2$

of detection rate of individual layers. It is something of a veridical paradox that a three-layer screening process in which one of the layers has a detection rate of only 50% can still deliver a system-level detection rate of around 90%. The only proviso is that the different layers of detection are fairly orthogonal.

The resilience of an orthogonal, cumulative-screening approach could help solve a long-standing conundrum in aviation security. Under current regulations and screening strategies, all screening layers have to meet all the detection requirements for a given application. A piece of detection equipment that is very strong in one area but not others will not be approved unless it is combined with another piece of equipment and presented for testing against the full range of threats as a single, black-box system. Moving to a cumulative screening approach would open the door to innovative screening equipment that might not necessarily detect the entire range of threats on its own, but – when combined with other, suitably orthogonal technologies – can help deliver a resilient, multi-layered screening process with a high system-level performance.

Implementing cumulative screening

From the logistical point of view, airport operators may be concerned that a replacing a cascading screening system with a cumulative screening system might impact negatively on operational efficiency, as it involves rescreeing items/persons even if they do not alarm at the first layer of screening. However, implementing cumulative screening with, for example, a three-layer system, does not mean that three times as many screening instances are required. Some savings are available. For example, if the threshold is two alarms out of three, and the first two layers produced no alarms, then even if the last layer generates an alarm, the threshold of two alarms out of three will not be reached. Therefore, the third layer of screening can be omitted. Conversely, if two alarms have already been generated, the item has already reached the threshold to be considered suspicious and the third layer can again be omitted.

Using this logic, the minimum required number of scans for each permutation of outcomes and for each threshold of cumulative screening is calculated in the accompanying software. Weighting these values by the frequency of each permutation of outcomes gives us a single number that describes the average number of screenings per bag (indicated in Fig. 10 for each threshold by the symbol N).

For a three-layered system with a 15% false positive rate on each layer, our analysis shows that 'two out of three' cumulative screening will require around 2.29 screenings per bag, compared to 1.16 screenings per bag for screening as currently implemented. In other words, cumulative screening would require around double the number of screening instances.

Furthermore, it should be emphasised that a pre-requisite for implementing cascading screening is the availability of sufficient, high-throughput screening technologies that are orthogonal with one another. Examples of these might include X-ray baggage scanners with 'on-the-fly' trace detection of explosive vapours, or automated analysis of images using computer vision and machine learning. Cumulative screening will also require a means to track screening results of individual items as they traverse through the multiple layers of screening.

The cost of ignoring correlation

If a system designer ignores the potential correlation/orthogonality of alarms between the different layers of a multi-layered screening process, then the system designer is assuming – either implicitly or explicitly – that alarm rates for the different layers are all randomly distributed across the range of test items. In this section, we investigate the potential difference between the expected and actual system performance if a system designer ignores the correlation/orthogonality between different layers.

To have a measure of both the true positive rate and true negative rate of a screening system, and to compare the behaviour of different decision thresholds, we describe the performance of each system using a single metric called the Euclidean 'distance to perfect', or EDTP. On an ROC plot, EDTP is simply the distance

Fig. 12 The Euclidian distance-to-perfect (EDTP) parameter, used to compare the overall performance of different systems and different system-thresholds

between a data point and the point corresponding to a perfect score, i.e., a true positive rate of 100% and a false positive rate of 0%. The concept is illustrated in Fig. 12.

A simulation was performed of a three-layered screening process where each layer has an average true positive rate of 80% and an average false positive rate of 20%. The correlation factors for true positives and false positives (ϕ_{TP} and ϕ_{FP} respectively) were both swept incrementally from -1.0 to +1.0.

In Fig. 13a, we see a heatmap of the EDTP for a system where the decision rule is that two or more alarms generate a system-level alarm (Case T2). The best result (point A) is obtained when the different layers are as orthogonal as possible for both true positives and false positives. Where the alarms are uncorrelated between layers is indicated by point B, and the worst-case scenario is point C.

In Fig. 13b, we see a heatmap of the EDTP where the decision rule is that three alarms are needed to generate a system-level alarm (Case T3). In this case, the best result (point D) is obtained when the different layers are as orthogonal as possible for false positives, and as correlated as possible for true positives. Where the alarms are uncorrelated between layers is indicated by point E, and the worst-case scenario is point F.

Fig. 13 Euclidian distance to perfect (EDTP) as a function of correlation factor, ϕ , for true positives (TPs) and false positives (FPs) for a three-layer screening process (each layer has an average true positive rate of 80% and false positive rate of 20%). For a decision threshold, $T \geq 2$, the best system is orthogonal TPs and FPs, whilst for $T = 3$ it is orthogonal FPs and correlated TPs. Annotated points are referred to in Table 2

The performance of the best, random and worst cases are described in Table 2, for two decision rules ($T \geq 2$ and $T = 3$) for the simulated three-layer screening process, and each situation corresponds to the points A to F in Fig. 13. For Case T2, a system designer who ignores correlation may calculate an expected average performance of [FPR 10%, TPR 90%], but the actual performance could be [FPR 2%, TPR 98%] in the best-case scenario, or [FPR 20%, TPR 80%] in the worst-case scenario. For Case T3, a system designer who ignores correlation may calculate an expected average performance of [FPR 1%, TPR 51%], but the actual performance could be [FPR 0%, TPR 69%] in the best-case scenario, or [FPR 10%, TPR 42%] in the worst-case scenario.

Conclusions

In this paper, a visually intuitive analysis of multi-layered screening processes is presented. The analysis is based on a software model that is provided as supplementary material. Novel aspects of this work include the use of realistic profiles of alarm distributions based on experimental observations of commercial security equipment, and a focus on the influence of correlation/orthogonality amongst the layers of screening. We also compare and contrast the performance of two different system designs: (i) cascading screening – which is commonly employed in aviation security today, and (ii) cumulative screening – which could be employed where orthogonal, high-throughput screening solutions are available.

The results show that a cumulative screening architecture can outperform a cascading one, yielding a significant increase in system-level true positive rate for only a modest increase in false positive rate. A cumulative screening process is also more resilient to weaknesses in the individual layers. Implementing cumulative screening based on a decision threshold of two or more alarms out of three would require approximately twice as many screening instances compared to a comparable cascading process. It would also require a means to track screening results of individual items/persons as they traverse through the multiple layers of screening.

The performance of a multi-layered screening process using the current cascading approach is maximised when the false positives are orthogonal across the different layers and the true positives are correlated. The system-level performance of

Table 2 Summary of best and worst performances of a three-layer screening process using extreme values of correlation/orthogonality, for two decision thresholds: $T \geq 2$ and $T = 3$. Each layer has an average detection rate of 80% and a false alarm rate of 20%. Points A to F are annotated in Fig. 13

parameter	$T \geq 2$			$T = 3$		
	best [A]	random [B]	worst [C]	best [D]	random [E]	worst [F]
ϕ_{FAR}	-1.0	0.0	1.0	-1.0	0.0	1.0
ϕ_{DR}	-1.0	0.0	1.0	1.0	0.0	-1.0
FPR	2%	10%	20%	0%	1%	10%
TPR	98%	90%	80%	69%	51%	42%
EDTP	0.03	0.15	0.28	0.31	0.51	0.59

a cumulative screening process, on the other hand, is maximised when both false positives and true positives are as orthogonal as possible. An important conclusion, therefore, is that the regulation and testing of security equipment should adopt systems-based thinking, as equipment-based thinking alone is insufficient to properly understanding of the overall performance of a multi-layered screening process. Another conclusion is that the aviation security community should develop high-throughput detection technologies that are orthogonal to those in use today – with a view to combining them in a cumulative screening approach. In this way, a step-change improvement in the system-level detection performance of multi-layered screening processes can be achieved.

In terms of future research, we point to the fact that the degree of correlation amongst different commercial screening equipment is currently not tested. This is a ‘known unknown’ that should be addressed, perhaps by developing a sampling procedure involving representative test articles (both threat and benign) that can be analysed by diverse categories of screening technologies.

In summary, optimising the system-level performance of a multi-layered screening process requires (i) knowledge of the degree of correlation of alarms between layers, and (ii) and judicious selection of layers depending on the decision rule for combining multiple screening results.

Funding This work was funded under the Horizon 2020 Framework Programme for Research and Innovation.

Code availability The software model is available as a Jupyter Notebook (supplementary material).

Declarations

Conflicts of interest The author declares no conflicts of interest.

Disclaimer The views expressed in this article are those of the author and may not be regarded as the position of European Commission.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- ACI-Europe (2020) Regulators and airport operators join forces to promote open architecture in airport security systems. Retrieved from <https://www.aci-europe.org/media-room/268-regulators-and-airport-operators-join-forces-to-promote-open-architecture-in-airport-security-systems.html>. Accessed 12 Apr 2021

- Anaconda, Inc. (2021) Anaconda | The world's most popular data science platform. Retrieved from <https://www.anaconda.com/>. Accessed 30 Mar 2021
- Babu VLL, Batta R, Li Lin L (2006) Passenger grouping under constant threat probability in an airport security system. *Eur J Oper Res* 168(2):633–644. <https://doi.org/10.1016/j.ejor.2004.06.007>
- Dzhongova E, Anderson D, de Ruiter J et al (2017) False alarm rates of liquid explosives detection systems. *J Transp Secur* 10:145–169. <https://doi.org/10.1007/s12198-017-0184-7>
- Jackson BA, LaTourrette T (2015) Assessing the effectiveness of layered security for protecting the aviation system against adaptive adversaries. *J Air Transp Manag* 48:26–33. <https://doi.org/10.1016/j.jairtraman.2015.06.009>
- Jacobson SH, Kobza JE, Easterling AS (2001) A detection theoretic approach to modeling aviation security problems using the knapsack problem. *IIE Trans* 33(9):747–759. <https://doi.org/10.1023/A:1010945832179>
- Johnston RG (2010) Lessons for layering. Retrieved from <https://www.asisonline.org/security-management-magazine/articles/2010/01/lessons-for-layering>. Accessed 3 Mar 2021
- Kobza JE, Jacobson SH (1996) Addressing the dependency problem in access security system architecture design. *Risk Anal* 16(6):801–812. <https://doi.org/10.1111/j.1539-6924.1996.tb00831.x>
- Lee AJ, Nikolaev AG, Jacobson SH (2008) Protecting air transportation: a survey of operations research applications to aviation security. *J Transp Secur* 1:160–184. <https://doi.org/10.1007/s12198-008-0014-z>
- Martonosi SE, Barnett A (2006) How effective is security screening of airline passengers? *Interfaces* 36(6):545–552. <https://doi.org/10.1287/inte.1060.0231>
- National Research Council (2013) Engineering aviation security environments—reduction of false alarms in computed tomography-based screening of checked baggage. The National Academies Press, Washington, DC. <https://doi.org/10.17226/13171>
- Nie X (2011) Risk-based grouping for checked baggage screening systems. *Reliab Eng Syst Saf* 96(11):1499–1506. <https://doi.org/10.1016/j.ress.2011.06.011>
- Nie X (2019) The impact of conditional dependence on checked baggage screening. *Eur J Oper Res* 278(3):883–893. <https://doi.org/10.1016/j.ejor.2019.04.034>
- Nie X, Batta R, Drury CG, Lin L (2009) Passenger grouping with risk levels in an airport security system. *Eur J Oper Res* 194(2):574–584. <https://doi.org/10.1016/j.ejor.2007.12.027>
- Nie X, Parab G, Batta R, Lin L (2012) Simulation-based Selectee Lane queueing design for passenger checkpoint screening. *Eur J Oper Res* 219(1):146–155. <https://doi.org/10.1016/j.ejor.2011.12.012>
- Project Jupyter (2021) Project Jupyter | Home. Retrieved from <https://jupyter.org/>. Accessed 30 Mar 2021
- Stewart MG, Mueller J (2013) Terrorism risks and cost-benefit analysis of aviation security. *Risk Anal* 33:893–908. <https://doi.org/10.1111/j.1539-6924.2012.01905.x>
- Stewart MG, Mueller J (2018) Risk and economic assessment of U.S. aviation security for passenger-borne bomb attacks. *J Transp Secur* 11:117–136. <https://doi.org/10.1007/s12198-018-0196-y>
- Vahčić M, Anderson D, Ruiz Osés M, Rarata G, Diaconu G (2019) Development of Inert, polymer-bonded simulants for explosives detection systems based on transmission X-ray. *Molecules* 24(23):4330. <https://doi.org/10.3390/molecules24234330>
- Wikipedia (2021) Swiss cheese model. Retrieved from https://en.wikipedia.org/wiki/Swiss_cheese_model. Accessed 10 May 2021
- Wright PD, Liberatore MJ, Nydick RL (2006) A Survey of Operations Research Models and Applications in Homeland Security. *INFORMS J Appl Anal* 36(6):514–529. <https://doi.org/10.1287/inte.1060.0253>