Agent-based vulnerability assessment at airport security checkpoints

A case study on security operator behavior

Janssen, Stef; van den Berg, Arjan; Sharpanskykh, Alexei

DOI
10.1016/j.trip.2020.100139

Publication date
2020

Document Version
Final published version

Published in
Transportation Research Interdisciplinary Perspectives

Transportation Research Interdisciplinary Perspectives 5 (2020) 100139

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: https://www.journals.elsevier.com/transportation-research-interdisciplinary-perspectives

Agent-based vulnerability assessment at airport security checkpoints: A case study on security operator behavior

Stef Janssen *, Arjan van den Berg, Alexei Sharpanskykh

Delft University of Technology, Kluyverweg 1, 2629 HS Delft, Netherlands

ARTICLE INFO

Article history:
Received 21 December 2018
Received in revised form 14 November 2019
Accepted 18 May 2020
Available online 11 June 2020

Keywords:
Agent-based modelling
Security
Vulnerability
Human decision making
Human performance

ABSTRACT

Despite enormous investments in airport security, terrorists have been able to find and exploit vulnerabilities at security checkpoints. Existing vulnerability assessment methodologies struggle with accounting for human behavior, and agent-based modelling forms a promising technique to overcome this limitation.

This paper investigated how the decision-making and performance of human operators can be taken into account while assessing vulnerability at an airport security checkpoint. To this end, an agent-based model was designed, in which the performance of security operators was modelled using a functional state model, while decision making was modelled using decision field theory. Passengers and an attacker that brings a weapon to the security checkpoint were also explicitly modelled as agents. Simulation results indicate that the highest skilled operators outperformed their lowest skilled counterparts on analyzing X-ray images, but performed worse on both searching luggage and performing patdowns. Furthermore, results showed that a high focus on speed of security operators leads to a decrease in luggage searches and therefore increased vulnerability.

More work is needed to calibrate and validate the simulation results, but initial results are promising. The agent-based model can be used by airport regulators and managers to understand the workings of their security checkpoint better and ultimately to reduce vulnerabilities.

1. Introduction

Despite enormous investments in airport security, terrorists have been able to find and exploit vulnerabilities at security checkpoints. In the years after 9/11, aviation has been targeted by several bombing attempts (BBC News, 2006; Edmunds, 2010; Burns, 2010), such as the shoe bomber (CNN, 2001). Each of those attempts exploited new vulnerabilities and bypassed the security checkpoint successfully. It is only after such an attempt that new regulations and procedures are developed to address the exploited weakness in the security checkpoint. This reactive approach leaves airports vulnerable to innovating attackers. This problem is well recognized within the scientific literature, but developing a method that accurately assesses all vulnerabilities in a security checkpoint is a challenging task.

The security checkpoint is operated by security operators that constantly have to perform cognitive tasks, such as detecting illegal items on an X-ray image (IATA, 2012). These operators also continuously have to make decisions, such as the decision to confiscate a potential weapon or not. Empirical research has shown that security operators do not necessarily follow protocol, but regularly bend and break the rules (Kirschenbaum et al., 2012a; Kirschenbaum et al., 2012b; Kirschenbaum, 2013; Kirschenbaum, 2015). They commonly ignore potential threats and alarms are often processed as false. Furthermore, the performance of security operators is dependent on a variety of factors, of which cognitive task demands and personality are two examples. These human factors affect the performance of the checkpoint as a whole and additional vulnerabilities may emerge from their behavior. Therefore any method that aims to systematically identify all vulnerabilities in a security checkpoint should include these cognitive aspects in the analysis.

The objective of this work therefore is to understand how the decision-making and performance of human operators influence vulnerability at an airport security checkpoint. To this end, we employ an agent-based modelling approach to identify and quantify the vulnerabilities of two typical airport security checkpoint setups. The contribution of this work is twofold. First, we define a novel agent-based model to assess vulnerability, in which we specify security operators' behavior by combining two different cognitive models. The performance of security operators on different tasks in the checkpoint is modelled using the functional state model (Bosse et al., 2008), and their decision-making process is modelled using decision field theory (Busemeyer and Townsend, 1993). The developed model can easily be adapted to test future concepts of security checkpoints, such as X-ray operators working remotely. These types of experiments are hard to perform directly at airports, as it may interrupt security operations. Secondly, by performing experiments with the model, we generate new insights with

* Corresponding author.
E-mail address: s.a.m.janssen@tudelft.nl. (S. Janssen).

http://dx.doi.org/10.1016/j.trip.2020.100139

respect to vulnerabilities at the security checkpoint. Three types of experiments are performed: experiments related to operator performance, experiments related to operator decision making and experiments related to different airport security checkpoint setups.

This paper is structured as follows. First, related literature about existing vulnerability assessment methodologies and human performance and decision making is reviewed in Section 2. Then, the agent-based model that we developed for this work is described in Section 3 and calibrated in Section 4. Three experiments were performed with the model and are described in Section 5. The first experiment is used to understand the influence of security operator performance on vulnerability, and the second for understanding the influence of security operator decision making on vulnerability. The third experiment is then used to investigate the effect of using different security checkpoint setups on vulnerability. Finally, the work is concluded in Section 6.

2. Related work

This section provides an overview of existing vulnerability assessment techniques, with a special focus on agent-based modelling. Furthermore, existing human performance and decision making models are discussed.

2.1. Existing vulnerability assessment methodologies

Following the ISO (I. Guide, 73, 2009) standards, vulnerability is defined as "intrinsic properties of something resulting in susceptibility to a risk source that can lead to an event with a consequence". Most vulnerability assessment methodologies are interested in quantifying vulnerability as a value between 0 and 100%.

The most common method to estimate vulnerabilities is to consult security experts involved in the field. These experts know about the equipment, procedures, and performance of the operators. They use this knowledge to quantify vulnerabilities on a predefined scale. Unfortunately, expert elicitation has proven to perform poorly in parameter estimation when there are dependencies in the system (Cooke and Goossens, 2008), while these dependencies inevitably exist (Cole and Kuhlmann, 2012).

Wu and Mengersen provide an overview of models that were used to aid security policy and planning (Wu and Mengersen, 2013). The discussed models range from simple probabilistic models (Chawdhry, 2009), to Bayesian models (Babu et al., 2006; Nie et al., 2009) and a fuzzy model (Akgun et al., 2010). None of these models explicitly incorporate the behavior of security operators, and therefore lack the capability to understand how the decision-making and performance of human operators influence vulnerability at an airport security checkpoint.

Several other methods exist as well. A failure tree models events in the form of an event tree (Benner, 1975), in which the nodes of the tree are events and the branches of each node specify the possible outcomes of the event. For each event in the tree, the likelihood of the different outcomes must be estimated, and vulnerability is finally calculated based on the combination of these events. The reliability of these methods hinges on the accuracy of those estimates which are not always accurate (Stroeve et al., 2013). Furthermore, a failure tree is unable to accurately model a complex socio-technical environment such as a security checkpoint.

The scenario-based approach is an approach that tries to identify threat scenarios to which airport security is most vulnerable (Cole and Kuhlmann, 2012). This method starts with identifying all threat elements that together describe the airport environment. For each of those scenarios, the countermeasures are identified as well. This allows users to identify scenarios in which a limited number of countermeasures are available. This approach has two major limitations. First of all, the airport environment is reduced to a linear set of relations between threat elements and countermeasures. Second, this method aims to identify the threat scenarios to which the system is most vulnerable, but only identifies the number of countermeasures related to each threat scenario, without factoring in the effectiveness of each measure.

Using penetration testing, airport defenses are physically tested. The results from those tests can be used to give better vulnerability estimates and develop better protocols. The big advantage of penetration testing is that it does not require any assumptions about the complexity of the system nor the behavior of the security operators. This guarantees that the vulnerabilities found during these tests coincide with reality, and can be beneficial to calibrate or validate other methodologies. However, a problem with penetration testing that it is labor-intensive and thus expensive (Bennett, 2015; Airport data & contact information, 2018). Due to these problems, airports can only test their defenses on a small subset of security scenarios.

2.2. Vulnerability assessment using agent-based modelling

Over the last few years, agent-based modelling has gained interest as an approach to assess airport security (Wilson et al., 2006). This is proposed as a risk management methodology called AbSRiM by Janssen et al. (2019b). In this methodology, security risks are assessed by building an agent-based model that can be used to simulate a threat scenario. Using Monte Carlo simulations, vulnerabilities can then be assessed. The agent-based modelling approach allows for modelling complex environments that emerge from the interactions between autonomous agents. Furthermore, human aspects can be explicitly modelled and therefore be taken into account in the assessment of vulnerability. On the contrary, agent-based models require an extensive modelling effort and this is a time-consuming process. As this approach forms a promising alternative to overcome the limitations of the other methodologies, we employed this methodology to assess vulnerabilities at security checkpoints.

2.3. Modelling security operator performance and decision making

Different studies have shown that the performance of security operators is not always optimal and that it is common for them to bend and break the rules (Kirschenbaum et al., 2012a; Kirschenbaum et al., 2012b; Kirschenbaum, 2013; Kirschenbaum, 2015; United States House of Representatives, 2011).

The performance of humans is dependent on a variety of factors, of which cognitive demands of a task and personality are two examples (Gonzalez, 2005; Hancock, 1989). Several computational models have been proposed in literature to model the performance of humans. Many of these models have a specific focus on aspect, like situation awareness (Endsley, 1995). The Functional State Model is a dynamic performance model that describes the performance of an agent as a function of task complexity, the state of the agent and its characteristics (Bosse et al., 2008). The model incorporates a large set of different factors, such as stress, exhaustion and situation awareness. The model was validated by empirical experiments with human operators from defense. This is also the model that we use in this work to model human performance, as it aims to incorporate a diverse set of factors.

Human decision making has been a long-studied field, and an extensive overview of modelling human decision making can be found here (Osman, 2010; Canellas, 2017). Two main streams can be distinguished when modelling human decision making: bounded and unbounded rationality (Canellas, 2017). In bounded rationality, decisions are made within a set of human constraints such as limited information or processing speed of the brain, while in unbounded rationality these constraints are not present. In this work, we also use the bounded rationality paradigm. Within the bounded rationality paradigm, several types of models are developed in literature: linear decision making models (Canellas, 2017), machine learning approaches (Gibson et al., 1997) and diffusion models (Busemeyer and Townsend, 1993; Ratcliff and McKoon, 2008). We focus on probabilistic decision making, which is shown to be well capable of account for human irrationality. Important models in this area are the decision field theory model (Busemeyer and Townsend, 1993) and the Ratcliff diffusion model (Ratcliff and McKoon, 2008). We used the decision field theory model in this work to model decision making of security operators, as it has strong empirical support and is famous for its ability to reproduce many known irrationalities in human decision making.

Only a few works exist that aim to model the behavior of security operators (Skorupski and Uchroński, 2015; Skorupski and Uchroński,

2017). This research models the effects of human factors on the performance of the security system by using a fuzzy inference system. However, their system is mostly based on expert opinions. Our work focuses on more detailed cognitive models of human security operators, and how they can be used to estimate vulnerability. Furthermore, we explicitly represent interactions between agents (security operators and attackers), and important security devices, such as body scanners.

3. Modelling the security checkpoint

This section describes the agent-based model that was developed to assess vulnerabilities at an airport security checkpoint, while focusing on human performance and decision making. The specification of the environment is discussed in Section 3.1, and the different types of agents are discussed in Section 3.2.

3.1. Environment

The environment of the model contains four different objects: luggage, weapons, sensors and equipment. *Luggage* has a *complexity level* that influences the task complexity of operators that interact with this luggage. The complexity level can either be *high* or *low*. Furthermore, luggage is owned by a passenger and may contain *explosive traces* (represented as a Boolean value) and/or a weapon.

Then, the *weapon* object conceptualizes a weapon that an attacker agent aims to bring past the security checkpoint. A weapon is of a certain *type*. The type of weapon can, for instance, be a ceramic knife or an explosive liquid. Similar to luggage, a weapon can contain *explosive traces*. Furthermore, a weapon has a *perceived risk*, r_{perc}, that indicates to which extent a security operator perceives objects that closely resemble the weapon as a risk to the airport. For instance, explosive liquids resemble a bottle of water and are therefore perceived as a low risk. Perceived risk is formalized as a real number between 0 and 1. While r_{perc} is different for each operator, we assumed that it is the same for everyone, and, therefore, included it as part of the weapon. Finally, a weapon can be on the body of an attacker, or in the luggage of an attacker. A full list of weapon types and their corresponding parameters is shown in Section 4.

Different sensors were defined in the model: X-ray sensor, Walk-through metal detector (WTMD), explosive trace detector (ETD) and body scanner. Each sensor has a probability to detect a specific weapon type, called base detection probability $p^{sensor}_{detect}(weapon\_type)$. This parameter is calibrated and shown in Section 4. Based on this detection probability, the sensor either detects or does not detect a weapon when presented, which can then be observed by operators. The X-ray sensor is an exception to this standard, as this sensor only allows operators to observe the luggage that is currently sensed by the sensor. In this case, the likelihood of detection is determined by the skill of the x-ray operator.

Finally, two types of equipment were defined in the model: *queue separators* and *X-ray systems*. Queue separators are used to guide passengers to the security checkpoint, while the X-ray system moves luggage forward through an X-ray sensor.

3.2. Agents

Three different agent types were defined: passengers, attackers, and operators. Each of these agents is human agents and is discussed in more detail below.

3.2.1. Passengers and attacker agents

Passengers and the attacker were defined similarly. They both do not exhibit sophisticated strategical behavior. Passengers carry luggage that they bring to the security checkpoint. Furthermore, passengers could carry *explosive traces* and they can own a weapon. This weapon can, as defined above, either be on the body of the passenger or in its luggage. We refer to a passenger that owns a weapon as an attacker, and as a passenger otherwise.

3.2.2. Security operators agents

A set of security operators that execute activities at the security checkpoint were defined: patdown operator, ETD check operator, luggage check operator, and X-ray operator. This section discusses the definition of the X-ray operator, as the other operator types are defined similarly.

Airport security is largely defined by regulations and guidelines defined by different regulatory institutes. For instance, the European Union has regulations for its member countries (Council of European Union, 2008a; Council of European Union, 2008b), the United States has the Aviation and Transportation Security Act (107th Congress, 2001), and the ICAO developed a security manual (ICAO, 2017).

Following these regulations and guidelines, each of these operators executes a fixed set of tasks and decisions. An X-ray operator inspects output generated by the X-ray machine and determines if there is a potentially illegal item. When this is the case, (s)he has to inform the luggage check operator, who then searches the luggage. Security operators do not necessarily follow this protocol, but regularly bend and break the rules. They commonly ignore potential threats and alarms are often processed as false (Kirschenbaum et al., 2012a; Kirschenbaum et al., 2012b; Kirschenbaum, 2013; Kirschenbaum, 2015; United States House of Representatives, 2011). Furthermore, humans cannot continuously perform optimally. It is dependent on a variety of factors, of which cognitive demands of a task and personality are two examples (Gonzalez, 2005; Hancock, 1989).

The performance of security operators on different tasks in the checkpoint is modelled using the functional state model (Bosse et al., 2008), and their decision-making process is modelled using decision field theory (Busemeyer and Townsend, 1993). In the case of the X-ray operator, the modelled task is inspecting output generated by the X-ray machine, while the modelled decision is that of informing or not informing the luggage check operator.

The functional state model, the decision field theory model, and their integration are discussed below.

3.2.2.1. Functional state model. To model the performance of security operators, the functional state model was selected (Bosse et al., 2008). While the model contains a set of 37 parameters, we only discuss the most important parameters here. For the other parameters, the reader is referred to the work of Bosse et al. (2008). The input for the model is the *task level* (*TL*), which is dependent on the *skill level* (*SL*) of the operator and the *task complexity* (*TC*) of the task at hand. For an X-ray operator, the task complexity represents how complex the luggage (s)he is currently investigating is. This dependency is modelled as follows:

$$TL(t) = \frac{TC(t)}{SL} \tag{1}$$

The output of the model is the *performance quality* (*PQ*) of the agent, indicating how well the operator is performing. A *PQ* of 1 corresponds to the baseline performance of an agent, while values lower than 1 correspond to performances that are worse than this baseline and values higher than 1 correspond to performances better than baseline. No theoretical bounds of *PQ* were provided in Bosse et al. (2008). However, typical *PQ* values in our simulation results are in the range of 0.5 and 1.5.

PQ is dependent on two factors: provided effort (*PE*) and task level (*TL*):

$$PQ(t) = \frac{PE(t)}{TL(t)} \tag{2}$$

PE is determined by the generated effort (*GE*) of the agent, recovery effort (*RE*) and noise effort (*NE*). The latter two parameters correspond to the ability of humans to decreases exhaustion, and the effort the human has to contribute to the noise in the environment respectively.

$$PE(t) = GE(t) - RE(t) - NE(t) \tag{3}$$

GE is the most important contributor to PE, and is ultimately defined by effort motivation (*EM*), among many other parameters. We refer to Eqs. (2) and (4) in the work of Bosse et al. for a complete deduction of *GE* (Bosse et al., 2008; Bosse et al., 2011).

Effort motivation is based on the current task level and the difference between experienced pressure (*EP*) and optimal experienced pressure (*OEP*). *EP* is similar to a person's stress level, while *OEP* determines how well a person can cope with a high *EP*. Finally, *EP* is, among other terms, related to generating effort above and below a critical point. The critical point is the amount of effort someone can generate without becoming exhausted. For an X-ray operator, *PQ* is reflected in the likelihood (s)he observes a weapon from the observations of the X-ray sensor. This is modelled as follows.

$$p_{detect}^{operator}(weapon\_type) = max\left(0, 1 - \frac{1 - p_{detect}^{x-ray}(weapon\_type)}{k \cdot PQ}\right) \quad (4)$$

The value $p_{detect}^{x-ray}(weapon\_type)$ corresponds to the base likelihood that a specific type of weapon is detected by an X-ray operator, which is calibrated in Section 4.2. The value $1 - p_{detect}^{x-ray}(weapon\_type)$ corresponds to the base probability of not detecting the weapon: the base false-negative rate. When performing well (i.e. a high *PQ*), X-ray operator improves on this base false-negative rate, and vice versa. We model this by dividing the base false-negative rate by the performance quality and a scaling factor *k*. The underlying assumption here is that the false-negative rate linearly decreases with increasing *PQ*. This false-negative rate is then transformed back to a detection probability by subtracting it from 1. To ensure that the value falls between 0 and 1, we take the maximum of 0 and the value obtained above.

The other operators at the security checkpoint use this performance model to execute the patdown activity, search luggage and perform an ETD test. The value of *k*, and other related parameters of the functional state model are calibrated in Section 4.

Two different personality types are introduced based on the work of Bosse et al.: personality I and personality II (Bosse et al., 2008). Bosse et al. extensively experimented with these two personality types and performed an in-depth analysis of their behavior. Type I has a relatively high *OEP*, meaning that it can cope well with high *EP* levels, while type II does not. This allows the first personality type to perform better under high pressure. We experiment with these personality types in our analysis.

3.2.2.2. Decision field theory. The decision-making process of the security operators was modelled based on the work of Busemeyer and Townsend (1993). The decision-making process in this model is an iterative process in which the operator constantly updates their *preferences* until the preference for one of the *options* exceeds a *decision threshold* value. This threshold value is one of the inputs of the model and its magnitude is related to the effort an agent spends on a decision. The higher the threshold value, the more time and energy the security operator needs to reach it.

During each iteration, the agent focuses on one of his *goals*. The selection of this goal is a random process, but the likelihood of the agent focusing on a goal depends on the *attention weight*. Once the attention of the agent is focused on one of his goals, the agent's preferences are updated based on the agents *beliefs* about how each of the options helps him in achieving the goal (s)he currently focuses on. The magnitude with which the preference for each of the goals is updated is known as *valence*. This valence is defined for each combination of goals and options.

Finally, the decision-making process is influenced by the agent's initial beliefs. This *initial preference* is the preference the agent has for each outcome before the decision process starts. An overview of this process is shown in the bottom part of Fig. 1.

For an X-ray operator, one decision is identified. If the X-ray operator observes a potential weapon (see also Section 3.2.2), (s)he has to decide if the luggage requires a search from the luggage check operator. The options for the X-ray operator are *inform* or *ignore*. Furthermore, three goals are defined for the X-ray operator based on existing literature (Sharpanskykh and Haest, 2016; Fairbrother, 2010).

- Accuracy
  - The operator wants to do their work as well and accurate possible. The importance of this goal may be dependent on pressure within the organizations or the agent's standards.
- Speed
  - The operator wants to do their job as fast as possible. The importance of this goal may be due to pressure within the organization to reach a certain throughput or the security operator wanting to minimize effort.
- Perceived Risk
  - It is the job of the security operator to minimize the risk of an attack. Perceived risk represents the beliefs an agent has about the potential consequences of the observed prohibited item. The importance of this goal may be dependent on the agent's beliefs about the likelihood of an attack and his risk aversion.

Both luggage check operators and physical check operators use this decision mechanism to determine if a passenger requires secondary screening when an illegal object was found. The ETD operator makes the same decision when explosive traces were observed. The other related parameters of this model are calibrated in Section 4.

3.2.2.3. Integration of models. We integrated the models by relating parameters of the functional state model to the decision field theory model. The relation between the models is shown in Fig. 1.

The *decision threshold* was set to be equal to the provided effort (*PE*) as defined in the functional state model. Provided effort denotes the effort that is contributed to the task by the agent. This relation means that the higher the provided effort, the more effort the agent wants to invest in making an accurate decision. This is based on findings by Busemeyer and Townsend (1993). Furthermore, we assumed that the *initial preference* of the X-Ray operator is according to regulations present at the security checkpoint, meaning that there is a strong initial preference to request a luggage check if needed. The next section describes how these parameters are calibrated.

4. Model sensitivity and calibration

In this section, the sensitivity of the functional state model and the decision field theory model is discussed. Furthermore, it is described how the overall model was calibrated. Different parameters had to be calibrated: parameters related to weapons, sensors, airport configurations, and operators. These are discussed in detail below.

4.1. Sensitivity analysis

We performed sensitivity analysis of both the functional state model and the decision field theory model. Fig. 2(a) shows how different task levels affect the performance quality and the provided effort in the functional state model. Results were obtained after the task level was kept constant for 20 s. At this point, the performance quality converged to an equilibrium value for any task level. From the figure, it becomes clear that both personality types have the highest performance quality around a task level of 250. The peak performance of personality type I is at a task level 230. At this point, it outperforms personality type II by 24%.

At task levels lower than 225 the performance quality of both personalities rapidly drops. This is mainly due to a lack of provided effort as can be seen in Fig. 2(b). In this range, personality type II outperforms personality type I by 20%. At task levels above 275, the performance of both personalities exponentially decreases. The provided effort of both agents stays approximately stable around 230, meaning that the agent

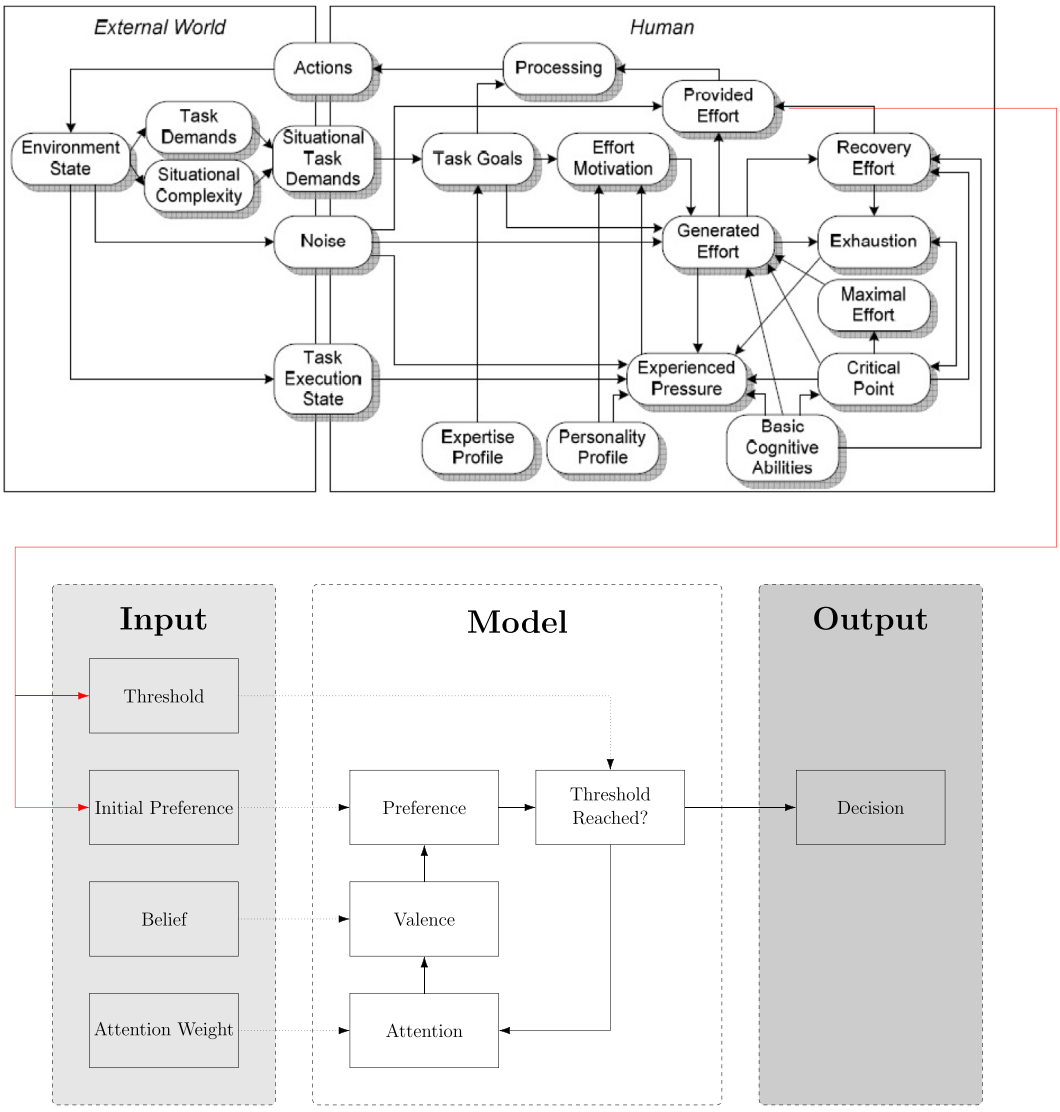


Fig. 1. An overview of the functional state model (Bosse et al., 2008) (top) and the decision field theory model (Busemeyer and Townsend, 1993) (bottom) used in this work. The integration between the two models is shown as well.

cannot provide more effort. At even higher task levels the performance quality drops.

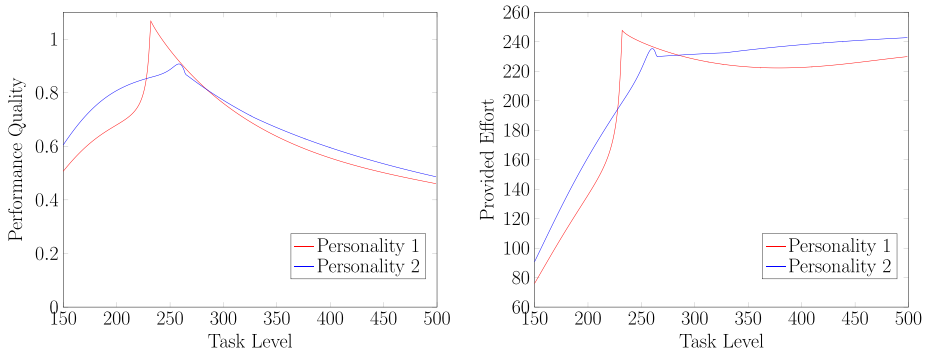
We investigated the sensitivity of the decision field theory model as well. The values in Table 3 were used with $c = 30$, but the initial preference for the inform decision was varied. The decision threshold was set to a uniform random value between 70 and 250, which is the range of provided effort values as observed above. Two scenarios were investigated: 1) a weapon with a perceived risk of 0 was observed, and 2) a weapon with a perceived risk of 1 was observed. A total of 1000 simulations were performed for each data point.

Fig. 3 shows how different initial preferences for the inform decision influence the decision of the X-ray operator. Both graphs have the same general shape. The choice to inform the luggage-check operator increases from a baseline value to 100% when the initial preference becomes 250. At initial preferences above 250, the operator chooses to inform the luggage-check operator 100% of the time. This is because the initial preference already exceeded the threshold value. The range of values for both scenarios is different. In scenario 1, the luggage-check operator is informed 93% of the time without any initial preference, while this is only 50% in scenario 2. In scenario 1 the dominant decision is to inform, as two out of three goals favor this decision. As the perceived risk is zero in scenario 2, there are effectively only two goals: accuracy and speed. Neither of these goals dominates the other, leading to a baseline of 50% inform decisions.

4.2. Weapon and sensor calibration

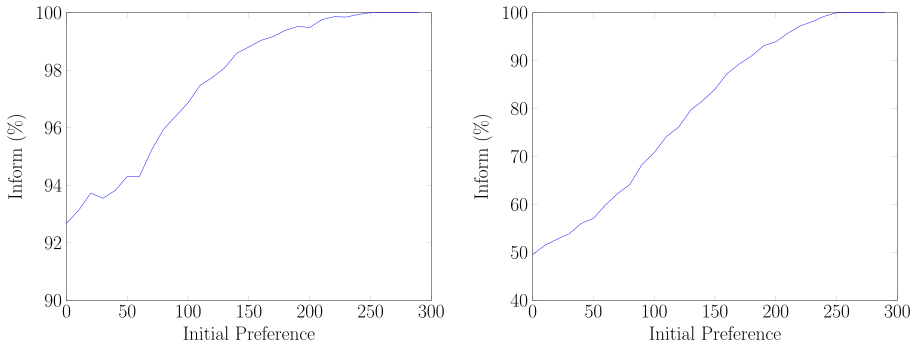
Table 1 shows the different weapon types used in this work. For each of the weapon types, it is indicated if they contain explosive traces and their perceived risk. It should be noted that the perceived risk represents the risk that is perceived by operators for objects that resemble the weapon. For instance, if a bomb is not recognized as such (like explosive liquids), the perceived risk is much lower. Bombs and fire-arms were assumed to have the highest perceived risks, while liquids were not perceived as a large risk, as operators continuously confiscate water bottles. Knives were perceived as a larger risk, but they are still commonly observed.

Table 2 shows the different detection probabilities for weapon-sensor combinations and weapon-activity combinations. The values for X-ray performance is based on literature (Wales et al., 2009), as well as the explosive bulk detection probabilities for body scanners (Grabell, 2011). No data could be found on how security operators perform on searching luggage and patdowns. These values were therefore based on assumptions.



(a) Performance quality for different task levels. (b) Provided effort for different task levels.

Fig. 2. The effect of changing task levels on performance quality and provided effort.



(a) $r_{perc} = 1.$ (b) $r_{perc} = 0.$

Fig. 3. The effect of changing initial preferences on the decision of X-ray operators to inform luggage-check operators when an illegal item was observed.

4.3. Airport configurations calibration

Two different airport configurations were defined based on IATA documentation (IATA, 2012): the regional airport and the international airport. There are two main differences between these configurations. First, the regional airport uses a WTMD whereas the international airport uses a body scanner. This choice of equipment impacts the detection rates of weapons hidden on the body of the attacker. The second difference is the communication between the X-ray operator and the luggage check operator. At the regional airport, there is the possibility to communicate directly, while at the international airport the luggage check operator is not in direct contact with the X-ray operator. The luggage check operator has to perform an X-ray himself/herself to determine where the weapon can be found.

Table 1
The different types of weapons with a description, an indication if explosive traces are present and their perceived risk.

Type	Explosive traces	r_{perc}	Description
Explosive bulk	Y	1	An improvised explosive device.
Explosive liquid	Y	0.1	Liquid explosives are often not directly recognized as a bomb.
Explosive powder	Y	0.2	Explosives in powder form are commonly not recognized as a bomb.
Gun	N	1	A standard handgun.
Knife	N	0.3	A small knife.
Ceramic knife	N	0.3	A small knife without metal.

4.4. Operator performance calibration

The task complexity (TC) of operators for the different tasks that they perform are calibrated in this section. To this end, we assumed three different skill levels (SL) for operators: 0.8 (low), 0.9 (medium), and 1.0 (high). These levels correspond to a realistic variation in skill between agents in the functional state model, based on experimentation with the functional state model and values found in literature (Wales et al., 2009). Furthermore, we assumed a base task level (TL_{base}) of 150, corresponding to a performance quality (PQ) of around 0.5. The base task level is the task level when the agent is not performing its activity and is based on the work of Bosse et al. (2008). An X-ray operator has about 1 s to identify potentially prohibited items in luggage and research has shown that the number of false negatives increases when the images become harder to interpret (Wales et al., 2009). Based on the same work, we assumed that the performance of an X-ray operator decreases with 5.47% for more complex luggage.

The scaling factor k was calibrated as follows. The value for $k \times PQ$ should be equal to 1, as an average operator performs according to the

Table 2
The detection probability $p_{detect}(weapon)$ for each sensor and activity.

	Expl. bulk	Expl. liquids	Expl. powder	Gun	Knife	Ceramic knife
WTMD	0.00	0.00	0.00	1.00	1.00	0.00
Body scanner	0.56	1.00	0.00	1.00	1.00	1.00
X-ray activity	0.735	0.645	0.645	0.875	0.675	0.675
Lugg. search activity	0.90	0.90	0.90	0.90	0.90	0.90
Pat down activity	0.90	0.90	0.90	0.90	0.90	0.90

base detection probability. We have performed 1000 simulations for each skill level, personality type, and task level to determine the mean performance quality of operators, and found that it corresponds to 0.59. We use this mean performance quality to finally find k to be equal to 1.68.

4.5. Operator decision calibration

For operator decision making, the following parameters had to be calibrated: initial preference, decision threshold, and valence. We used attention weight as a parameter to experiment with. The decision threshold is equal to the provided effort *PE* as suggested in the work of Busemeyer and Townsend (1993). The valences and initial preferences are shown in Table 3. The initial preference for the *inform* decision was assumed to be the decision threshold of the agent multiplied with a constant of $c_{pref} = 0.95$. This indicates that the X-ray operator has a strong preference to follow rules and regulations. Furthermore, the valences related to speed and accuracy are $\pm c$ for both options. We chose $c = 30$ such that the mean decision time of X-ray operators corresponds to times reported in literature (Wales et al., 2009). Finally, the valences for the perceived risk goal were made dependent on the perceived risk of the observed weapon. We assumed this to be a multiplication between c and the perceived risk. The parameters of the other operators were determined similarly.

5. Experiments and results

We performed experiments with the model to assess vulnerabilities at different security checkpoint setups. The setup of the experiments is discussed first, followed by a discussion of results.

5.1. Experimental setup

The model was implemented in the AATOM simulator, which is a Java-based airport terminal operations simulator (Janssen et al., 2019a). It is agent-based and contains several calibrated presets and templates of basic airport terminal components that can readily be used. No other simulator that we know of contains such a combination of agent-based modelling and pre-calibrated airport-specific components.

As specified in the model description, four operator agents were defined in the model: patdown operator, ETD check operator, luggage check operator, and X-ray operator. We used a single security lane setup, and passengers were generated for a single flight with up to 100 seats. A single attacker was introduced among the passengers that went through the security checkpoint.

The following parameters were varied in the execution of the experiments.

- *Attacker parameters*
  - *Weapon.* The weapon the attacker uses is one of the weapons shown in Table 1.
  - *Weapon Location.* The attacker has the option to hide the weapons on his body or in his luggage.
- *Checkpoint Configuration.* The checkpoint configuration is either the regional airport or the international airport.
- *Operator parameters*
  - *Skill Level.* The skill level of the agents is either 0.8 (low), 0.9 (medium) or 1.0 (high).
  - *Personality Type.* The agents either have personality I or II, based on the work of Bosse et al. (2008).
  - *Attention Weights.* The attention weight for each goal is set to 0.33 (low), 0.5 (medium) or 0.67 (high). The weights are normalized so that they add up to one after they are selected.

A total of $N = 15{,}000$ simulation runs were performed, while using a uniform random assignment of the above parameter values.

Table 3
Calibration of the decision parameters for the X-ray operator.

	Initial Pref.	Accuracy	Speed	Perc. risk
Inform	$c_{pref} \cdot DT$	c	$-c$	$c \cdot r_{perc}$
Ignore	0.0	$-c$	c	$-c \cdot r_{perc}$

5.2. Results

The results are discussed as follows. We define vulnerability as the proportion of attackers that moved past the security checkpoint with their weapon. These attackers did not receive secondary screening and their weapon was not confiscated. We first show how the skill level and personality type of security operators influence their performance. Then, we show how different attention weights of the decision field theory model influence the decisions made by the operators. Both these results are an indication of the vulnerability of the security checkpoint, as both performance and decision making directly influence the number of secondary screenings and weapon confiscations. Finally, an overall vulnerability assessment of the different checkpoint configurations is conducted and a discussion is provided.

5.2.1. Performance of operators

The performance quality of X-ray operators and luggage check operators can be found in Fig. 4. The performance quality of security operators is directly related to the vulnerability of the security checkpoint. A low performance quality of any of the operators leads to a higher vulnerability, as items are detected with a lower probability. As can be seen in the figure, *PQ* increased with skill level for X-ray operators. The agents with the highest skill level (of 1) outperformed the agents with the lowest skill level (of 0.8) with 5.2%.

Different results were observed for luggage check operators. The operators with the highest skill level were outperformed by the agents with the lowest skill level by 4.0%. This also seems counter-intuitive but can be explained from the mechanisms of the functional state model. If the task level becomes too low, the performance quality drops, as skilled agents are not motivated enough to generate effort. For operators with a lower skill level, the task is more challenging and they are more motivated to put in the effort. This lead to the counter-intuitive result that the most skilled agents were not top performers on this relatively simple task. This result may seem counter-intuitive but is caused by the fact that agents perform (relatively) simple tasks and find it hard to motivate themselves to put in enough effort. Following the functional state model, operators with a higher skill level, experience a lower task level for the same task as their lower-skilled counterparts. Generated effort is, among other parameters, based on the motivation of the operator, which in turn is partially determined by the task level. Because the task level is lower for higher-skilled operators, the effort motivation decreases, which decreases the provided effort. Our simulation results have shown that this negative effect on performance quality of decreased motivation is larger than that of an increased task level for lower-skilled operators. Section 3.2.2 provided a discussion of the different variables in the functional state model.

These results are not unique to the Functional State Model and our model. Hackman and Oldham proposed a so-called Motivating Potential Score (Hackman and Oldham, 1976) which is a framework that is widely used in literature. MPS is, among other terms, composed of skill variety. This is strongly related to what we have defined as skill level in our paper and explains the connection between motivation and skill level. Furthermore, jobs with a high MPS, have a positive effect on motivation, performance and job satisfaction (Singh et al., 2016). This then relates motivation to task performance.

The differences between the performance quality of analyzing X-ray images and checking luggage can be explained as followed. Analyzing X-ray images is a difficult cognitive task for humans. A large number of stimuli have to be processed and illegal items have to be identified at a high speed. An operator performing a luggage check has more time to

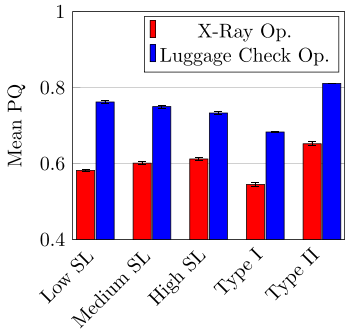


Fig. 4. The mean performance quality (and their 95% confidence intervals) for the different skill levels and personality types.

execute the task at hand. Typically, they take around 90 s, while X-ray operators only have a few seconds for their task. This allows the luggage check operators to generate more effort and therefore reaching higher performance quality.

Furthermore, agents that cannot cope with pressure well (type II) outperform agents that can better cope with pressure (type I) by 20%. The results of the different personality types shown in this figure are an aggregate of all skill levels. Personality type II has a relatively low *OEP*, which is closer to the actual experienced pressure than the high *OEP* of personality type I. The difference between these values determines the effect on effort motivation. A low difference leads to a low reduction in effort motivation, while a high difference leads to a high reduction of effort motivation. As mentioned before, a lower effort motivation finally leads to a lower performance quality.

5.2.2. Decision making of operators

We analyze the decision-making process of X-ray operators. When an X-ray operator detects a potential weapon, the agent has two options. The first option is to ignore that the potential weapon was observed, while the second option is to inform the luggage check operator. When a potential weapon was detected, luggage check operators were informed correctly 93.7% of the time on average. This number varied based on the attention weights for each of the goals, as shown in Fig. 5. Not searching luggage when it contains a weapon, directly increases the vulnerability of the system.

One of the reasons for an X-ray operator to not inform the luggage check operator is that it might not perceive the potential weapon as an actual weapon. For instance, liquid explosives might resemble a water bottle. While a water bottle is illegal according to checkpoint regulation, regulations are not always strictly enforced by security operators (Kirschenbaum et al., 2012a; Kirschenbaum et al., 2012b; Kirschenbaum, 2013; Kirschenbaum, 2015). Not informing the luggage check operator then leads to faster processing of passengers, which is of enormous economic importance for airports.

From the figure, it becomes apparent that the attention weight for speed was the most dominant parameter in the inform decision. Varying this parameter from low (0.33) to high (0.67), lead to a 12% decrease in luggage searches. The second most important parameter is the attention weight for accuracy. Increasing this parameter from low to high, caused an 11% increase in luggage searches. The attention weight for risk was less dominant. An increase from low attention to high caused a 5.3% increase in luggage searches. This parameter was less influential, as many potential weapons are not perceived as a large risk by the operators. Speed and accuracy, on the other hand, played a more important role in the decision-making process. While not shown, results for decisions by other types of operators followed similar trends.

Jesus performed a questionnaire among security operators at a regional airport to determine how they make trade-offs between security and efficiency (Jesus, 2018). One of the main findings of his research was that operators could be classified into three categories: 1) passenger level of service operator, 2) security-focused operator, and 3) efficiency-focused operator. About 13% of the surveyed employees fell into the last category. These employees mostly focused on improving the efficiency of checkpoint operations and barely on security.

While the results of Jesus are not readily comparable to our results, we do observe an interesting similarity between them. Both results indicate that some employees mostly focus on executing their work efficiently (i.e. high attention weight for speed), which then results in increased vulnerabilities.

5.2.3. Different checkpoint setups

The performance of the security checkpoint for different weapons and locations are shown in Fig. 6. In this figure, the distribution between three potential outcomes of a scenario are shown: *vulnerability* (weapon not confiscated and no secondary screening), *secondary screening* (regardless of weapon confiscation) and the situation in which the *weapon* was *confiscated* while no secondary screening was conducted.

From this figure, it becomes clear that some weapons were never confiscated at the regional airport. These weapons cannot be detected by the equipment used to scan the passengers. None of the explosives smuggled on the body get detected by the WTMD and the same holds for ceramic knives. Explosives only got detected by a random ETD check, which lead to a secondary screening in 10.1% of the cases. Furthermore, knives can be taken through the checkpoint at the regional airport without large consequences. Most often, the knife got confiscated and the attacker could try again at a different time as the chances on a secondary screening were found to be almost zero. The regional airport performed best on detecting guns in luggage. These weapons were confiscated 84.8% of the time when they were located in the luggage (as compared to 70.0% in the international airport) and immediately lead to a secondary screening. This becomes 90.7% when the attacker carried the weapon on their body.

At the international airport, only one type of weapon remained undetected. Smuggling explosive powder through a body scanner had a success rate of 88.6%. The only measure against it was a random ETD check. Furthermore, liquid explosives and powders hidden in luggage were confiscated only 32–34% of the time. Even when these items were confiscated, the security operator did not necessarily recognize these items as bomb parts and allowed the attacker to move on. Bulk explosives, on the other hand, were detected in 50% of the cases and lead to immediate secondary screening. An attacker bringing a gun was very unsuccessful at the international airport. The attacker was most successful when locating the gun in their luggage, but this only had a success rate of 30%. Knives could best be brought hidden in the luggage as well. In that case, they were only confiscated 36% of the time and the chances of secondary screening were minimal. However, the potential impact of a knife past the security checkpoint is far more limited than that of other weapons investigated in this work.

The regional airport outperformed the international airport on checking luggage for all weapons. In the regional airport, 62.6% of the weapons in the luggage are confiscated, whereas in the configuration of the

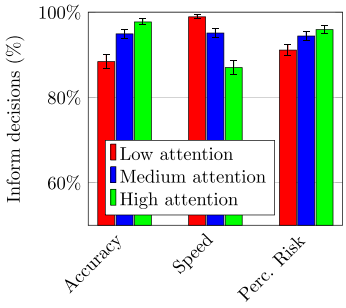


Fig. 5. Percentage of correct *inform* decisions (and their 95% confidence intervals) for the different attention weights.

international airport this is only 42.6%. The main reason for this is the lack of communication in the configuration of the international airport. The X-ray operator flags luggage for a search, but the luggage check operator has to identify the weapon on the X-ray image himself. This extra step in the process caused a loss in performance of 32% and occurs solely because two officers independently had to recognize a weapon on an X-ray image instead of just one.

5.2.3.1. Relation with related work. ABC News reported in 2015 that in 95% of trials undercover investigators were able to smuggle mock explosives or banned weapons through checkpoints (Fishel et al., 2015). Two years later, this percentage decreased to a still extremely high value of around 80% (Kerley and Cook, 2017). These problems do not only exist in the United States. The Telegraph reported late 2014 that airport security failed to detect half of the dangerous weapons at Frankfurt airport (Huggler, 2014). The vulnerabilities that we found in this work (see Fig. 6) are close to the vulnerability as described in these public reports. However, to the best of our knowledge, there is no public data available that evaluates the effectiveness of security checkpoints specifically for different weapon types, as we do in our work. We have used all data that was available to calibrate the capabilities of the different sensors to detect each weapon type in our model. However, more work is needed to validate our results. This can be done by performing penetration tests with different weapon types at security checkpoints, and comparing these results with our simulation results.

5.3. Discussion of results

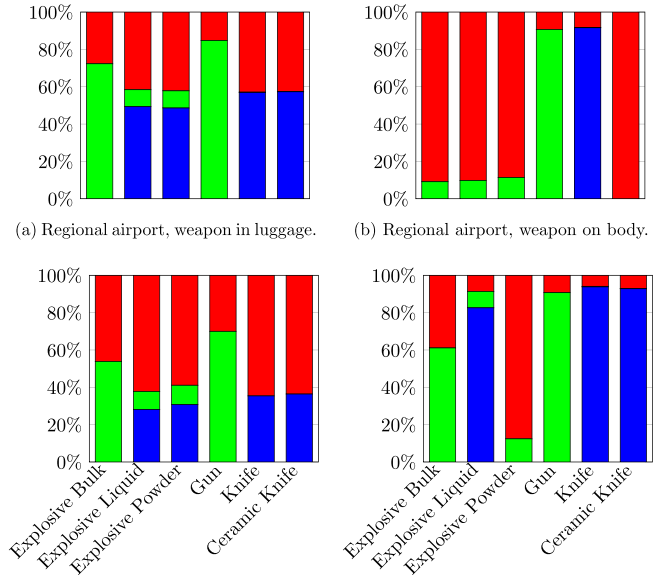
As mentioned in Section 4, the calibration of the model was based on a set of simplifying assumptions. These assumptions influenced the magnitude of the resulting vulnerabilities in different checkpoint configurations. The calibration of the model can be improved by performing field tests to determine different parameters. For instance, the performance of operators for searching luggage can be evaluated by providing security operators a set of luggage containing legal and illegal objects. Furthermore, perceived risks of objects can be evaluated for different operators using a similar method. Decision making of operators can be calibrated better by performing choice experiments, such as the one performed by Jesus (2018).

While vulnerability estimates are inherently hard to validate, some researchers performed real-life experiments (Ford, 2017; Gholami et al., 2017). This form of validation is a direction of further research for this work. Both calibration and validation of the model can still be improved, but the proposed model is still valuable for airport security practitioners, as it can be used to generate improved results when more data becomes available.

We did not consider all security mechanisms that are present in airports. For instance, intelligence agencies can detect attackers before they arrive at the security checkpoint. Behavior detection officers (United States Government Accountability Office, 2013; Winter and Cora, 2015) are also capable of detecting suspicious behavior at the security checkpoint and perform secondary screenings based on that. Furthermore, more strategic attacker behavior in which the attacker chooses the right type of weapon for the checkpoint configuration can be considered as well.

Agent-based modelling is an important tool to better understand complex systems. Using our model, vulnerabilities caused by imperfect human decision making and performance were identified. Understanding how these vulnerabilities emerge enables airport security managers and policymakers to improve their security policies and reduce vulnerabilities. For instance, in Section 5.2.1 we found that operators with a low skill level outperform their higher-skilled counterparts on checking luggage. This is a surprising result that hiring officials can take into account while hiring security operators. Furthermore, higher-skilled operators outperform lower-skilled operators on the more difficult task of examining X-ray images. We also quantify the effect of different skill levels on the performance of operators (see Fig. 4). This result can be used as a basis for hiring officials to hire operators for specific positions. Furthermore, they can use this result to better understand the effects of the composition of their current team on the performance of the security checkpoint.

Our model can also be used to test future concepts of security checkpoints. For instance, when X-ray officers do their work remotely, our model can be adapted with relative ease to determine the performance of such a setup. These types of experiments cannot easily be performed at airports, as it may interrupt security operations. Furthermore, experiments with humans are known to be hard to perform due to the diversity of human behavior. Using our model, these experiments can be performed more easily. This can, for instance, be done by hiring operators with the right personality type and skill set, or by taking these aspects into account while planning operators.



(a) Regional airport, weapon in luggage. (b) Regional airport, weapon on body.

(c) International airport, weapon in luggage. (d) International airport, weapon on body.

Fig. 6. The performance of the checkpoint setups for the different weapons defined in this work. Performance is shown in terms of vulnerability (weapon not confiscated and no secondary screening; red bar), percentage of secondary screening (with or without weapon confiscation; green bar), confiscated weapons (without secondary screening; blue bar).

Our agent-based approach is more time consuming to perform than most other vulnerability assessment methodologies and requires a large amount of data for calibration. Other vulnerability assessment methodologies form better alternatives in cases with a lack of time or data, but our approach is particularly suitable to investigate vulnerabilities in which human behavior plays a role. A more in-depth discussion about the advantages and disadvantages of the use of agent-based modelling is discussed by Janssen et al. (2019b).

6. Conclusion

In this paper, we investigated how the decision-making and performance of human operators can be taken into account while assessing vulnerability at an airport security checkpoint. We developed an agent-based model, in which the performance of these operators was modelled using the functional state model, while decision making was modelled using decision field theory.

Simulation results indicated that the highest skilled operators outperformed their lowest skilled counterparts on analyzing X-ray images, but performed worse on both searching luggage and performing patdowns. This lead to similar differences in security checkpoint vulnerabilities as well. These skilled operators found their tasks too easy and are unable to motivate themselves to put in the required effort. Furthermore, the goals the operator focuses on during the decision-making process were found to influence vulnerability. A high focus on accuracy or perceived risk for the X-ray operator lead to an increase in luggage searches, and therefore reduced vulnerabilities. However, a high focus on speed leads to a decrease in luggage searches and therefore increased vulnerability. The developed model can be used to assess the effect of human behavior and decision making on the performance of current and future security checkpoint procedures, which is often impossible in real-life experiments. More work is needed to calibrate and validate the model and simulation results, but initial results are promising.

This work can be extended by investigating how other types of security measures (i.e., behavior detection officers) influence the vulnerability of the security checkpoint. The vulnerability with respect to other threat scenarios (i.e., a bomb attack before the security checkpoint) can be investigated as well. Finally, the model can be calibrated better by using classified data on sensor performance, operator performance and attacker behavior.

References

107th Congress, 2001. Aviation and transportation security act. https://www.gpo.gov/fdsys/pkg/PLAW-107publ71/pdf/PLAW-107publ71.pdf.

Airport data & contact information. https://www.faa.gov/airports/airport-safety/airportdata-5010/ (accessed: 2019-11-12).

Akgun, I., Kandakoglu, A., Ozok, A.F., 2010. Fuzzy integrated vulnerability assessment model for critical facilities in combating the terrorism. Expert Syst. Appl. 37 (5), 3561–3573.

Babu, V.L.L., Batta, R., Lin, L., 2006. Passenger grouping under constant threat probability in an airport security system. Eur. J. Oper. Res. 168 (2), 633–644.

BBC News, 2006. Airlines terror plot disrupted. http://news.bbc.co.uk/1/hi/uk/4778575.stm (accessed: 2019-11-12).

Benner, L., 1975. Accident investigations: multilinear events sequencing methods. J. Saf. Res. 7 (2), 67–73.

Bennett, B., 2015. Red team agents use disguises, ingenuity to expose tsa vulnerabilities. http://www.latimes.com/nation/nationnow/la-na-tsa-screeners-20150602-story.html (accessed: 2019-11-12).

Bosse, T., Both, F., Van Lambalgen, R., Treur, J., 2008. An agent model for a human's functional state and performance. Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. 2. IEEE Computer Society, pp. 302–307.

Bosse, T., Both, F., Hoogendoorn, M., Jaffry, S.W., Lambalgen, R.v., Oorburg, R., Sharpanskykh, A., Treur, J., De Vos, M., 2011. Design and validation of a model for a human's functional state and performance. International Journal of Modeling, Simulation, and Scientific Computing 2 (4), 413–443.

Burns, J.F., 2010. Yemen bomb could have gone off at east coast. http://www.nytimes.com/2010/11/11/world/europe/11parcel.html.

Busemeyer, J.R., Townsend, J.T., 1993. Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. Psychol. Rev. 100 (3), 432.

Canellas, M.C., 2017. Decision Making With Incomplete Information. (Ph.D. thesis). Georgia Institute of Technology.

Chawdhry, P.K., 2009. Risk modeling and simulation of airport passenger departures process. Winter Simulation Conference, Winter Simulation Conference, pp. 2820–2831.

CNN, 2001. Shoe bomb suspect to remain in custody. http://edition.cnn.com/2001/US/12/24/investigation.plane/ (accessed: 2019-11-12).

Cole, M., Kuhlmann, A., 2012. A scenario-based approach to airport security. Futures 44 (4), 319–327.

Cooke, R.M., Goossens, L.L., 2008. TU Delft expert judgment data base. Reliab. Eng. Syst. Saf. 93, 657–674.

Council of European Union, 2008a. Council regulation (EU) no 300/2008. http://data.europa.eu/eli/reg/2008/300/oj.

Council of European Union, 2008b. Council regulation (EU) no 1998/2015. http://data.europa.eu/eli/reg_impl/2015/1998/oj.

Edmunds, N.G., 2010. Indictment. https://www.asser.nl/upload/documents/DomCLIC/Docs/NLP/US/Abdulmutallab-Indictment-06-01-2010.pdf.

Endsley, M.R., 1995. Toward a theory of situation awareness in dynamic systems. Hum. Factors 37 (1), 32–64.

Fairbrother, J.T., 2010. Fundamentals of Motor Behavior. Human Kinetics, Champaign, IL.

Fishel, J., Thomas, P., Levine, M., Date, J., 2015. Exclusive: undercover dhs tests find security failures at us airports. https://abcnews.go.com/US/exclusive-undercover-dhs-tests-find-widespread-security-failures/story?id=31434881 (accessed: 2018-09-30).

Ford, B., 2017. Real-world Evaluation and Deployment of Wildlife Crime Prediction Models. (Ph.D. thesis). University of Southern California.

Gholami, S., Ford, B., Fang, F., Plumptre, A., Tambe, M., Driciru, M., Wanyama, F., Rwetsiba, A., Nsubaga, M., Mabonga, J., 2017. Taking it for a test drive: a hybrid spatio-temporal model for wildlife poaching prediction evaluated through a controlled field test. Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, pp. 292–304.

Gibson, F.P., Fichman, M., Plaut, D.C., 1997. Learning in dynamic decision tasks: computational model and empirical evidence. Organ. Behav. Hum. Decis. Process. 71 (1), 1–35.

Gonzalez, C., 2005. Task workload and cognitive abilities in dynamic decision making. Hum. Factors 47 (1), 92–101.

Grabell, M., 2011. Just how good are the tsa's body scanners? https://www.propublica.org/article/just-how-good-are-the-tsas-body-scanners

Hackman, J.R., Oldham, G.R., 1976. Motivation through the design of work: test of a theory. Organizational behavior and human performance 16 (2), 250–279.

Hancock, P.A., 1989. A dynamic model of stress and sustained attention. Hum. Factors 31 (5), 519–537.

Huggler, J., 2014. Frankfurt airport security 'failed to detect 50 per cent of dangerous weapons', finds undercover inspection. https://www.telegraph.co.uk/news/worldnews/europe/germany/11308518/Frankfurt-airport-security-failed-to-detect-50-per-cent-of-dangerous-weapons-finds-undercover-inspection.html.

I. Guide, 73, 2009. Risk Management-Vocabulary. p. 551.

IATA, 2012. Checkpoint of the Future - Blueprint 2014, Report.

ICAO, 2017. Aviation Security Manual (Doc 8973 – Restricted). ICAO, Montreal, Canada.

Janssen, S., Sharpanskykh, C.R., Alexei, Langendoen, K., 2019a. Aatom: an agent-based airport trminal operations model simulator. Proceedings of the 51st Computer Simulation Conference, SummerSim 2019, Berlin, Germany, July 22–14.

Janssen, S., Sharpanskykh, A., Curran, R., 2019b. AbSRiM: an agent-based security risk management approach for airport operations. Risk Anal. 39 (7), 1582–1596.

Jesus, M.S., 2018. Trade-off analysis between security and efficiency of airport operations. http://resolver.tudelft.nl/uuid:35b2b53f-09b9-438e-b429-499a890f4643.

Kerley, D., Cook, J., 2017. Tsa fails most tests in latest undercover operation at US airports. https://abcnews.go.com/US/tsa-fails-tests-latest-undercover-operation-us-airports/story?id=51022188 (accessed: 2018-09-30).

Kirschenbaum, A.A., 2013. The cost of airport security: the passenger dilemma. J. Air Transp. Manag. 30, 39–45.

Kirschenbaum, A.A., 2015. The social foundations of airport security. J. Air Transp. Manag. 48, 34–41.

Kirschenbaum, A.A., Mariani, M., Van Gulijk, C., Lubasz, S., Rapaport, C., Andriessen, H., 2012a. Airport security: an ethnographic study. J. Air Transp. Manag. 18 (1), 68–73.

Kirschenbaum, A.A., Rapaport, C., Lubasz, S., Mariani, M., Van Gulijk, C., Andriessen, H., 2012b. Security profiling of airport employees: complying wih the rules. Journal of Airport Management 6 (4), 373–380.

Nie, X., Batta, R., Drury, C.G., Lin, L., 2009. Passenger grouping with risk levels in an airport security system. Eur. J. Oper. Res. 194 (2), 574–584.

Osman, M., 2010. Controlling uncertainty: a review of human behavior in complex dynamic environments. Psychol. Bull. 136 (1), 65.

Ratcliff, R., McKoon, G., 2008. The diffusion decision model: theory and data for two-choice decision tasks. Neural Comput. 20 (4), 873–922.

Sharpanskykh, A., Haest, R., 2016. An agent-based model to study compliance with safety regulations at an airline ground service organization. Appl. Intell. 45 (3), 881–903.

Singh, A., Singh, S.K., Khan, S., 2016. Job characteristics model (jcm): utility and impact on working professionals in the UAE. Int. J. Organ. Anal. 24 (4), 692–705.

Skorupski, J., Uchroński, P., 2015. A fuzzy model for evaluating airport security screeners' work. J. Air Transp. Manag. 48, 42–51.

Skorupski, J., Uchroński, P., 2017. A fuzzy model for evaluating metal detection equipment at airport security screening checkpoints. Int. J. Crit. Infrastruct. Prot. 16, 39–48.

Stroeve, S.H., Blom, H.A., Bakker, G.B., 2013. Contrasting safety assessments of a runway incursion scenario: event sequence analysis versus multi-agent dynamic risk modelling. Reliab. Eng. Syst. Saf. 109, 133–149.

United States Government Accountability Office, 2013. TSA Should Limit Future Funding for Behavior Detection Activities. U.S. Government Accountability Office.

United States House of Representatives, 2011. A Decade Later: A Call for TSA Reform.

Wales, A., Halbherr, T., Schwaninger, A., 2009. Using speed measures to predict performance in X-ray luggage screening tasks. Security Technology, 2009. 43rd Annual 2009 International Carnahan Conference on, IEEE, pp. 212–215.

Wilson, D., Roe, E.K., So, S.A., 2006. Security checkpoint optimizer (sco): an application for simulating the operations of airport security checkpoints. Proceedings of the 38th Conference on Winter Simulation, Winter Simulation Conference, pp. 529–535.

Winter, J., Cora, C., 2015. Exclusive: TSAs secret behavior checklist to spot terrorists. https://theintercept.com/2015/03/27/revealed-tsas-closely-held-behavior-checklist-spot-terrorists/ (accessed: 2018-03-27).

Wu, P.P.-Y., Mengersen, K., 2013. A review of models and model usage scenarios for an airport complex system. Transp. Res. A Policy Pract. 47, 124–140.