Modelling and analysis of dynamic relations and trade-offs between security and efficiency in airport security operations

Thesis report

A. J. P. Knol

TU Delft

Modelling and analysis of dynamic relations and trade-offs between security and efficiency in airport security operations

Final thesis report

by

Arthur Knol

to obtain the degree of Master of Science
at the Delft University of Technology,
to be presented on Friday February 23, 2018 at 10:00.

Student number:	4170350	
Project duration:	May 15, 2017 – February 23, 2018	
Thesis committee:	Prof. dr. K. G. Curran,	TU Delft, chair
	dr. O. A. Sharpanskykh,	TU Delft, supervisor
	ir. O. Stroosma	TU Delft

TU Delft

List of acronyms

DMAT	Diffusion Model Analysis Toolbox
FPR	False Positive Rate
GSA	Global Sensitivity Analysis
ICAO	International Civil Aviation Organization
IED	Improvised Explosive Device
LEADSTO	Language and Environment for Analysis of Dynamics by SimulaTiOn
pax	Passengers
PM	Preliminary Model
PVT	Psychomotor Vigilance Test
RDM	Ratcliff Diffusion Model
RTHA	Rotterdam The Hague Airport
ROC	Receiver Operating Characteristic
RT	Response Time
S	Success
SDT	Signal Detection Theory
TPR	True Positive Rate
TVC	Threat, Vulnerability, Consequence

Contents

1

Introduction

In September 1970, a plan was announced that would mark the start of a new era in aviation travel. A series of aircraft hijackings in the 1960s led president Nixon of the United States to unveil a plan that would "deal effectively with piracy in the skies"[5]. Although at that time domestic hijackings were happening at a rate of more than two per month in the USA, no plan to ensure security of aviation had ever been executed before. One year later, ICAO formalized the plan and ever since then, all commercial aircraft passengers know what "passing through the security check" means.

Nowadays, the presence of security measures is indispensable in airport terminals. The 9/11 attacks in 2001 caused the most recent significant shock to the view on security in the aviation industry. Terrorists and criminals see aviation as a useful target for attack. Successful attacks on airports or aircraft can cause large numbers of fatalities and can have big economic impact. It is therefore undeniable that security measures need to be in place. But these security measures have a large impact on another important quality factor of airports: the efficiency of the operations.

This year at Schiphol was busier than ever before. During the holidays in May and the summer holidays, extremely large queues before the security check due to holiday traffic caused large waiting times and even missed flights[26]. Every passenger who has been in such a large queue may have thought: "What if I could pass the security check quicker?" If there was no security check, passengers could easily walk to their gate. But in order to ensure security in air traffic, every large international airport follows strict regulations, which require to check all passengers on forbidden items. Airports are constantly deliberating to ensure the security in the airport terminal and on-board of aircraft, while under pressure to have an efficiently performing organization. The recent passenger peaks during holiday seasons showed that airports have troubles in operating efficiently, while maintaining good performance in terms of security[26].

Clearly, there is a certain relation between the security and efficiency of airports. Decisions that managers make for airport security operations, for instance concerning passenger flows, often affect both performance areas. Airport security managers and operators make *trade-offs* between security and efficiency: focusing too much on security might be at the cost of the efficiency of the airport, while vice-versa the same is true. If these relations and operational factors are portrayed quantitatively, this would provide assistance to managers to make more informed decisions regarding their airport operations strategy. But can this relation be quantified? How exactly do the security measures influence the airport terminal efficiency? If a certain new security measure is taken, what will be the result of this change in terms of efficiency? And how will this affect the performance in terms security?

All of these questions are within the context of one subject: *the relations and trade-offs between security and efficiency in airport security operations*. The goal of this MSc thesis research is to identify and analyze these trade-offs and relations in the context of airport security operations.

The report starts with a literature review in Chapter 2. The next chapter is Chapter 3, in which the research objective and methodology are presented. In this chapter also the research questions and the scope are determined. The methodology from Chapter 3 describes that first a preliminary exploration will be performed, to map the possibilities of the research. This exploration is performed in Chapter 4. The following chapter is Chapter 5, in which the final model is designed. Experiments with the final model are set-up and analyzed in Chapter 6. Chapter 7 discusses these results and elaborates upon the implications of these results, while finally presenting recommendations

2

Literature review

In the literature review report "Analysing Security and Efficiency of Airport Terminal Operations: Literature review", previously performed research was identified that could be used as input for this thesis[36]. This chapter will provide an overview of the most important and influencing findings of that literature review report.

The first part of this chapter will be an introduction to the concepts of *security* and *efficiency*. The definitions of these two concepts and how they are measured in airport terminal context will be described in respectively Section 2.1 for security and Section 2.2 for efficiency. The next section considers trade-off theories: Section 2.3 the objective conflicts that occur between security and efficiency and discusses trade-off theories that describe these conflicts. The final section is Section 2.4. This concluding section provides background information about theories and models that can be used for implementation during the research.

2.1. Security risk and analysis methods in airport terminal context

In this section security risk and methods to analyze security risk are treated. As a start it is important that a clear definition of security risk is provided. This is done in Section 2.1.1. When the definition of security risk is known, methods to analyze security risk can be discussed. This is done in Section 2.1.2.

2.1.1. Definition of security risk

Security is defined by Merriam-Webster as "the quality or state of being secure, such as freedom from danger and freedom from fear or anxiety" [41]. The risk is than the description of the threat to the state of being secure. It is important to know the difference between the concepts *security* and *safety*. Matthias Springer describes the difference as follows: "Safety stands for accident avoidance, and security for crime prevention"[40]. Security risk is generally regarded as a function of the three risk parameters: threat, vulnerability and consequence. It can be represented as in equation 2.1[5].

$$\text{Risk} = f(\text{Threat, Vulnerability, Consequence}) \tag{2.1}$$

Equation 2.1 describes risk as a function of the nature of a Threat (the source of danger, the attack), the Vulnerability (probability of success, given the threat/attack occurs) and the Consequences (result of the threat/attack) related to a particular attack scenario. This is called the TVC framework[6]. Threat and vulnerability can both be defined in terms of likelihood of occurrence, or probability. Consequence is rather expressed in terms of potential impact or severity. The risk equation is often described in terms of the multiplicity of the three risk factors, as can be viewed in equation 2.2.

$$\text{Risk} = \text{Threat} \times \text{Vulnerability} \times \text{Consequence} \tag{2.2}$$

From equation 2.2 could be derived that if one of the three risk parameters equals zero, the total risk is zero as well. Equation 2.2 should not be seen as a mathematical formula, but rather as a model to demonstrate a concept. The reason for this is that currently no neutral units of measurement exist for defining a threat, vulnerability or consequence. Furthermore, the three risk factors have a nature to be interdependent, which makes it even harder to compute them as single elements[61]. There is however much literature that describe methods risk can be analyzed. These will be elaborated upon in Section 2.1.2.

2.1.2. Security risk analysis methods

Security risk can not be computed in a straightforward way. Frameworks exist to help analyzing security risk. To be able to investigate the dynamic relation between security risk and efficiency, it is important to have a good understanding on how risk is analyzed. In this section the well-known TVC framework for analyzing risk is introduced. After having introduced the TVC framework, its shortcomings will be discussed and alternative risk analysis methods will be proposed in this section.

TVC framework

The TVC framework provides a good starting point for analyzing risk. The three factors of the TVC framework will be elaborated upon here.

Threat

The Oxford Dictionary[14] has two definitions for the word threat: "A statement of an intention to inflict pain, injury, damage, or other hostile action on someone in retribution for something done or not done" or "A person or thing likely to cause damage or danger". A threat in the context of an airport terminal can thus be seen as one or multiple persons who intent to inflict harm or loss to the terminal or the passengers in the terminal.

When analyzing the total threat to a system, one needs to analyze two things: what are the potential threats (threat identification) and what is the likelihood of each threat (threat likelihood estimation).

To identify potential threats for airport terminals, one can take a look at historical attacks or other events that occurred at airport terminals, or at comparable locations in which masses of people congregate, like train stations or concert halls. It is however very hard to be complete when identifying potential threats, since opponents will always try to come up with new attack means or methods.

The likelihood of a threat is in most cases out of control of the system on which the potential attack is performed. It is mostly influenced by the threat actor(s). According to van Impe[61], the rating of the threat actor is dependent on a number of factors, among which:

- The skill level of the attacker;
- The motive of the attacker;
- The opportunity: is the attacker in the possession of the required capabilities, the necessary access and the financial resources?

An example of a very simple, absolute method to determine the likelihood of a threat is by using data with information of terrorist attacks from the past. The likelihood of a certain attack can be determined through dividing the number of that type of attacks, by the amount of time in which these attacks happened. But this is a very limited method on which only few conclusions can be drawn.

Another way of analyzing the likelihood of a threat, is by determining the relative likelihood. This is often seen in the news, when a high threat level is announced[44]. This happens when one of the influencing factors for threat actors (skill, motive or opportunity) is estimated to be higher than normal.

Vulnerability

The vulnerability of a system is the probability that all defence measures in the security scenario fail[32]. Thus, if the threat happens, vulnerability is the probability that the attack is successful. Information and data of the vulnerability of a security system is mostly unpublished and classified. Five commonly used methods for analyzing the vulnerability of a system are the following:[60]

- Expert elicitation: by consulting security experts with inside knowledge of a system, a good basis can be formed for estimating the vulnerability in a threat scenario. However, a proper parameter estimation is difficult to determine in this way due to dependencies;
- Failure trees: the nodes of an event tree are events and the branches specify possible outcomes. A failure tree is a structured way to quantify the probability of specific series, but it is very dependent on the accuracy of the estimates;
- Vulnerability logic diagrams: a diagram shows the steps that an attacker takes to reach his goal and at each step the effectiveness of the countermeasures is graded as low, medium or high. The vulnerability

is then categorized on a predefined scale. This is not an exact method, thus has low accuracy, but it can help in providing first estimates and in prioritizing;

- Penetration testing: by testing the system with actors, accurate vulnerability estimates can be found. In penetration testing, corrections should be made to compensate for potential foreknowledge that the actor has. Furthermore, an actor can never act in the exact same way as an attacker would;

- Data analysis: although it is hard to find a direct measure that indicates the vulnerability of a system, there are derived metrics that give an indication of the performance with respect to vulnerability. An example is the true alarm rates: if a certain protocol change in the system leads to a higher ratio of true alarms, while all other factors have not changed, this can be regarded as an improvement in the vulnerability status[38].

As described, a vulnerability becomes concrete when a potential attacker finds out that there is one. This indicates the interdependence of vulnerability and threat probability: if a system is known to be vulnerable, there is a higher probability that an offender plans his attack on that system. For that reason, threat and vulnerability are sometimes taken together as "attack probability", or "attack likelihood"[61]. It is however important to note of the difference between a threat and a vulnerability: a threat is beyond control of the system, while the system can try to improve its own vulnerability.

Consequence

The consequence is the impact that a threat would have, if the attack succeeds. The total impact of an attack consists of the following two aspects:[6, 32].

- Direct losses: these are the losses which are directly related to the attack and of which the impact can be measured directly. It might include physical damages to assets, but also human fatalities and injuries;

- Indirect losses: these are the losses which are not a direct result of the attack, but which are secondary or even third level losses. These losses are sometimes related to the attack in a complex pathway. Examples are decreased number of future passengers, business disruptions and other economic impact. Indirect losses can best be quantified based on estimated direct losses and complemented with historic data.

The objective of consequence analysis is to quantify the potential impact of an attack. To do so, consequences are identified as the worst reasonable consequence that could be generated by a specific threat[3]. The consequence of an attack is thus strongly connected to the type of threat: for example, a bomb with a large radius will have a larger consequence than one with a smaller radius. However, by imposing adequate security policies, processes and procedures, the impact of an attack can be limited[61]. Consequence analysis should therefore be regarded as a different step than threat identification or threat likelihood estimation.

Limitations of the TVC framework

One limitation of the TVC framework is the fact that the three risk factors can not be multiplied with each other, because they often have large interdependencies. These interdependencies are very important to take into account, and the TVC framework is not particularly designed for doing so. In principle it is possible to capture dependencies by estimating conditional probability distributions for T, V and C. But this does not yet solve the fact that the interaction between threat, vulnerability and consequence is much more dynamic than captured by the TVC framework. There exist feedbacks between the three factors: the way choices that are made for one factor, affect the other factors[43]. The threat assessment used by the Department of Homeland Security of the United States, RAMCAP, is an example of such a step-by-step approach that makes use of the TVC framework and in which the dynamic interaction between the three factors is not incorporated[3].

Another important shortcoming is that the results of the risk analysis can not directly be used to optimally allocate defensive resources. Establishing valid numbers for threat, vulnerability and consequence is a time-consuming process, but when it is finished, the result does not help to provide an advice on how to protect the system against the threat in the best way.

Another critic is the fact that the TVC framework is too direct. Cox criticizes that "trying to directly assess probabilities for the actions of intelligent antagonists instead of modelling how they adaptively pursue their goals in light of available information and experience can produce ambiguous or mistaken risk estimates"[11]

Brown and Cox[9] go one step further and state that it is unjustified to use the TVC framework, as it can mislead risk analysts who are investigating terrorism. Attacker have different information than defenders do. Therefore, the values for Threat, Vulnerability and Consequence that attackers act upon will always differ from the values used by the defender. Even if the defender's experts are thoroughly trained, well calibrated, unbiased assessors; values calculated from different information sources, will always defer. It might therefore be a good idea to take a look into alternative security risk analysis methods, which have an approach that also incorporates the attacker's point of view.

Alternative security risk analysis methods

A potential solution to the direct and one-sided nature of the TVC framework is Game Theory. In the TVC framework for risk analysis, the "game" between attacker and defender is defined as follows: the allocation of defensive resources to protect potential targets is performed first by the defender. Next an attacker decides which targets to attack, estimating what the defender has done. In game theory, the probable consequences of the choice strategy of both attacker and defender are modelled, and the expected utilities of all probable consequences are assessed. In that way, game theory models can clarify the nature of the interacting decisions made by attackers and defenders. It can thus distinguish clearly between strategic choices (made by an attacker), and random variables (which are beyond control of the attacker and defender). This is in contrast with the TVC framework, in which attacker decisions are modelled as random variables or as uncertain attributes of threats. By combining it with the TVC framework, game theory can help to produce more effective risk management recommendations for allocating defensive resources[12]. The challenge however with game-theoretic approaches are that they require a lot of assumptions and tend to overrationalize strategies and choices[60].

Another approach to the risk analysis problem that has been initiated is agent based modelling of the airport security system[64]. Instead of rationalizing what possible outcomes are, the agent-based approach focuses on modelling the attacker and defender as different agents. These agents are modelled to have different autonomous behaviour, but also to coordinate their behaviour: they sometimes cooperate with other agents, and sometimes need to negotiate with other agents. In this way, underlying socio-technical processes can be modelled more realistically than when using the TVC framework[32]. By providing the agents with a certain strategy and placing them in a proper environment, interesting patterns on global level can emerge resulting from the autonomous behaviour and interactions of the agents on a local level. In agent-based modelling, cognitive and social models can be implemented, which results in more realistic estimations of security risks. For example, the above described game theory can be used as a negotiation process where agents are players and make moves[51]. This all makes agent-based modelling a promising, but not yet very well explored method for security risk analysis[60].

2.1.3. Security risks in airport terminal context

Section 2.1.1 and 2.1.2 describe security risk in a more general sense: how is it defined, and how can it be analyzed. In this section, the coupling to the context of the airport terminal is made. What are the threats in airport terminals? And what is done to reduce vulnerability and limit the consequence of a potential threat? These questions are answered in this section.

Threats on airports

The reason for necessity of security is the fact that there are parties that pose threats on airport terminals. Therefore, before analyzing security and efficiency, the root cause for why security is necessary on an airport should be investigated.

In the literature review by Knol, different threats were analyzed and were assessed on the relative risk that they impose.[36] The threat that imposed the most risk was considered to be the most interesting to start focusing on in the research. The different threats were subdivided into two groups: before the security check, in the so-called *non-sterile area* and after the security check, in the so-called *sterile area.* The assessment of relative imposed risk of the different threats was done based on the following criteria:

- Threat probability (T);
- Vulnerability (V);
- Consequence (C);
- Impact of security measures on vulnerability.

The following threat scenario was the most interesting one to investigate, following from the assessment:

A passenger tries to conduct a forbidden item through the security check

The success probability of this threat scenario is considered very high, the consequence significant, but most importantly: by varying security measures, the airport security system can have a large impact on the vulnerability to this threat scenario. It is chosen not to name the passenger "an attacker", since it is not known what the intentions of the passenger with the forbidden item are. To ensure security, every forbidden item should be filtered by the security check, despite of the intentions of a passenger or attacker. For this reasons, the above described threat scenario will be the focus of the investigation. The threat scenario will from now on be referred to as Threat Scenario under Investigation (TSI).

Security measures on airports

In the previous paragraph a threat scenario was chosen that will be used for investigation (TSI), after having assessed different scenarios on risk imposing criteria. This paragraph serves to provide an introduction about the countermeasures to this threat: the security measures. It serves to answer the question: what actions can airport managers take to increase or decrease the measures, that should help protect their passengers and their assets?

On airports, many security measures are present, visible and invisible. As already discussed, the airport can be divided into the non-sterile and the sterile area[56]. Examples of visible security measures in the non-sterile area are patrolling security agents or video surveillance.

To go from the non-sterile area to the sterile area, persons and luggage need to go through a security check. There are different checkpoints for passengers and for employees and other insiders of the airport terminal. In these checkpoints, security agents check if the person carries any object that does not comply to the rules of allowed objects in the sterile area. These rules have become more and more strict over the years and are mostly constituted as a reaction on a certain threat or attack[37]. The security agents perform these checks using equipment that can observe forbidden items in luggage or clothes. When necessary, people are searched manually by a security agent[5].

Next to the well-known security equipment like walk-through metal detectors and x-ray luggage screening, more specialized screening devices and procedures are available. These more elaborate devices and procedures could increase the security of airports to a large extent. An example is the introduction of X-ray body scanners instead of WTMDs, like Schiphol introduced as from 2010[22]. However, these are often costly and time-consuming devices, which results in longer processing times, increased operational costs and larger taskforce of security personnel[38]. It is therefore important for airport (security) management, to consider which security measures they should take to guarantee safety, while working in a time - and cost efficient way.

In literature, the security measures that can be taken by an airport are basically divided into four groups: Human factors, Facility & equipment, Responsibilities & procedures and Extra measures[55, 56, 67]. This subdivision is further clarified in Table 2.1.

Table 2.1: Security measure categories and examples

Security measure category	Subcategory	Examples of security measures
Human factors	Number of workers	# screeners, # screeners per shift, shift duration;
	Quality of workers	Basic quality, training quality;
	Motivation of workers	Work environment, wage level, stress, fatigue;
Facility & Equipment	Luggage screening	X-ray machine, explosives detector, manual search;
	Passenger screening	WTMD, hand-held metal detector, private search;
	Screening area	# security lanes, m^2 security check area;
Responsibilities & Procedures	Responsibilities	Airport authority, contracted screening company, government and policy makers;
	Procedures	Screening checkpoint configuration, Passenger distribution over day, Procedures for: pax screening, carry-on luggage screening, prohibited item screening;
Extra measures		Shatterproof glass, bomb sniffing dogs, Train all agents to SWAT teams.

Some of the examples of security measures from table 2.1 are easier to implement or change than others, and some are more expensive than others. The purpose of Table 2.1 is to provide a clear overview of what possible security measures that could be altered.

2.2. Airport terminal efficiency metrics and analysis methods

Although performing well in terms of security has a very high priority for airports and their management, it is not their only focus. Airport managers also have to consider the impact their security measures have on the efficiency of the airport, in order to have a smoothly operating airport terminal. Managers distinguish two types of efficiency: financial and operational efficiency[59]. Financial efficiency concerns e.g. profit margin and revenues for landing, parking and cargo handling/logistics. Operational efficiency refers to the practical aspects of airports: passengers handled, total terminal area, etc. Since the dynamic relation between security operations and the efficiency related to security operations is the main topic of this report, the main focus will be on operational efficiency.

Before considering the influence of the security measures on efficiency, it is important to understand how efficiency is measured and why it is important. Section 2.2.1 gives an answer to the latter, while Section 2.2.2 introduces metrics used to determine of airport terminal efficiency.

2.2.1. Importance of airport terminal efficiency metrics

The most basic definition of economic efficiency is: using the available resources (inputs) to maximize the production of goods and services (output)[52, 59]. Airports thus have a high efficiency if they produce a relatively large output (e.g. many handled passengers / flights), with a relatively small input (e.g. few employees or low cost). According to Scotti[50] there are four different stakeholders that will benefit from an efficient airport:

- *Airport management:* need to have efficient operations in order to stand strong in the competitive environment of airports;
- *Airline managers:* are interested in efficient operational activities of the airports they operate on;
- *Municipalities:* want to attract business and tourists to their region, and an efficient airport is a pré for that goal;
- *Policy makers* who design airport improvement programs can use information on efficiency to make optimal decisions about resource allocation.

To the four stakeholders identified by Scotti, one final important stakeholder that benefits from efficient airports can be added: the passengers. Efficient airports will eventually save passengers time and cost, which is to their benefit.

2.2.2. Airport terminal efficiency metrics

Many attempts have been made to quantify the operational efficiency of airports. In the performed literature review that preceded this research, many efficiency metrics to measure airport terminal operation performance and their outputs and inputs are introduced and reviewed[36]. The focus of this research will be the trade-off between security and efficiency of the airport when considering TSI. Therefore, the used efficiency metrics in these research will be those that are used to measure the performance of the security check system. These are subdivided into generic measures and alternative, more subjective measures.

Generic efficiency measures

The most basic efficiency performance measures that describe the efficiency of the security measures on an airport are: time to gate, average time spent in the security check and average time spent per activity in the security check (luggage drop, physical check, etc.)[13]. As a result of the processes within the security check, it is also of interest to consider the average and maximum queue length, and the average and maximum time spent in the queue for the security check[25, 63]. Except from only looking at the average time spent at a process, the distribution of time spent during the check can also indicate important factors. [35]. Long processing and waiting times can have as a consequence that passengers miss their flight. Although the cause of missed flights can very well be any other than the security check, an airport with a large amount of missed flights will be considered to be an inefficient one.

Another output can be found when concerning the alarm rates of both security employees and automated security equipment: airports want to decrease the false alarm rate, without compromising on probability of detection of real attackers [38]. One could for example change the sensitivity of a scanner, which probably increases the false alarm rate, but can also increase probability of detection of forbidden items[67]. When these measures are related to the type of flight, type of passenger or time of the day (peak or not), interesting conclusions can be drawn.

A final, currently more frequently used measure of efficiency is the level of service (LoS) of an airport[1, 29]. This concept is specified by IATA and is defined to be a combination of average waiting times and average space (surface area) per passenger. Based on these parameters, airport performance is measured by placing airports in one of the four categories: under-provided, sub-optimum, optimum, or over-designed. IATA provides guidelines on how to establish facilities in such a way that airports fall within the category optimum.

Alternative efficiency measures: subjective measures

In Section 2.2.1, it was explained that there are many different stakeholders that benefit from an efficient airport. An important stakeholder is the passenger: if a passenger can choose between two equally secure airports, he/she probably prefers the efficient one. The aspects mentioned in the generic efficiency measures are important for passengers. When choosing an airport to fly with, passengers will prefer an airport with low waiting times and want to avoid a high probability of missing their flight. These factors are objective measures that are equal to each passenger. However, for passengers there also exists something as subjective measures, which are different from person to person.

An example of such a subjective measure is passenger satisfaction: what is the opinion of the passengers about the processes that happen at the airport? A high passenger satisfaction of a certain airport might be reason enough for a passenger to prefer that airport. An example of a way in which passenger satisfaction can remain high under difficult circumstances, is by displaying the waiting times: people have less stress when they have an estimate of the time that they still have to spend in the queue[8]. Certainly with the continuously evolving security measures and the increasing crowds at airports, the opinion of passengers becomes more important to consider[23]. Also, an increased passenger satisfaction could lead to a better performance on other outputs: satisfied passengers may be less likely to complain and more likely to buy things at the airport.

Another example of a subjective measure that is more concerned with security, is people's perception of security risk[42]. This works in two ways. On the one hand, there is the perception of security for potential attackers. The picture that potential attackers have about the security system might be different from the picture that the airport terminal tries to put: if potential attackers think that a certain security component is more vulnerable, it is more likely that a threat will take place on that component. On the other hand, there is the perception of risk that passengers have. If passengers do not feel safe on an airport, there is a chance that they choose another airport or another means of transport. Airport terminal managers can compensate for this by changing the security measures such that people feel more secure on the airport.

2.2.3. Review of efficiency metrics

In Section 2.2.2 different efficiency metrics are set out. For this research an efficiency measure is required that is of high interest to airports, but also easy to measure objectively. The alternative efficiency measures are of very high interest to airports lately, as was found in discussions with Schiphol and RTHA. However, because to assess these properly a more psychological study should be preformed, it is preferred to measure airport efficiency in a more directly measurable quantitative way. Processing time and queuing time in the security check are directly measurable, quantitative measures. These measures are also of major influence to the LoS concept and to passenger satisfaction. Furthermore they draw a clear picture of the overall performance of the security check. Lastly, they are more tangible concepts than False Positives and LoS are. Therefore, security processing time and queuing time before security are regarded as suitable metrics for measuring efficiency performance of the airport.

2.3. Security-efficiency objective conflicts: the trade-offs

In the previous sections, security and efficiency in the context of airport terminal operations were explicated separately. But in this research the objective is not only to investigate security and efficiency, but also to get to the root cause of the objective conflicts between these two performance areas: what is the essence of the trade-off between security and efficiency within airport terminal operations?

Limited literature is to be found addressing the security and efficiency trade-off specifically, let alone

in context of airport terminal operations. Some literature can be found when one dives deeper into the fundamentals behind this trade-off. Based on work of Herbert Simon[53], Robert Hoffman[28] and David Woods[66], two fundamental trade-offs have been identified, that are concerned in the airport terminal operations security-efficiency trade-off. These trade-offs will be discussed in this chapter: Section 2.3.1 introduces the Acute-Chronic Goal Responsibility Trade-off and Section 2.3.2 presents the Speed-Accuracy Trade-off.

2.3.1. Acute-chronic goal responsibility trade-off

Within work systems, different roles and responsibilities are distinguished. These different roles have different subsets of goals. The agents in these roles cooperate when they share the same goal, but conflict when they have different goals. Fundamental or chronic goals, like security, tend to be sacrificed when an increasing pressure arises to achieve acute goals, "faster-better-cheaper" goals. In such a way, macrocognitive work systems[1] become blind to risks. Acute goals, like passenger throughput and cost savings, can be measured directly, while security is measured in the long run. Furthermore, chronic goals like security are harder to measure and act like "values". Hoffman and Woods name this the *"Acute-Chronic Goal Responsibility Trade-off"* and calls this family of empirical laws "Bounded Responsibility". It is the task of a manager to make sure that there is sufficient interaction between the different roles within a work system. If there is no interaction between the different agents that all have their own responsibilities and goals, these agents "tend to work at cross purposes in the face of goal conflicts".

In 2001 the Institute of Medicine of the United States suggested a new strategy for health care in the US[66]. Six goals needed to be achieved at the same time: the national health care should be - Safe, Effective, Patient-centered, Timely, Efficient and Equitable. The idea is to find best practises, "silver bullets": systems that simultaneously advance multiple goals and do not conflict with others. Woods describes however, with a dramatic example of a conflict between acute production goals an chronic safety risks, that such a system that finds simultaneous advancement, will never work. His example is the Columbia space shuttle accident in 2001. The investigation board of this accident found that the accident was the result of pressure on acute goals, which drew away attention and investments from chronic goals, like controlling safety risks. Hollnagel summarizes the conflict between these two goals, by commenting that: "If anything is unreasonable, it is the requirement to be both efficient and thorough at the same time - or rather to be thorough when with hindsight it was wrong to be efficient". To cope with this empirical problem, Woods suggests that when selecting systems, in order to advance all goals, chronic goals should always be put first, with a second concern for efficiency and timeliness of methods. Furthermore he states that the quality of a trade-off depends on two parameters: (i) the discrimination power; how well can someone make an objective judgment and (ii) placement of decision criteria: make sure that the placement of a decision criterion is dynamically matched with the assessment of changing risk and uncertainty.

From Hoffman and Woods, it can be concluded that the acute-chronic goal responsibility trade-off must be handled by management that properly divide the roles of different agents within a work system and ensure reciprocity. Furthermore, managers must focus on chronic goals, which in the airport terminal operations context are to be secure, passenger-centered and equitable. The second concern is in airport terminal operations to be effective, timely and efficient. To do otherwise would result in a tendency of higher valuation of immediate, easy-to-measure consequences, and thus an unintentional sacrifice of chronic values. When making decisions in the trade-off, managers should focus on an iterative process of making objective judgments, and make sure that their decision criteria are matched with possible implications for risk and uncertainty.

2.3.2. Speed-accuracy trade-off

After having treated an abstract, high level trade-off in Section 2.3.1, the other trade-off that will be treated in this chapter is a more low level one. The *"Speed-Accuracy Trade-Off"* concerns behaviour that people encounter daily: performing a certain task, with some time pressure. In almost every task that humans perform, they need to trade between accuracy and speed. If one performs a task too quickly, he probably solves the problem inaccurate. If some more time is taken, one can be more accurate in performing the task. Usually this trade is solved without complications, but when there is a limited amount of time, this trade-off becomes interesting[33].

[1] Macrocognitive work systems are systems designed to support interdependencies among humans and intelligent machines to carry out joint cognitive work, as sensemaking, replanning, mental projection to the future, and coordination[28]

When performing a task in which mistakes of any sort can be made, fast responses are associated with more mistakes and slow responses with fewer mistakes. This is the case because in a fast response, all relevant information has not been processed completely by the task performer. This is generally true for people. But the number of mistakes that are made for a given amount of time pressure differs from person to person.

When under the pressure of a limited amount of time, many people tend to focus on speed. To solve the problem quickly is then emphasized over the accuracy with which the problem is solved. When the focus is however too much on speed, this can lead to missing the goal of the task completely. And when one misses the goal of the task, it is no longer relevant that the task is performed quickly. But on the other hand, there are also people who perform their task too slowly and who can produce more and better results when encouraged to increase working pace.

Different aspects are of influence on the speed-accuracy trade-off. An obvious influencing aspect is the quality of the worker: if two persons perform a task within the same time, but one person makes fewer mistakes than the other person, the quality of the first person is higher. A less obvious influencing aspect, is the so-called "arousal condition" of a person. Drinking a cup of coffee for example, increases the arousal condition of a person, which makes the person shift to a quicker, but less accurate result. For different tasks, a certain optimal arousal condition can be found. A sprinter for example, will have a better performance if he is highly aroused, while a table pool player performs better at a lower arousal rate[45]. Other examples of aspects that influence the speed-accuracy trade-off are attention and stress[65].

In almost every task there is both a speed and an accuracy requirement. This is also true in airport terminal operations. The speed-accuracy trade-off is an example of one of the high-level trade-offs that managers need to deal with when concerning the acute-chronic goal responsibility trade-off that was described in Section 2.3.1. But also on low-level airport operations, the speed-accuracy trade-off comes into play. It is something security officers are faced with constantly. It is the task of x-ray screening security officers to choose for a bag if it needs to be directed to a bag checking security officer, or if the bag is cleared for boarding. X-ray screening security officers can not take unlimited time to check every bag completely and should, if possible, not send every bag to the bag checker, but on the other hand may never clear a bag with a prohibited item in it. The same accounts for body screening security officers: they can not take unlimited time, but may not miss any forbidden item. But this also applies to security personnel apart from the security check: video proctors or patrolling officers should survey as many situations as possible, while not missing any dangerous situation. The best approach is to perform the task as fast as possible, without sacrificing accuracy. But how to do so with different persons with different characters in the airport terminal work system is a challenge.

2.4. Agent-based modelling: autonomous and diverse modelling of agents in a dynamic environment

Agent-based modelling was briefly introduced in Section 2.1.2 as a promising alternative way of measuring security risk. One of the benefits of agent-based modeling is the ability to represent socio-technical factors of agents. Cognitive and social models can be used, resulting in diverse agents who can make decisions autonomously. Furthermore, the model is able to express the dynamics of the system, as the environment and the agents can change over time.

This research will use the features of agent-based modelling and simulation. But before being able to exploit these possibilities properly, an investigation needs to be done to literature that provides models of autonomous (dynamic) behaviour and properties of diverse types of agents. In this section, these possibilities are explored. Firstly two different concepts are presented that canb e used to model autonomous decision making. In Section 2.4.1, Signal Detection Theory is explained and the concept of ROC-curve is introduced. The second decision concept is the diffusion model, which is explained in Section 2.4.2. The following section is Section 2.4.3, in which a *fatigue model* is explained: a model that accounts for tiredness of operators is introduced which is capable of capturing the dynamics within the system. But not only operators should be modelled realistically; passengers are also diverse agents. To account for this, different types of passenger agents that are present in the airport terminal are discussed in Section 2.4.4.

2.4.1. Signal detection theory

Signal Detection Theory (SDT) is a theory that was developed in the 1950s and models decision making processes for detection decisions. The theory states that a certain stimulus is presented to the decision maker, upon which the decision maker response with either yes or no ("yes, the stimulus was present", or "no, the stimulus was not present"). It is however not always completely clear for the decision maker if the stimulus

was present or not. For example, some forbidden items may be very obvious and easily detectable, others may be very hard to detect. Also, not all allowed items are the same: e.g. luggage with only allowed items may contain an allowed item that appears to be a forbidden item. Signal detection theory accounts for this uncertainty by assuming that the sensory process (which transforms physical stimulation into internal sensations) has a continuous output based on random Gaussian noise. When the signal is present, the signal is combined with that noise. By assumption, the noise distribution has a mean, $\mu_n = 0.0$, and a standard deviation, $\sigma_n = 1.0$. The mean of the positive signal-plus-noise distribution, μ_s, and its standard deviation, σ_s are dependent on the sensitivity of the sensory process and the strength of the signal. The result is two Gaussian probability distributions, of which an example with $\mu_s = 1.0$ and $\sigma_s = 1.0$ can be viewed in Figure 2.1.

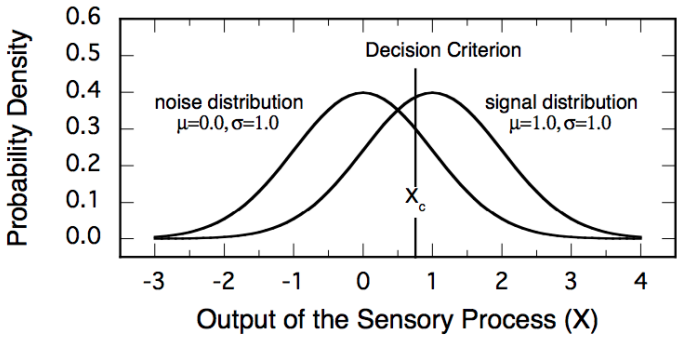


Figure 2.1: Gaussian probability functions of getting an output without (left Gaussian) and with (right Gaussian) a signal present

The vertical line in Figure 2.1 is the decision criterion, X_c, or a *threshold*. Outputs which are higher than X_c lead to "yes" as a response, outputs lower than X_c will have "no" as a response. It can be viewed that because of the noise, the items without a signal can receive "yes" as a response, since the right tail of the noise distribution contains probabilities lower than the X_c. This is called a False Positive (FP), or a Type I error. On the other hand, items with a signal may receive "no" as a response, since the left tail of the signal distribution contains probabilities higher than X_c. This is called a False Negative (FN), a miss, a vulnerability or a Type II error. An overview of correct and false responses of decision makers is given in Table 2.2

Table 2.2: Overview of calibrated parameters implemented in AATOM model

	Response = "Yes"	Response = "No"
Signal Present	True Positive / Hit	False Negative / Miss / Type II error
Signal Absent	False Positive / False Alarm / Type I error	True Negative / Correct Rejection

The described concept is a method to design a so-called Receiver Operating Characteristic, or ROC-curve. An ROC-curve describes the relation between the two operating characteristics True Positive Rate (TPR) and False Positive Rate (FPR) while modifying the threshold (X_c in Figure 2.1). In Figure 2.2 different ROC curves are displayed within one figure[49].

In Figure 2.2, security operator A has an FPR of 0.1 and a TPR of 0.9. A_2 is the same operator as A, but A_2 has chosen a lower threshold for threat. As a consequence, the operator will have a higher TPR ($\pm$ 0.95), but at the cost of a higher FPR (0.2). When comparing operator B with operator A, one sees that A is a much better performing operator than B: for the same TPR of 0.9, A has an FPR of 0.1, while B has an FPR of almost 0.8. One sees that the curve on which B can move its threshold is closer to the 0-curve than the curve of A. The 0-curve is the worst level on which that an operator can perform. On the 0-curve, FPR is equal to TPR, which means that the security operator can not distinguish forbidden passengers from allowed passengers and is just guessing randomly. The value 0 is the value for the sensitivity parameter d'. This parameter is defined as:

$$d' = z(\text{TPR}) - z(\text{FPR}) = \frac{\mu_f - \mu_a}{\sqrt{\frac{1}{2}(\sigma_f^2 + \sigma_a^2)}} \tag{2.3}$$

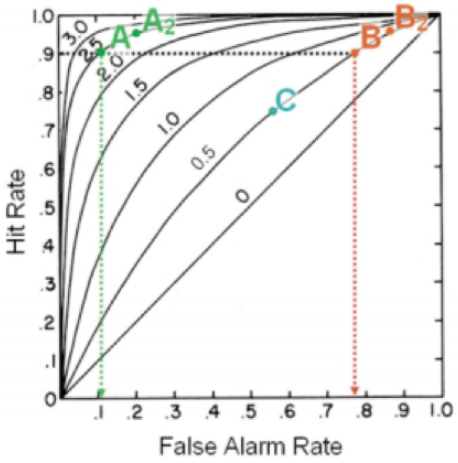


Figure 2.2: ROC-curves for different d'

As can be deduced from equation 2.3, a large value of d' corresponds to a large difference between TPR and FPR. This means the higher d', the better the security operator can distinguish between forbidden and allowed passengers, hence the better the performance of the security operator. A has a d' of 2.5, while B has a d' of 0.5. Operator C has different values for TPR and FPR, but has the same d' and thus has an equal performance when compared to operator B.

All security operators and sensors on airports can be modelled to have their own ROC-curve. The exact d' value for security operators or sensors at airports is non-disclosed, but it can be assumed that its range is between $0.5 < d' < 3.0$[5]. The value for d' can be distinct for different security operators based on their quality. Furthermore, the value for d' can be different for security operators over time. For example, training of security personnel results in higher values for d', and an analysis could be performed to the relative improvement.

2.4.2. Diffusion model for autonomous decision making

Many choices in daily life are decisions that need to be made between two possibilities. An example is: will I go to work by bike or by car. In such a choice, trade-offs are considered. Both choices have advantages. The decision is finally made when a collection of impulses and well considered arguments together form enough evidence to make the choice. And this decision is made within a certain amount of time, since one can not wait all day with choosing if he will travel by bike or car.

The considerations that security operators face during their work can be compared to the above described daily life situation. An X-ray scanner for example, is shown an image of carry-on luggage. After an (efficiently used) amount of time he needs to make a choice: is there enough evidence to send the carry-on luggage through to the luggage check operator, or can the luggage pass. Sometimes it is very obvious that an extra check should be performed (when he discovers a knife), or that no check is needed (when the luggage is empty). But sometimes the operator is in doubt, and he makes his choice based on certain impulses. The same is true for luggage checkers and physical checkers: based on observed impulses the operators decide whether there is enough evidence to confiscate the item.

A well-described method in literature to model this type of two-choice decisions is the *diffusion process model*[10, 46, 47, 54]. The diffusion model is often referred to as the Ratcliff Diffusion Model (RDM), after researcher Roger Ratcliff who is a major contributor to this model. The basic principle behind the diffusion model is the integration of accumulated noisy evidence over time[62]. If enough evidence has been gathered to point to one of the two options, the process stops and the output is a decision. The process of accumulation is determined by two forces: (i) a tendency to drift toward one of the two boundaries (drift rate, v), and (ii) a stochastic component determining the direction of the step to one of the boundaries and the size of this step. If enough evidence has been accumulated to reach the upper response boundary (threshold, a) or the lower response boundary (often 0 by assumption), then the decision is made. The decision can be biased: starting point z determines the proximity of the bias to either response boundaries. The diffusion process is visualized in Figure 2.3.

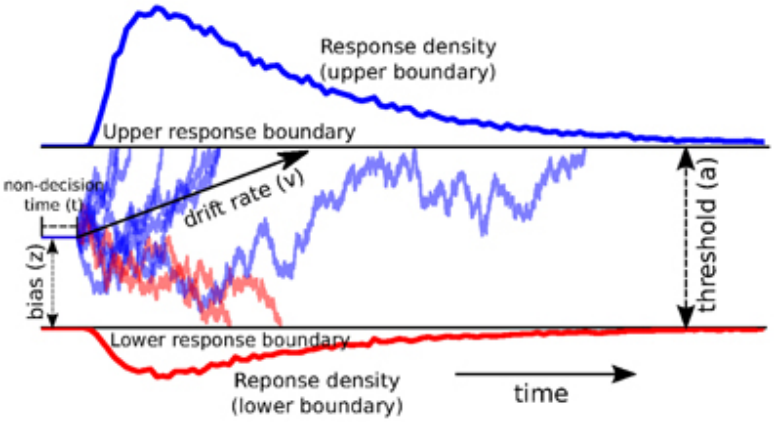


Figure 2.3: Graphical representation of arbitrary diffusion processes[57]

Next to the decision process determining variables already discussed (v, a and z), one other variable is visible in Figure 2.3: the non-decision time, from now on denoted as T_{er}. The non-decision time is the time it takes between the start of the decision and the actual start of the accumulation of evidence. T_{er} can be used to represent other processes involved, such as motor reaction time.

The discussed parameters do not necessarily have fixed values. It is common to assume that z, v and T_{er} have are variable between trials. A summary of the introduced diffusion process variables is given in Table 2.3.

Table 2.3: Seven free parameters of the Ratcliff Diffusion Model[62]

Parameter type	Sym.	Parameter	Interpretation
Decision process	a	Boundary separation	Speed-accuracy trade-off (high a means high accuracy)
	z	Starting point	Bias for either response ($z = a/2$ is neutral)
	v	Drift rate	Amount of input information; Quality of the stimulus
Non-decision	T_{er}	Non-decision time	Sum of all other processes involved (e.g. motor RT)
Intertrial var.	s_z	Intertrial range of z	Participant's variability in bias
	s_t	Intertrial range of T_{er}	Participant's variability in non-decision time
	η	Intertrial SD of v	Variability in stimulus quality (or attention / motivation)

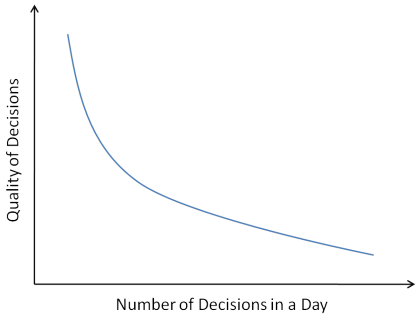
As described in Table 2.3, varying the boundary a can be used as a speed-accuracy trade-off. Furthermore, the drift rate v can be made dependent on the input information of the decision maker. Certain impulses can induce a high drift rate, while others imply a relatively lower drift rate. For the starting point, bias z, the following is true: the lower the bias z, the larger the amount of positive evidence that is required to come to a positive outcome of the decision process. This implies a relatively small chance on a positive and thus a larger chance on a negative outcome. For a high starting point, a high bias z, the inverse is true.

It can be viewed in Table 2.3 that the diffusion model contains many (seven) free parameters. In order to have the model work properly, these parameters need to be calibrated well. There is literature about algorithms that can be used for calibrating the diffusion model to experimental data. An example of such a calibration algorithm is DMAT, created by Vandekerckhove and Tuerlinckx (2008)[62]. When using this algorithm, the only required data is the response of the operator and the time in which he made the response, and the set-up of the experiment if the experiment consists of different set-ups. In their article, Vandekerckhove and Tuerlinckx provide a clear introduction and explanation of their calibration algorithm.

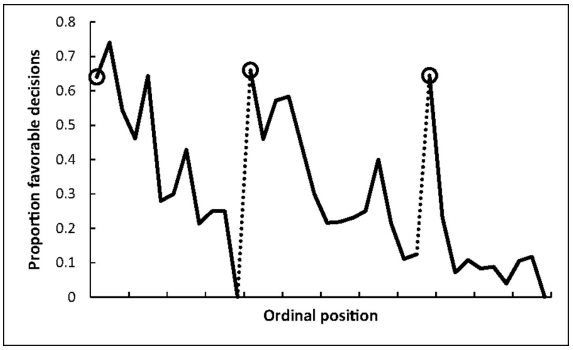
Concluding, the diffusion model can be used as a cognitive model for the autonomous decision making process of the agents. Furthermore it introduces the possibility to model diverse agents by specifying one or more of the parameters differently. When one or more of these free parameters are chosen to be variable over time, the cognitive model becomes a dynamic one. An example possibility that can be used to combine the diffusion model with is the fatigue model that will be described in the next section, Section 2.4.3.

2.4.3. Biomathematical fatigue model for dynamic agent behaviour

In research it was shown that every decision a person makes on a day, either if he chooses which shirt to wear or the decision to start a war or not, costs mental energy. Or in psychology language: decisions deplete willpower[58]. People only have a certain amount of mental energy on a day to make decisions. In psychology this effect is called *Decision Fatigue* [4]. The effect of depleting willpower or decision fatigue is visualized in Figure 2.4.



(a) Qualitative visualization of decision fatigue



(b) Favorable decisions of judge per time on day[58]

Figure 2.4: Graphs visualizing the concept of Decision Fatigue

Figure 2.4a is an example qualitative description of the decreasing quality of decisions as a consequence of an increasing number of decisions in a day. Figure 2.4b is a real track of decisions performed by an American judge during one day. On comparable cases, the decisions of the judge became less and less favorable for the convicted as the judge was working longer. The probability of approving parole fluctuates during the day and is lowest after a long session and just before break. [15]

Much research has been performed to the consequence of decision fatigue. In research at Stanford University, it was found that being mentally depleted leads to a reduced ability to make trade-offs[15]. Anderson published an article about decision fatigue leading to decision avoidance[2].

If there is one job in which many (important) decisions need to be made in a day, it is the job of security operator. Therefore it is very interesting to investigate how the decisions made on a day exhaust a security operator. The decision fatigue resulting from all these decisions may very well influence the performance of the security operator, and thus the performance of the airport security system. Therefore, implementing a fatigue model is an interesting research subject. And since agent-based modelling allows for this type of autonomous, diverse models in a dynamic environment, it can be implemented in the model.

McCauley et al. designed and validated a biomathematical model that accounts for the effects of sleep and sleep loss on waking neurobheavioral performance[17]. A set of seven differential equations model the varying performance of a person over time. The performance is dependent on the time the person woke up on the particular day and the sleep the person has had in the nights before that day. The performance is measured in terms of lapses a person has during a Psychomotor Vigilance Test (PVT). In this way, a "fatigue score" is assigned to the person on a certain point in time, indicating the performance of the person. A high fatigue score means a tired, and thus worse performing operator.

In a very recent paper, Walsh et al. (2017) use this fatigue score to implement fatigue in the diffusion model[39]. In their research the fatigue score determines the drift rate, and in that way steers the choice and decision time of a person during the PVT. They validated their results by performing experiments with humans. The fatigue diffusion model showed excellent goodness-of-fit to the human data: the model has shown to be able to predict complete distribution of the response times. Thereby they successfully combined the McCauley biomathematical model with the Ratcliff Diffusion Model.

Although the PVT is not exactly the same as a two-choice decision that security operators make, there are some similarities. In both cases the operator of the test needs to respond quickly, while preventing wrong responses. Furthermore, both the PVT and the two-choice decisions that security operators face, are vigilance tests: tests that require sustained attention and concentration and not too much cognitive capability. In vigilance tests the low required cognitive capability has as a consequence that the alertness of the person that performs the test reduces over time. And that is exactly what the above mentioned combination of models describes. Because of these similarities, the McCauley biomathematical fatigue model is a promising model to implement in agent-based modelling. In this way, use is made of the possibility of agent-based models to incorporate dynamic models.

2.4.4. Diverse passengers in airport terminals

The previous two sections considered the autonomy and diversity of security operators. But the security system of an airport is not only dependent on the properties and performance of the security agents. As Kirschenbaum states: To assume the “passive passenger” syndrome during this (security) process would be a grave error in judgment as the evidence points in the opposite direction"[35]. He states that passengers should not be seen as passive objects that must be delivered to their flights while all conducted in the same way through certain (security) processes. In the security check, the average screening time per "good" passenger is 20 to 30 seconds. But to work with such an average time, Kirschenbaum states, is not per definition useful. The reason for this is that there is a large differentiation between passengers: some problematic passengers will take 1 minute per person, while very problematic passengers can take 5 - 10 minutes per person. Four types of passenger behaviour are distinguished: (i) a passenger passing without incident, (ii) a passenger directly accepting the orders to remove prohibited items, (iii) a passenger who will "negotiate" shortly before accepting the orders or (iv) a passenger refusing or arguing to comply. Kirschenbaum concludes with contrasting the currently used "perfect scenario" with an "imperfect scenario". In the perfect scenario, all passengers require an average time of 20 to 30 seconds to pass the security check. In the imperfect scenario a small portion of passengers (± 10 - 20%) require either one or five minutes of time, while the other portion of passengers is quicker than 20 seconds. He finds that the very problematic, small group of ± 10 - 20% of the passengers that requires 5 minutes, costs the airport twice as much security screening time as the entire other group of good and little problematic passengers. The relevance of accounting for diversity of passengers has hereby been shown.

Kirschenbaum furthermore provides data on how different types of passengers behave differently in the security system. In the article, data is provided on the number of checks per different passenger types. The data shows that in charter flights, the hand luggage of every second or third passenger needs to be checked. At the same time in business flights, the hand luggage of only every seventh to ninth passenger needs to be checked. These data points can readily be used when implementing diverse passengers in an agent-based model, of which Kirschenbaum has shown the importance.

3

Research objective and methodology

As was described in the introduction, this research will focus on the analysis of the trade-off between security and efficiency in airport security operations. But one can start analyzing this trade-off, an outline should be drawn of how the research will be performed. This chapter serves to determine the research objective and proposes a methodology that will be followed to reach this objective.

In the first section of this chapter, Section 3.1 the research gaps that followed from the literature review in Chapter 2 are summarized. This is followed by Section 3.2, in which the research objective for this MSc thesis will be presented and a research question is formulated. After having formulated the research objective and research question, the scope within which the research will be performed in specified in Section 3.3. The chapter concludes by presenting a methodology in Section 3.4, that will be used to answer the research question.

3.1. Research gaps

Following form the literature review in Chapter 2, several research gaps can be identified to focus the research on. The research gaps are summarized in this section.

- **To use socio-technical modelling to represent agents making trade-offs:** the agent-based modelling paradigm provides the opportunity for socio-technical modelling. Using this, the diverse actors and the interactions between them can be represented properly;
- **To implement trade-off theories[1] in a model concerning airport terminal security and efficiency:** in literature about trade-off theories, interesting knowledge is found that was not earlier linked explicitly to the airport performance areas security and efficiency;
- **To capture in a single (agent-based) model the security-efficiency trade-off in airport security operations context and use it for decision making:** using agent-based modelling and simulation, new insights can be obtained that can help in making better decisions airport security operations;
- **To make the results of agent-based modelling experiments explicit and insightful for practical use:** experimental results on itself are not necessarily insightful for airport terminal managers. A tool can be constructed which explicitly represents the results in an insightful way, which will eventually help airport security managers to make better informed decisions concerning their strategy;

Furthermore, the literature study investigated the context in which this research could be performed. The threat scenario to be focused on was identified to be "A passenger tries to conduct a forbidden item through the security check", from now on referred to as TSI. Next to this, efficiency metrics that can be used to measure the performance of the security system of airport terminal operations have been identified. The efficiency will be measured in two ways: as the total time it takes a passenger to get through the security check, and as the average waiting time a passenger experiences before the security check. To measure security performance, a security risk assessment framework has been introduced that can be used as a guideline.

[1] Acute-chronic goal responsibility trade-off, Speed-accuracy trade-off

3.2. Research objective and questions

Having identified these research gaps, a research objective has been formulated that aims to fill in the above described research gaps. The objective of this research is defined as:

To identify and analyze the **dynamic relations** and **trade-offs** between **security and efficiency in airport security operations**, by applying an **agent-based modelling** approach

To reach this objective, a research question has been formulated. Answering this research question helps in reaching the objective of the research. The central research question in this research is:

How could the dynamic relations and trade-offs between security and efficiency in airport security operations be identified and analyzed, by applying an agent-based modelling approach?

To be able to better answer this research question, it was divided into four subquestions. These subquestions can be found below.

1. *What are influencing factors of security and efficiency that should be taken into account?*
2. *How can the influencing factors from (1) be modelled in an agent-based simulation model, such that emerging patterns can be identified?*
3. *How can trade-off theories be used to analyze and identify trade-offs between security and efficiency, measured respectively in vulnerability and processing time and waiting time?*
4. *Can a tool be developed for better informed decision making concerning airport security operations, using the model from (2)?*

The scope in which the research (sub)questions will be answered are further clarified in Section 3.3.

3.3. Scope of research

From the research questions it becomes clear that the research will focus on the performance areas of security and efficiency within the airport operations and the interaction between these two areas. These performance areas are broad concepts. Therefore, before starting to answer the research questions, it is important to determine the scope within which the performance areas of security and efficiency will be investigated. This scope will be elaborated upon in this section.

3.3.1. Security

In Chapter 2 a framework for assessing security risk was identified, called the TVC methodology. This methodology assesses security risk by identifying different threats and determining the threat probability (T), the vulnerability (V) and the consequence (C) of the threat.

In the literature review it was identified that one threat scenario, named TSI will be investigated. This threat scenario was defined as "A passenger wants to conduct a forbidden item through the security check". In this section it will be presented what a possible threat scenario assessment of TSI would look like.

1. Consequence assessment The consequence of a threat scenario is the impact that it would have if the attack succeeds. This impact consists of two aspects: direct losses (of which the impact can be measured directly) and indirect losses (second or higher level losses)[6, 32]. The objective of consequence assessment is to quantify the potential impact of an attack[3].

The consequence of TSI are the direct and indirect losses after a successful completion of the threat. If the forbidden item is successfully carried through the security check, the consequences are losses behind the security check (e.g. in aircraft). Since it is expected that the magnitude of this consequence is only to a small extent related to the efficiency of the security process, consequence assessment will be left outside of the scope of this research.

2. Threat likelihood assessment The threat likelihood is the probability that an attack actually takes place. Threat likelihood assessment is performed by special forces which investigate the probability that groups or individuals want to perform an attack. To assess this likelihood is regarded to be outside the scope of this research.

3. Vulnerability assessment The vulnerability of a system is the probability that all defence measures in the security scenario fail[32]. Or in other words: given that the threat occurs, vulnerability is the probability that the attack is successful. It is thus the aim of the security system of airports to keep their vulnerability as low as possible, and to catch as much forbidden items as possible (hits).

Information and data on the vulnerability of the airport security system is mostly classified for security reasons. But a lot of security measures are visible. The working principle of these measures is largely understood. It is therefore possible to model and assess these measures.

For TSI it is chosen to focus on one particular part of the vulnerability assessment, namely the security check. It will be checked if the security operators are able to find the forbidden item in the luggage or on the body of the passenger who caries the item. It does not matter if the passenger carrying the forbidden item is someone with bad intentions, or just an ignorant passenger who accidentally does not obey the rules. Because the intentions of the passenger with the forbidden item are not known, all forbidden items should be filtered by the security check. Every forbidden item that passes through the security check without being detected is thus a vulnerability of the system.

In this research, the vulnerability of the system is measured as follows:

$$\text{Vulnerability} = 1 - \frac{\text{\# forbidden items taken in security check}}{\text{total \# forbidden items presented at security check}} \tag{3.1}$$

4. Risk calculation As was described in the literature review, risk is calculated by multiplying threat probability with vulnerability and consequence. Since for the proposed investigation of TSI the threat probability and consequence are assumed to be outside of the scope, vulnerability is what determines the risk of TSI. A higher vulnerability will yield a higher security risk, and a lower vulnerability means a lower security risk. The performance in terms of security of an airport will thus be measured as the vulnerability of the security check system to TSI.

3.3.2. Efficiency

In the review of efficiency metrics in the literature review, it was explained that the most suitable measure for the efficiency performance of the security system are *processing time of the security system* and *queuing time before the security system*. The reason for this is that these metrics can be quantified in an objective way, measured directly, they are tangible concepts and they draw a clear picture of the overall performance of the security check.

3.4. Methodology

Having set the research objective and the scope, the methodology that will be used to reach the research objective can be presented. This is done in this section, by first introducing the environment in which the modelling and simulation will take place in Section 3.4.1, followed by an explanation on the model design framework in Section 3.4.2.

3.4.1. AATOM modelling and simulation environment

The central modelling and simulation environment in this thesis research is AATOM. AATOM stands for Agent-based Airport Terminal Operations Model and is a microscopic agent-based model that simulates movement and operations in the airport terminal[30]. It is designed to be a platform for performing studies to airport terminal operations. AATOM is created by PhD candidate Stef Janssen and is currently under development. This thesis research is one of the first researches that uses AATOM as an experimental environment. Therefore, next to using AATOM as an experimental environment, a significant part of this thesis research has also been dedicated to contribute to the development of the basis of the AATOM model.[2] In this section the benefits of AATOM will be discussed, followed by an explanation on how experiments will be performed within the AATOM simulator. For a detailed explanation of the model and all its constituent elements, the reader is referred to Janssen et al. (2017) [30].

[2] An important part of the research is the contribution to the AATOM model and the AATOM simulator. An overview of the contribution to the AATOM model and simulator is given in Appendix E

Benefits of AATOM

The benefits of using AATOM as a modelling and simulation environment are summarized below.

- **To ability to analyze emerging patterns on the global level by modifying local model parameters.** Airports are complex socio-technical systems, in which *interactions* happen between agents (social) and the environment (technical systems). These interactions are non-linear and complex. Because AATOM makes use of the agent-based modelling paradigm, it is capable of taking these non-linearities into account, resulting in a detailed and realistic model;

- **The possibility to implement complex social models.** Because the model is built up from the bottom, *autonomous* agents can be designed. These agents can have goals and are capable to reason make their own decisions. This distinguishes agent-based models from conventional models. And on airports autonomy of the agents is a very important factor: all security operators have different personalities and work in their own way (e.g. more focused on speed or more focused on accuracy). Furthermore, they get tired during the day which influences their performance. These are just a few examples of autonomous behaviour that can be modelled very well using AATOM.

- **The possibility to model diverse agents.** Conventional models tend to model passengers as "passive objects" which need to be led through the security check[35]. However, not all passengers are the same: e.g. some are quicker in dropping carry-on luggage at the security check than others. Passengers are thus not passive objects, they are *diverse*, autonomous agents. The same accounts for operators: some will get more tired during the day than others. Operators are also diverse, autonomous agents.

- **Scenarios can be tested that could not be tested in real practice.** Parameters can be adjusted without direct consequences. Airports would for example never experiment with a very low strictness of security checks, but it might be an interesting study to examine the consequence of such a modification for the efficiency of the airport terminal operations.

- **AATOM is an agent-based model in which all constituent entities of the airport terminal are represented**. In AATOM all elements of an airport that are required for this research are present. Among the modelled agents are passengers and security operators. The environment exists of all types of security equipment (WTMD, X-ray scanners, conveyor belts) and typical airport areas (security queues, gates) and any flight schedule can readily be implemented.

- **An AATOM simulator exists in which experiments can be performed.** An AATOM simulator, implemented in JAVA can be used for performing experiments and analyzing results.

AATOM can thus be used to modify parameters that influence airport terminal operations, and with the output of AATOM emerging patterns can be analyzed and KPIs of security and efficiency can be measured taking into account non-linear interactions. Next to this, the influence of diverse and autonomous agents can be modelled at a detailed level, which helps in better resembling reality. This all can be performed using a model that is specially tailored for airport terminal operations, and using a corresponding AATOM simulator that can be used for experiments.

Calibration of AATOM

Using the building blocks of AATOM, any arbitrary airport terminal can be represented. But all airports in the world are different and operate under different circumstances. Therefore, a choice should be made what type of airport will be modelled in AATOM. The TU Delft has good connections with Rotterdam The Hague Airport (RTHA), and this small international airport is open for sharing security and efficiency data with TU Delft. Furthermore, researches from TU Delft have been allowed to perform measurements and surveys at RTHA themselves. Because of this great opportunity of obtaining useful data, the choice has been made represent (the security check of) Rotterdam The Hague Airport in AATOM. The process of making sure that AATOM properly represents RTHA is called the *calibration* of AATOM.

During the different modelling iterations, the data provided by RTHA will be used to calibrate the model. But there are some baseline parameters that are equal for all models and case studies. These parameters are base conditions that describe the circumstances under which the airport terminal operates. These parameters are the *flights departing from the airport* (see Appendix B), *arrival distribution of passengers at the airport* and the *processing time distribution of a passenger dropping and collecting his luggage at the security check*. The calibration of these three parameters is given in this section, and is summarized in Table 3.1.

Table 3.1: Overview of calibrated parameters implemented in AATOM model

Parameter	Standard value	Variable range
Departure flight time schedule	RTHA, October 5th 2017, 05:00 - 10:00	-
Load factor of aircraft	0.9	-
Arriving passengers at $t_{non-peak_1}$	10%, from 1:40:00 - 1:20:00 $\sim \lambda e^{-\lambda x}$	-
Arriving passengers at t_{peak}	80%, from 1:20:00 - 0:40:00 $\sim \lambda e^{-\lambda x}$	-
Arriving passengers at $t_{non-peak_2}$	10%, from 0:40:00 - 0:20:00 $\sim \lambda e^{-\lambda x}$	-
# opened security lanes	3	-
$t_{luggagedrop}$	$N(54.6, 36.09)$	-
$t_{luggagecollect}$	$N(71.5, 54.95)$	-

Using a normal distribution for $t_{luggagedrop}$ and $t_{luggagecollect}$ induces the possibility that a negative value is drawn randomly from the distribution This would mean that that specific passenger has a negative processing time at the specific activity. If this occurs, this value is corrected to zero, resulting in $t_{luggagedrop}$ or $t_{luggagecollect}$ equal to 0 for that specific passenger. For all other time distributions described in the rest of the report, this truncation will be done in the same way.

Experiments in AATOM

As was explained in Section 3.2, the objective of the research is to investigate the trade-off between security and efficiency performance in airport security operations for a certain threat scenario. To investigate a threat scenario, AATOM will be extended by implementing the threat scenario in the model. Security operators are responsible of preventing this threat by filtering forbidden items in the security check. The decision making of the security operators will be modelled to a detailed level using (dynamic) cognitive models. Furthermore, diverse agents (passengers and operators) will be implemented in the model.

Having created an extended version of the AATOM model, the next step is to perform Monte Carlo simulations in the AATOM simulator, built in a Java environment. As was explained, the AATOM simulator is calibrated to RTHA, meaning that the terminal environment of RTHA is constructed. This layout can be viewed in Figure 3.1.

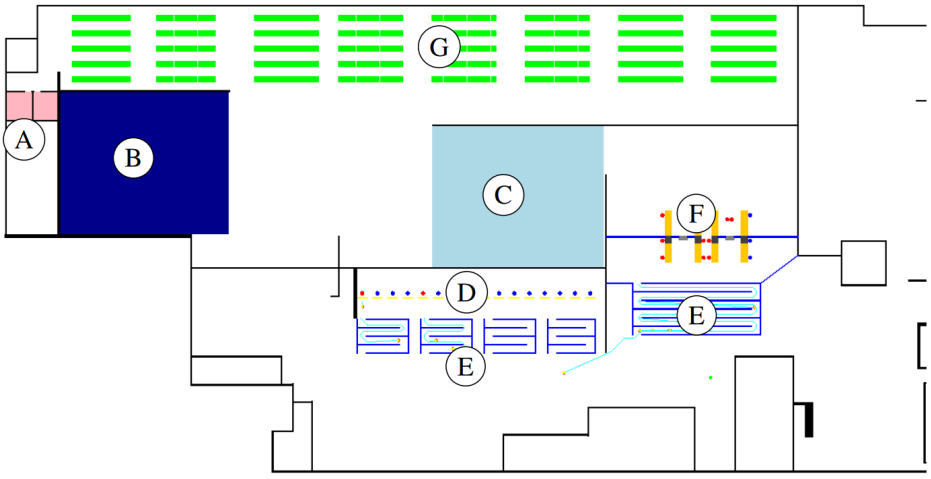


Figure 3.1: Terminal environment of RTHA modelled in the AATOM simulator, used for case studies. A, B and C are facility areas, D is the check-in area, E are queuing areas, F is the security checkpoint and G is the gate area.

When experiments are performed with different sets of parameters, for every set of parameters at least 100-250 simulations will be performed to reduce the random variations in the results.More simulations would be preferable, but one simulation costs approximately 7-8 minutes, and the available simulation time is limited.

The goal is to identify emerging patterns from the simulations. The specific sought-for emerging patterns are patterns that describe the relationship between security and efficiency. The found relations may behave differently for different parameter settings. To investigate the effect of the different parameter settings, sensitivity analysis is performed. Using sensitivity analysis, regions within the found relations will be identified

within which the behaviour of the system is homogeneous. The difference in behaviour between regions will be identified and analyzed. An example case study is to investigate an operator's focus on efficiency. What are the consequences in terms of security performance if an operator is focused completely on efficiency and not on security? And how will the efficiency performance change if the operator chooses to focus more on security?

But the sensitivity analysis will not only be performed by differentiating one parameter (e.g. security focus) within the trade-off. Using sensitivity analysis, it will also be analyzed how the trade-off holds in different circumstances. It could be possible that a relation is found, but once the circumstances of the environment are adapted, the relation behaves differently. For example: how does the relation, found in the above described example case study, change during different times of the day. If operators are more tired, how will this affect the identified trade-off? In AATOM it is possible to model different circumstances, which will help in providing a complete picture of the identified trade-offs. How these models are designed and investigated will be described in the following chapters. But first Section 3.4.2 will give a high level description of the steps that will be set for modelling and simulation in this research.

3.4.2. Modelling steps

During the research a general framework for the model design is followed. This framework can be found in Figure 3.2

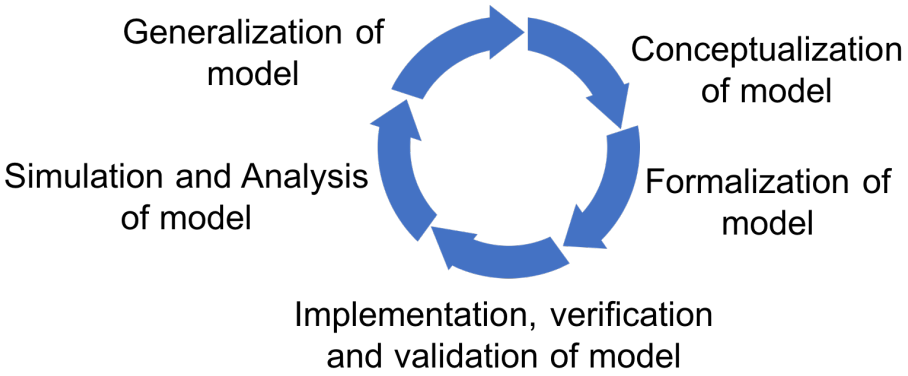


Figure 3.2: General methodology for model design

The methodology from Figure 3.2 is an iterative process. Once the model is designed, created, validated and simulations have been performed, one analyzes the results and generalizes these. In the generalization, one aims to find results that can be used for confirming a hypothesis or developing a theory. If this is not yet possible using the results, one tries to find in the generalization where the model needs to be improved in order to be able to confirm the hypothesis in the next iteration.

Because this study is one of the first researches to trade-offs in airport terminal operations, first a preliminary exploration will be performed. The goal of this preliminary exploration is to map the possibilities of the research and to determine which direction is most interesting. The preliminary exploration is done by performing one full iteration of the model design framework, consisting of five different modelling steps in Figure 3.2. This preliminary exploration will be performed in Chapter 4. After the preliminary exploration has been performed, the final model for this research will be designed in Chapter 5. The design of the model in Chapter 5 is a second iteration of the modelling spiral and comprises conceptualization, formalization and implementation, verification and validation of the final model. Chapter 4 and 5 serve to answer research subquestion 1 and 2 from Section 3.2. Using the final model, case studies will be set-up in which the behavior of the model is analyzed. These experiments and the results are described in Chapter 6. The generalization of the second iteration will be performed in Chapter 7. In this chapter a discussion that refers back to the main research question is presented by answering the final two research subquestions, 3 and 4. Furthermore the implications of this research are explained and recommendations for further research are given in this chapter. The report is concluded in Chapter 8.

4

Preliminary exploration

The preliminary exploration is done by performing one full iteration of the model design framework, consisting of five different modelling steps. This section is subdivided into five different sections that all describe a different component of the model design framework. The first step is to design the conceptual model. This preliminary conceptual model describes how the first analysis of the trade-off will be performed using a model and will be discussed in Section 4.1. This is followed by the formal model description in Section 4.2, which gives a mathematical description of the model using the formal language LEADSTO. In 4.3 the implementation of the model in AATOM simulator is treated. Here the verification and validation of the model in the AATOM simulator is performed, resulting in some interesting findings about the mismatch between reality and the simulator. The next section is Section 4.4, which treats the results of the experiments that were performed with the preliminary model. The final section is Section 4.5, in which the generalization of the preliminary simulation results is performed. This section answers the question: how can the preliminary results be used for further modelling?. What is still missing in the current results and what should be improved in the current model to obtain better results?

4.1. Preliminary conceptual model

The purpose of the preliminary model is to create a mapping of the investigation possibilities that are available, before making a choice on what aspect will be focused in the final model. In Chapter 3 it was described that the first focus will be on a threat scenario called TSI. The proposed conceptual model thus consists of the additions to the AATOM model that are necessary for implementing TSI into AATOM. These additions and the reasoning behind them are explained in this section.

The explanation of the conceptual model is performed in four steps. Step 1 is to explain the necessary additions to the AATOM model. Once the necessary additions have been introduced, the next step is to describe the interactions between the added entities. This is done by presenting a graphical representation of the conceptual model in Step 2. Having designed the model, Step 3 is to define what the outputs of the model will be in terms of security and efficiency, and how these will be analyzed. The final step, Step 4 is to color in the boxes: in this step the newly introduced parameters of the model will be calibrated, which makes the model ready for use.

4.1.1. Expansion of AATOM model

The necessary additions can be split into two different types of additions. The first addition is that there are *passengers with forbidden items and passengers with only allowed items* and the second addition is that there are *operators who need to distinguish between "forbidden" and "allowed" passengers.* These two types of additions will be treated in this step.

"Forbidden" and "allowed" passengers

TSI is a threat scenario in which someone with wrong intentions wants to conduct a forbidden item through the security check. To make sure that this threat does not result in a success, airports have constructed guidelines on checking for forbidden items that should be intercepted in the security check. The list of forbidden items contains clearly dangerous objects, like weapons and bombs. But terrorists can also make their own

bomb, called an Improvised Explosive Device (IED). IEDs can be made from material that on it's own is not necessarily considered dangerous. To prevent attacks with improvised weapons, the list of forbidden items also contains all kinds of daily used items, like bottles of water or bandage scissors. The consequence of this is that terrorists are not the only passengers that carry forbidden items with them in their pockets or in their luggage. Passengers without wrong intentions may also have objects that need to be intercepted in the security check.

For this model, two types of passengers are distinguished. *Forbidden passengers* are passengers that carry forbidden items when they pass through the security check. Independent of carrying the items accidentally or having intentions: all forbidden passengers should be filtered and their items should be confiscated. *Allowed passengers* are passengers who do not carry forbidden items when they pass through the security check, and who thus do not require extra checks.

In the model, forbidden items can either be in the *luggage* of the forbidden passenger, or on the *body* of the forbidden passenger (e.g. in his or her pocket, or directly connected to the body). In the preliminary model, forbidden passengers have a forbidden item both in their luggage and on their bodies. The forbidden items of forbidden passengers can thus be detected both in physical and in luggage checks.

In the model, the body and the luggage of the all passengers are assigned a certain *threat level*. The threat level of the body is denoted with p_t (short for *physical threat*, the threat level of the luggage is denoted with l_t (short for *luggage threat*. Forbidden passengers generally have relatively high threat levels, passengers generally have relatively low threat levels. For simplicity, in the preliminary model passengers get assigned the same threat level to their body as to their luggage.

Security operators and equipment distinguishing between forbidden and allowed passengers

Security operators have the responsibility to detect forbidden items at the security check. Passengers who carry a forbidden item on their body should be detected by the WTMD and their forbidden item should be found by the physical check operator. Carry-on luggage that contains forbidden items should be detected by the X-ray scanner and operator, and the forbidden item should be removed by the luggage check operator. At the same time, it is important that the security system only intercepts real forbidden items. The time spent on allowed items should be minimized, and thus allowed passengers should not have an unnecessary check by a physical check operator, and allowed luggage should not be unnecessarily checked by a luggage check operator. To do so in the best way and to make as few mistakes as possible, security operators are trained and security equipment is designed to intercept as much forbidden items as possible, while letting allowed items pass without superfluous checks.

In the preliminary model, signal detection theory is used as a theory that describes the detection process of security operators and security equipment. In order to distinguish between forbidden and allowed passengers, security operators get assigned a certain *thresholds for threat*, $thres_{\text{threat}}$ (corresponding to X_c in Figure 2.1). Using the security screening devices in the security check, the security operators examine the threat level of the body p_t and luggage l_t of the forbidden / allowed passenger and compare this threat level to their own threshold for threat. These probability of the threat levels that the passengers get assigned follow a Gaussian distribution. The probability distribution of the threat levels of allowed passengers follow the standard normal distribution by assumption, as usual in signal detection theory: $\mu_a = 0.0$, $\sigma_a = 1.0$ The probability distribution of the threat levels of forbidden passengers (μ_f and σ_f) are calibrated on data on airport security detection performance, as will be explained in Step 4.

If the threat level p_t or l_t is higher than the threshold $thres_{\text{threat}}$, the security operator / equipment will give "yes, the person / luggage is forbidden" as an answer, if p_t or l_t is lower than the threshold $thres_{\text{threat}}$, it will give "no, the person / luggage is not forbidden" as an answer.

The threshold for threat $thres_{\text{threat}}$ that is maintained for the security is *not fixed*. This means that the value of $thres_{\text{threat}}$ can be varied. A low value for $thres_{\text{threat}}$ will result in a high true positive rate (TPR): almost every forbidden passenger is caught. But on the other hand, this also results in a large number of false positives (FPR) and thus in inefficiencies. A low $thres_{\text{threat}}$ can thus be seen as a *security focused system*. Inversely, a higher threshold for threat results in a reduced number of false positives, but in an undesired increase of forbidden passengers that successfully bring forbidden items into the aircraft. This can be regarded as an *efficiency focused system*.

4.1.2. Graphical representation of conceptual model

To clarify the relations between all the different entities in the conceptual model that were described in Step 1, a graphical representation of the conceptual model of TSI is designed in this step. In this graphical represen-

tation, it is shown how the airport security system handles a passenger with a forbidden item. This graphical representation can be found in Appendix A.

The graphical representations are built from agents and objects. Every agent and object has the form of figure 4.1.

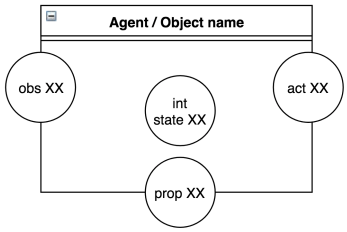


Figure 4.1: Conceptual agent or object

On the top of the square the name of the agent or the object is specified. On the left border of the square, circles with input states for the agent or object are defined. If the agent or object has observed something, this state is adjusted. The circles inside the square are internal states of the agents, which can be predefined or can be influenced by observations. On the lower border of the square, the properties or states of the agents are defined that can be observed by other agents or sensors. On the right border activity states can be specified. These are the outputs of the agents, resulting from observations and internal states. The lines and arrows between the different agents and objects define the relations between them.

The graphical representation clearly depicts dynamic relations between states of agents and the interaction between agents. It is therefore used as a guideline for constructing the formal model in Section 4.2. Next to this it also provides a clear overview of the conceptual model for the reader.

4.1.3. Security and Efficiency outputs and performance analysis

In Step 1 and 2 the preliminary conceptual model for security and efficiency has been introduced. The objective of this research is to construct a model for security and efficiency that is dependent on airport security operation parameters that influence both. To know the influence of these parameters, the output of the model needs to be measured. In this section it is presented how the security and efficiency outputs are measured, using simulations and analyses in AATOM.

Security output and performance analysis

As discussed in Chapter 3, the security risk is measured as the vulnerability of the system. In the mode no distinction is made between passengers who carry forbidden items and do not have bad intentions, and passengers who carry forbidden items and do have bad intentions. This means that to ensure security, the aim is to reduce the missed forbidden items. Vulnerability of the security system of the airport terminal described in the conceptual model can then be denoted as:

$$\text{Vulnerability} = \frac{\text{Missed forbidden pax}}{\text{Total forbidden pax}} = \frac{\text{Total forbidden pax - Hits}}{\text{Total forbidden pax}} = 1 - TPR \tag{4.1}$$

Forbidden passengers can be caught in different stages of the security check. This can occur in the walk-through metal detector and the subsequent physical check, in the ETD check or in the X-ray scan and the subsequent luggage check.

The vulnerability is thus equal to (1 - TPR) of the complete security system. As was described in Step 1, TPR is dependent on the distribution of the threat levels of the forbidden passengers, and on the chosen threshold $thres_{\text{threat}}$. This means that for the preliminary conceptual model, if the number of simulations n approaches infinity, TPR is deterministic, dependent on the threat level distributions of forbidden and allowed passengers, and on the chosen $thres_{\text{threat}}$. Vulnerability, and thus the security performance, can then be calculated analytically. The analysis of the security performance thus does not yet incorporate non-linear interactions or dynamic properties, which are the strengths of agent-based modelling and AATOM (as explained in Chapter 3). In a later stage of the research these properties should be explored more elaborately.

The analytically calculated values can however be used as a part of verification of the model in this stage by comparing them to the output for TPR of the simulation results with the AATOM simulator, as will be explained in Section 4.3.

Efficiency output and performance analysis

As was described in Chapter 3, it is most straightforward to quantify efficiency as throughput of the airport terminal system. The more passengers pass through the security check in the same amount of time, the higher the efficiency. This means that the shorter the average time it takes for passenger to reach the gates, the higher the throughput and thus the efficiency. Because the average time-to-gate can be easily split into the components of which it is composed, average time-to-gate per passenger will be used as the efficiency measure. It is composed of the four different processes that the passenger passes (check-in, security check, border control and possibly going to a facility), the times spent in the queue before these processes, and the walking time between these processes. The equation then becomes:

$$\text{Time-to-gate} = t_{check-in} + t_{Qcheck-in} + t_{security} + t_{Qsecurity} + t_{border} + t_{Qborder} + t_{facility} + t_{walking} \quad (4.2)$$

In equation 4.2, Q stands for queuing process before the described process.

In the current conceptual model, the implemented threat scenario will only have influence the processes in and around the security check. The research will therefore be focused on the implication of TSI for $t_{security}$ and $t_{Qsecurity}$. The security check is however not a single processes; it can be subdivided into separate processes that happen at the security check. This subdivision is formulated in equation 4.3.

$$t_{security} = t_{luggagedrop} + t_{WTMD} + t_{phyiscalcheck} + t_{luggagecollect} + t_{luggagecheck} + t_{walking_{security}} \quad (4.3)$$

Because the processes in security are dependent on the interactions between the passengers and security operators, this is a non-linear process of which the output can not be computed analytically. Therefore, the AATOM simulator will be used as the tool to get results for efficiency implications of TSI.

4.1.4. Calibration of expanded AATOM model

Having chosen the additions for the expanded AATOM model, specified the interactions and determined how these will be analyzed, a last step needs to be set before experiments can be performed with the model and the simulator. The additions explained in Step 1 add a couple of parameters to the model, which need to be calibrated in order to make sure that the experiments produce useful results. The parameters in question and their calibration are explained in this step. At the end of this step a table is presented in which all newly added calibrated parameters in this step are summarized (Table 4.2).

Threat level distribution of forbidden passengers: μ_f, σ_f

As was explained in Step 2, the probability distributions of the threat levels of passengers with allowed and forbidden items is determining for the system. It determines the number of hits and misses (influencing vulnerability) and the number of false positives and true negatives, influencing efficiency. The threat level probability distribution of both allowed and forbidden passengers is assumed to be a normal distribution, following signal detection theory. The relative difference between the noise and the signal distribution is determining for the behaviour of the system and not the absolute difference. Therefore, the noise distribution (read: threat level distribution for allowed passengers) can be assumed to be equal to the standard normal distribution (hence $\mu_a = 0$, $\sigma_a = 1$). Having fixed this, the signal distribution can be determined.

Because much data on security performance of airports is undisclosed, it has not been possible to retrieve data on hit rates (TPR) and false positive rates (FPR) directly from RTHA or other airports. In literature however data is found which is summarized in Table 4.1.[16, 18, 24, 27]

Table 4.1: Overview of calibrated parameters implemented in preliminary expanded AATOM model[16, 18, 24, 27]

	Security focused airport	Efficiency focused airport
Hit Rate (TPR)	97 %	45 %
False Positive Rate (FPR)	20 %	0%

This data can be used to calibrate the signal distribution (read: threat level distribution for forbidden passengers). Because calibrating to a value of 0 % for FPR is not possible using a Gaussian distribution, the FPR of an efficiency focused airport will be set to 0.1 %. Using an error minimization algorithm it could be determined that the threat level distribution for forbidden passengers will have a mean of $\mu_f = 2.949$ and a standard deviation of $\sigma_f = 1.12$. Verification and validation of these values will be performed in Section 4.3.

Threshold for threat for security system: $thres_{\text{threat}}$
As mentioned in Step 1, the threshold for threat $thres_{\text{threat}}$ can be varied. A low threshold means a relatively security focused system, and a high threshold means an efficiency focused system. However, since a realistic system is considered, there are boundaries up to which $thres_{\text{threat}}$ can be adjusted. These boundaries are determined by the data from Table 4.1. The lower boundary for $thres_{\text{threat}}$ is determined by the security focused airport from Table 4.1, the upper boundary for $thres_{\text{threat}}$ by the efficiency focused airport. It can be shown that using the above defined values for threat level distribution of forbidden passengers, and calibrating on the data from Table 4.1, that $0.842 < thres_{\text{threat}} < 3.090$. A threshold of $thres_{\text{threat}} = 0.842$ can then be seen as a 100% security focused system, a $thres_{\text{threat}} = 3.090$ as a 100% efficiency focused system. Between this boundaries the value for $thres_{\text{threat}}$ can be adjusted to the desired amount of % focus on security (or efficiency). As a standard value for $thres_{\text{threat}}$ the center between these boundaries is taken: $thres_{threat} = 1.966$, corresponding to a 50% focus on security.

Security operator performance: d'
As was shown in equation 2.3, d' can be computed using the values for μ_f, σ_f, $\mu_f a$ and σ_a. When doing so, the result is $d' = 2.778$, which fits in the range of $0.5 < d' < 3$, although rather high. A lower d' means a worse performing airport, and a higher d' means a better performing airport. To investigate what the influence is of competence of personnel or training, d' can be varied and the results can be analyzed.

Forbidden passengers ratio: $p_{forbidden}$
The number of passengers who contain forbidden items in their luggage and/or on their body is very determining for the performance of the security system. Many forbidden items result in extra checks, causing a worse performance regarding efficiency. And also, many forbidden items make the vulnerability number extra significant: the absolute number of missed forbidden items becomes higher. The forbidden passengers ratio $p_{\text{forbidden}}$ is the probability that a passenger is a forbidden passenger. Using the same data Kirschenbaum uses in his research to the cost of airport security[35] it was found that on average 20.67 % of the passengers require an extra check. Clearly, these checks can be either false positives or hits. Assuming the center threshold of $thres_{\text{threat}} = 1.966$, gives a FPR of 2% and a TPR of 81%. This gives that 76.4% of the checked passengers only contained allowed items, and 23.6% contained forbidden items. It will thus be set that $p_{\text{forbidden}} = 0.236$. As was explained in Step 1, for the preliminary model it is assumed that a forbidden passenger has a forbidden item both in his luggage and on his body.

Random check probability: $p_{\text{randomcheck}}$
Next to shifting the $thres_{\text{threat}}$ more towards a security focus, and to improve operator performance d', there is another, easy to configure way of increasing the probability of detection of forbidden items: random checks. Random checks are defined in this model as the execution of a luggage check or physical check, regardless of the result of respectively the x-ray operator or the WTMD.

If the airport security system performs no random checks, $p_{\text{randomcheck}} = 0$, and if the airport security system performs a luggage check and physical check on every passenger, $p_{\text{randomcheck}} = 1$. How airports exactly perform their random checks is non-disclosed, but increasing the number of random checks is a quick and easy security measure to improve the security performance. There is no direct indication that airports always perform random checks (except for the ETD check, which is left out of the scope). The standard value will therefore be $p_{\text{randomcheck}} = 0$. In the model it can however be varied for values of $0 < p_{\text{randomcheck}} < 1$.

Processing times of operator checks: $t_{\text{luggagecheck}}$, $t_{\text{physicalcheck}}$, $t_{\text{x-ray}}$, p_{box}
The processing times of the operator checks are of important influence on the efficiency of the airport security system. It is therefore important for the result of this research that these parameters are calibrated to values that are close to reality. Using data assembled at RTHA about the processing times in the security check, the parameters for processing times of operator checks ($t_{\text{luggagecheck}}$, $t_{\text{physicalcheck}}$, $t_{\text{x-ray}}$) were calibrated. In collaboration with another MSc thesis research that was being performed in parallel, it was found that $t_{\text{luggagecheck}}$ and $t_{\text{physicalcheck}}$ had a significant fit following the respective distributions of $N(104.67, 80.86)$ and $N(43, 20.96)$. For $t_{\text{x-ray}}$ it was not possible to find a distribution that fitted significantly to all the data points - the data was too scattered[1]. It was then found that the reason for the wide spread of the data for $t_{\text{x-ray}}$ could lie in the fact that different passengers use a different number of boxes, significantly influencing the amount of required time for the X-ray scan per passenger. Therefore a new calibration was performed, in

[1] The data on $t_{\text{x-ray}}$ received from RTHA indicated the total time to scan all the luggage of one passenger in the X-ray scanner.

which four distributions were fit to passengers who uses one, two, three or four boxes in the X-ray scan (no passengers used more than four boxes). Doing this, the fit is significant. The distribution processing times for different number of boxes can be found in Table 4.2. In the same Table, the probabilities for number of boxes are presented.

Table 4.2: Overview of calibrated parameters implemented in AATOM model

Parameter	Symbol	Standard value	Variable range
Threat level distribution allowed passengers	μ_a, σ_a	$N(0, 1)$	-
Threat level distribution forbidden passengers	μ_f, σ_f	$N(2.949, 1.12)$	-
Threshold for threat of security system	$thres_{\text{threat}}$	1.966	[0.842, 3.090]
Security operator performance	d'	2.778	[0.5, 3]
Forbidden passengers ratio	$p_{\text{forbidden}}$	0.236	-
Random check probability	$p_{\text{randomcheck}}$	0	[0, 1]
Luggage check processing time	$t_{\text{luggagecheck}}$	$N(104.67, 80.86)$	-
Physical check processing time	$t_{\text{physicalcheck}}$	$N(43.00, 20.96)$	-
X-ray check processing time, one box	$t^1_{\text{x-ray}}$	$N(10.28, 5.06)$	-
X-ray check processing time, two boxes	$t^2_{\text{x-ray}}$	$N(16.44, 9.08)$	-
X-ray check processing time, three boxes	$t^3_{\text{x-ray}}$	$N(20.82, 11.04)$	-
X-ray check processing time, four boxes	$t^4_{\text{x-ray}}$	$N(21.00, 11.98)$	-
Probability for use of one box at X-ray scan	p^1_{box}	0.317	-
Probability for use of two boxes at X-ray scan	p^2_{box}	0.485	-
Probability for use of three boxes at X-ray scan	p^3_{box}	0.188	-
Probability for use of four boxes at X-ray scan	p^4_{box}	0.010	-

4.2. Preliminary formal model

For formal specification and analysis of dynamic properties of the multi-agent system model that implements TSI, the formal language LEADSTO will be used. LEADSTO is a language that enables one to model direct temporal dependencies between two state properties in successive states, which are called dynamic properties[19]. The advantage of using LEADSTO is that it is an executable language. Before specifying the formal model, a short introduction of LEADSTO will be given below. For a further explanation of the structure and semantics of the LEADSTO language, the reader is referred to Bosse et al.[19].

If two state properties are defined, α and β, and variables e, f, g and h are specified as non-negative real numbers, then the following relation between α and β is a LEADSTO expression:

$$\alpha \rightarrow_{e,f,g,h} \beta \tag{4.4}$$

In equation 4.4 α, the antecedent, and β, the consequent, are state properties. The two state properties are linked by variables e, f, g, and h. This link is described by Bosse et al as follows:[19]

"If state property α holds for a certain time interval with duration g, then after some period of delay between constants e and f, state property β will hold for a certain time interval with duration h"

Another advantage of LEADSTO is that it can be easily depicted graphically. The above description of Bosse et al of the link between α and β is graphically depicted in Figure 4.2. Specifying the formal model will help during the implementation of the model in the simulator, which will be treated in Section 4.3.

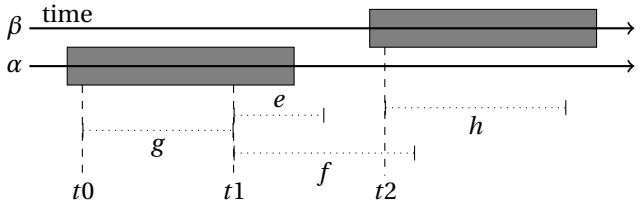


Figure 4.2: The timing relationships in LEADSTO

The used sorts in the preliminary formal model are:

Sort	Explanation	Elements
AGENT	a set of all agents	{operator, passenger}
OPERATOR	a set of all operators	{op_{xray}, op_{lug}, op_{phy}}
PASSENGER	a set of all passengers	{pax}
OBJECT	a set of all objects	{lug, WTMD, X-ray}
PROPERTY	a set of all properties	{agentproperty, objectproperty}
AGENTPROPERTY	a set of all properties of agents	{thres_{threat}, rc_{pax}, p_t}
OBJECTPROPERTY	a set of all properties of objects	{rc_{lug}, l_t}
LOCATION	a set of all locations	{l_{xray}, $l_{op_{xray}}$, $l_{op_{lug}}$, l_{WTMD}, $l_{op_{phy}}$}
INSTRUCTION	a set of all instructions	{perform_check(object / passenger)}
TIME	a set of linearly ordered time points	{t_{xray}, t_{lug}, t_{WTMD}, t_{phy}}

The used predicates are (in sequence of appearance in PM1 & PM2):

Predicate	Description
observed_own_location(operator, location)	The agent *operator* has observed its own *location*
observed_location_of(operator, location, object / passenger)	The agent *operator* has observed the *location* of an *object* or a *passenger* agent
output_random_check(object)	The output of the random check performed by an *object*
communicated_from_to(operator, operator, instruction)	The agent *operator* has communicated an *instruction* to another *operator* agent
set_property_of(property, object / passenger, boolean)	The *property* of the *object* or *agent* is set to *true* or *false*
property_of(property, object / passenger)	The *property* of an *object* or *passenger* is invoked
observed_property_of(operator, property, object / passenger)	The agent *operator* has observed the *property* of an *object* or *passenger* agent
perform_check(object / passenger)	Instruction:perform check at an *object* or *passenger*
arrested(object / passenger)	The agent *operator* arrests the *object* or the *passenger* agent

The used constants are (in alphabetic order):

Constant	Description	Range
$l_{op_{lug}}$	Location of op_{lug}	($\mathbb{R}$, $\mathbb{R}$)
$l_{op_{phy}}$	Location of op_{phy}	($\mathbb{R}$, $\mathbb{R}$)
$l_{op_{xray}}$	Location of op_{xray}	($\mathbb{R}$, $\mathbb{R}$)
l_t	Threat level of luggage	*see Table 4.2*
l_{WTMD}	Location of WTMD	($\mathbb{R}$, $\mathbb{R}$)
l_{xray}	Location of X-ray	($\mathbb{R}$, $\mathbb{R}$)
p_t	Threat level of passenger	*see Table 4.2*
p_{random}	Random check probability	[0, 1]
rc_{pax}	Passenger is randomly checked	boolean
rc_{lug}	Luggage is randomly checked	boolean
t_{lug}	Luggage check duration	N(104.67, 80.86) s
t_{phy}	Physical check duration	N(43.00, 20.96) s
t_{WTMD}	WTMD check duration	0.1 s
t_{xray}	X-ray check duration	*see Table 4.2*
thres_{threat}	Threshold for threat of operator	[0.842, 3.090]

PM1: X-ray sensor, X-ray operator and luggage check operator performing luggage check

PM 1.1: Determining random check for luggage

observed_own_location(X-ray, l_{xray}) & observed_location_of(X-ray, l_{xray}, lug) →
$_{[0,0,1,1]}$ output_random_check(X-ray)

PM 1.2: Output random check

output_random_check(X-ray) →
prob(p_{random}, ($_{[0,0,1,1]}$ communicated_from_to(op_{xray}, op_{lug}, perform_check(lug)) &
$_{[0,0,1,t_{end}]}$ set_property_of(rc_{lug}, lug, true))

PM 1.3: Observing threat level of luggage in X-ray

observed_own_location(op_{xray}, $l_{op_{xray}}$) & observed_location_of(op_{xray}, l_{xray}, lug) &
observed_property_of((rc_{lug}, lug) & rc_{lug} = false → $_{[0,0,1,t_{xray}]}$ observed_property_of(op_{xray}, l_t, lug)

PM 1.4: Information transfer between X-ray operator and luggage check operator

observed_property_of(op_{xray}, l_t, lug) & property_of($thres_{threat}$, op_{xray}) & $l_t \geq thres_{threat}$ →
$_{[0,0,t_{xray},1]}$ communicated_from_to(op_{xray}, op_{lug}, perform_check(lug))

PM 1.5: Observing threat level of luggage in luggage check

communicated_from_to(op_{xray}, op_{lug}, perform_check(lug)) & observed_own_location(op_{lug}, $l_{op_{lug}}$) &
observed_location_of(op_{lug}, $l_{op_{lug}}$, lug) → $_{[0,0,1,t_{lug}]}$ observed_property_of(op_{lug}, l_t, lug)

PM 1.6: Arrest luggage with excessive threat level

observed_property_of(op_{lug}, l_t, lug) & property_of($thres_{threat}$, op_{lug}) & $l_t \geq thres_{threat}$ →
$_{[0,0,t_{lug},1]}$ arrested(op_{lug}, lug)

PM2: WTMD sensor and physical check operator performing physical check

PM 2.1: Determining random check for passengers

observed_own_location(WTMD, l_{WTMD}) & observed_location_of(WTMD, l_{WTMD}, pax) →
$_{[0,0,1,1]}$ output_random_check(WTMD)

PM 2.2: Output random check

output_random_check(WTMD) →
prob(p_{random}, ($_{[0,0,1,1]}$ communicated_from_to(WTMD, op_{phy}, perform_check(pax)) &
$_{[0,0,1,t_{end}]}$ set_property_of(rc_{pax}, pax, true))

PM 2.3: Observing threat level of passenger in WTMD

observed_own_location(WTMD, l_{WTMD}) & observed_location_of(WTMD, l_{WTMD}, pax) &
observed_property_of((rc_{pax}, pax) & rc_{pax} = false → $_{[0,0,1,t_{WTMD}]}$ observed_property_of(WTMD, p_t, pax)

PM 2.4: Information transfer between WTMD and physical check operator

observed_property_of(WTMD, p_t, pax) & property_of($thres_{threat}$, WTMD & $p_t \geq thres_{threat}$ →
$_{[0,0,t_{WTMD},1]}$ communicated_from_to(WTMD, op_{phy}, perform_check(pax))

PM 2.5: Observing threat level of passenger in physical check

communicated_from_to(WTMD, op_{phy}, perform_check(pax)) & observed_own_location(op_{phy}, $l_{op_{phy}}$) &
observed_location_of(op_{phy}, $l_{op_{phy}}$, pax) → $_{[0,0,1,t_{phy}]}$ observed_property_of(op_{phy}, p_t, pax)

PM 2.6: Arrest passenger with excessive threat level

observed_property_of(op_{phy}, p_t, pax) & property_of($thres_{threat}$, op_{phy}) & $p_t \geq thres_{threat}$ →
$_{[0,0,t_{phy},1]}$ arrested(op_{phy}, pax)

4.3. Implementation, verification and validation of preliminary model

In this section the implementation, verification and validation of the preliminary model will be performed.

4.3.1. Implementation of model

With the preliminary conceptual and formal model specified, the model can be implemented in the AATOM simulator. The implementation consisted of the following parts:

- Distinguishing between forbidden and allowed passengers
- Assigning threat levels to forbidden and allowed passengers following the defined distributions
- Assigning a (variable) threshold for threat for security operators and equipment
- Allowing d' to vary between 0.5 and 3.0
- Allowing $p_{\text{randomcheck}}$ to vary between 0 and 1
- Adding an analyzer for the time spent in security check and queue
- Adding an analyzer for hit rate (TPR)
- Adding an analyzer for false positive rate (FPR)

4.3.2. Verification of model

Having performed these implementation steps, verification of the model should take place. The goal of the verification is to make sure that the conceptual model is implemented correctly in the AATOM simulator and that the model works according to the specification explained in the conceptual model.

A verification step has been performed for all additions. First the security performance of the model was verified. Forbidden and allowed passengers got assigned a different color in the model and with sample counting could be verified that the distinction was implemented well, and that $p_{forbidden} = 0.236$. Next the threat levels for both allowed and forbidden passengers were printed and a histogram was plotted, which followed the implemented distributions and were thus verified to work correctly. This directly meant that d' was correct, as d' is dependent only on the threat level distributions. Analyzers were added to count the number of random checks and the hit rate and false positive rate for both luggage and bodies of passengers. The comparison of the threat levels with $thres_{\text{threat}}$, done by the WTMD and the security operators was verified to be working correctly. This was done by printing the calculated difference by the WTMD or security operators and printing if the passenger was sent for an extra check (physical or luggage search). The output hit rates and false positive rates of both the x-ray + luggage search and WTMD + physical search had the same value. This was to be expected since the luggage and the bodies of all passengers were assigned the same threat level, and $thres_{\text{threat}}$ was equal for every operator / equipment as well. As a final step, the values for hit rates and false positive rates were compared to the analytically calculated values: as was explained in Step 3 of Section 4.1, the implemented signal detection theory also allowed for analytic computation of hit rates and false positive rates. The simulation result correspond to the analytically computed results, which means that the model is verified.

4.3.3. Validation of model

The validation of the efficiency performance was less straightforward. During the validation an interesting pattern emerged from the model. As was explained in Chapter 3, the distributions of the duration of the security processes were calibrated on data from RTHA. However, RTHA provided more data: RTHA also supplied information about the passenger throughput rate in the security check. It was stated that the throughput rate in the security check of RTHA lies between 2.5 and 3.0 passengers handled per security lane per minute[48]. Furthermore RTHA states that the mean is 2.6 passengers per lane per minute. The data for throughput rate can be used as a parameter to verify if the model works properly: if the simulations are run using the calibrated model, does the throughput rate match reality as well?

When implementing the processing time distributions of Table 3.1 and 4.2 and setting the variable parameters to standard values, the resulting throughput rate was found to be between 1 and 2 passengers per security lane per minute. This is clearly lower than the RTHA range of 2.5 - 3.0 passengers per security lane per minute. This means that the current state of the AATOM simulator is not yet sufficiently resembling reality:

it is currently impossible to implement realistic values for the processing times, and at the same time obtain realistic values as an output for security throughput rate. The reason for this could lie in some simplifications that are currently implemented in the security check of the simulator. For example, only three passengers can drop or collect their luggage at the same time. Furthermore, passengers can not yet move their luggage box to another table in order to make free space for following passengers. Furthermore, the physical interactions of passengers bumping into each other are consequences of the navigational model implemented in the simulator, and are not validated using data about walking patterns in airport security checks. These are some examples of simplifications in the AATOM simulator which could be the reason for the mismatch between the implemented processing time parameters and the output security throughput rate.

To solve for this problem, a correction factor CF is introduced in the preliminary model, which reduces all processing times with a certain number. CF is multiplied with the mean of the normal distributions of the processing times and CF^2 is multiplied with the standard deviations. In this way, the proportions between the processing time distributions remain the same, while the lower times ensure a realistic throughput rate. Using standard values of the calibrated parameters, the resulting throughput rate equals 2.6 passengers per security lane per minute when $CF = 0.65$. Another datapoint that was received from RTHA was the fact that during rush hour, with a throughput rate of 2.6 passengers per security lane per minute, a queue of approximately 100 passengers would arise. When CF is set to 0.65 in the simulator, the rush hour queue indeed contains approximately 100 passengers. This is another validation of the fact that this value of CF is correct. Therefore, $CF = 0.65$ will be the value that will be used in the experiments with the preliminary model.

4.4. Preliminary experiments, results & analysis

Having implemented, verified and validated the model, the preliminary experiments can be performed. In this section the set-up of the experiments will be discussed and the results will be analyzed. In the analysis also some attention is paid to validation of the model and the experiment results.

To use agent-based modelling in a newly developed simulator to analyze trade-offs in airport terminal operations is an innovative study. Because of the novelty of the research, it is important to first explore the possibilities for the research to determine which direction is most interesting. The goal of the preliminary experiments is thus to map the possibilities and study what the possible options for research are.

As preliminary experiments, three different preliminary case studies are performed which should lead to a better understanding of the possibilities of trade-off analysis with AATOM and the AATOM simulator. The results should lead to a mapping of the possibilities and could indicate which direction is most interesting to investigate further.

In the three preliminary experiments the relations and the trade-offs between security and efficiency are analyzed. The results of the experiments make these relations explicit and also show how it holds for the different simulation modes.

As a first experiment, the influence of the implemented signal detection theory (explained in Section 4.1) will be explored. It is expected that when varying the value of the threshold for threat, $thres_{\text{threat}}$, the focus will shift from security to efficiency or vice-versa. By doing so, the first experiment aims to find a first relation between security and efficiency in airport terminal operations. But $thres_{\text{threat}}$ is not the only parameter that influences the performance of the system. Another parameter that was identified in Section 4.1 to influence the performance of the system was operator performance, d'. The response of the system to a variable value for d' is therefore the second experiment. As a final exploratory experiment, it is chosen to not only dive into performance of the system, but explore a rather easy to implement security measure: random checks. The question is: what happens to the performance of the system in terms of security and efficiency when $p_{\text{randomcheck}}$ is set to be a variable? And how does the system behave differently when compared to preliminary experiment 1 or 2?

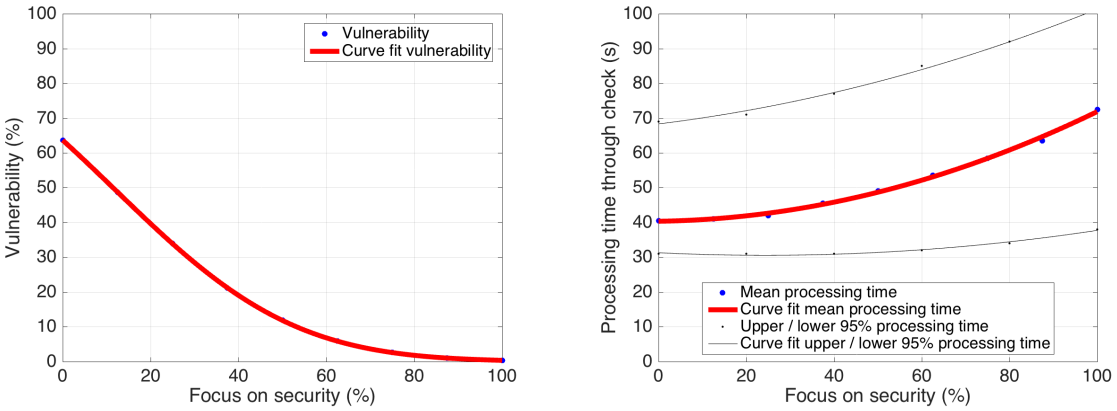
By performing these three experiments, it is aimed to have a proper mapping of the possibilities for the research to the trade-off between security and efficiency. This can later be used for the design of the final model, in Chapter 5.

4.4.1. Varying focus of system by varying value for threshold for threat, $thres_{\text{threat}}$

As was explained in Section 4.1, varying $thres_{\text{threat}}$ will have large influence on TPR and FPR. Nine different scenarios were set up based on nine different values for $thres_{\text{threat}}$, in a range between [3.090, 0.842] with a decreasing step size of 0.281. This corresponds to a respective focus on security from 0% to 100% with a 12.5% size increasing step (or a focus on efficiency from 100% to 0% with a 12.5% size decreasing step).

For all scenarios 100 Monte Carlo simulations are performed, which makes a total of 500 simulations. Every simulation approximately 1,000 passengers passed the security check, which results in a total of 500,000 data points about the time agents spent at the security check.

The result of the simulations can be found in figure 4.3.



(a) Vulnerability of airport system for different threat thresholds $thres_{\text{threat}}$

(b) Efficiency of airport system for different threat thresholds $thres_{\text{threat}}$

Figure 4.3: Efficiency and security performance for different threat thresholds $thres_{\text{threat}}$

In Figure 4.3a the analytically computed vulnerability is plotted against the focus on security, hence the threshold of the threat. It can be viewed that, corresponding to the implemented signal detection theory, the more focus on security (and thus the lower $thres_{\text{threat}}$), the lower the vulnerability of the airport security system. This is due to the fact that a relatively larger part of the threat level distribution falls below the threshold when the threshold is high.

When validating the results of Figure 4.3a, one finds that a 0% focus on security leads to a vulnerability of > 60 %. In Table 4.1 was however stated that realistic vulnerability on airports would not be above 55 %. The high vulnerability is in Figure 4.3a for low security focus is due to the structure of the system. When the X-ray operator and the luggage check operator both have a $thres_{\text{threat}}$ 3.09, they will both only catch 45 % of the forbidden items, resulting in a lower combined hit rate. On the other side of the spectrum, with high focus on security (low $thres_{\text{threat}}$), the same happens but then for extreme low vulnerability when compared to Table 4.1 One should thus be careful with drawing conclusions about regions with an extreme low or high focus on security.

In Figure 4.3b the efficiency performance is plotted, by plotting the processing time in the security check against the focus on security. It is seen that a higher focus on security results in more time spent in security. This relation is the strongest for an security focus of > 50%. For an security focus of < 50%, an increase in focus on security (decreasing $thres_{\text{threat}}$) hardly results in a decrease of security time. It is seen that the 95% confidence interval is reasonably large, which means that there is large variation in security time. As the value for processing time increases, the uncertainty increases as well.

When plotting processing time against vulnerability for a variable value for $thres_{\text{threat}}$, Figure 4.4 is obtained.

In Figure 4.4 a first explicit relation between security and efficiency is visualized. It is seen that when varying values of $thres_{\text{threat}}$, vulnerability is traded-off against efficiency. Decreasing the vulnerability is at the cost of security time, and vice-versa. This was to be expected, as this is mainly as a consequence of the implemented values of security focused airports and efficiency focused airports from Table 4.1.

It can be seen that if vulnerability is approaching 0, the function follows a vertical asymptote. Inversely, the relation also follows a horizontal asymptote when security time is decreased. This allows for the region of three typical regions. The first region is the almost vertical area in Figure 4.4. In this region a high value of efficiency performance can be gained at low security cost. At the other end of the spectrum, the opposite is true. The third region is almost horizontal, and a high value of security performance can be gained at low efficiency cost. In between there is a region in which significant amounts of security performance are traded against significant amounts of efficiency performance. This will be called the *critical trade-off region*.

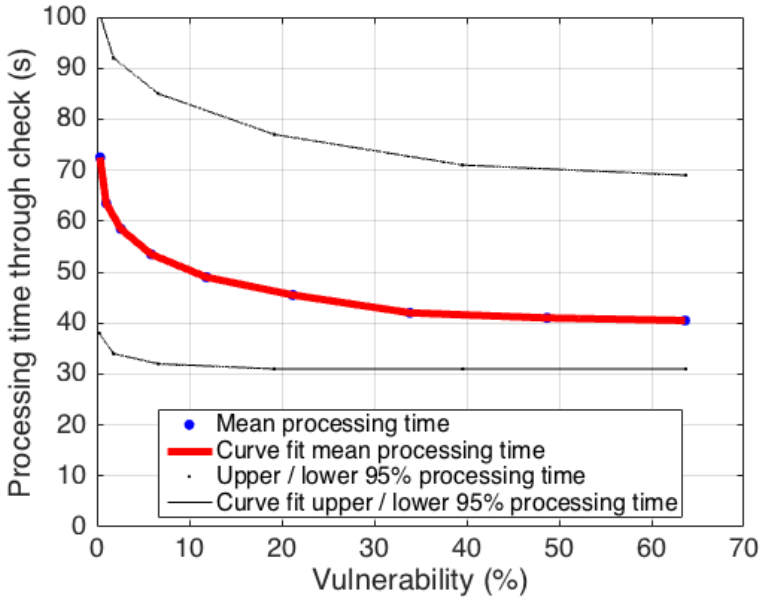


Figure 4.4: Efficiency of airport system vs. vulnerability for different threat thresholds $thres_{\text{threat}}$

In Figure 4.4 this region is approximately at 2.5 % < Vulnerability < 12 %. These values occur for 1.403 < $thres_{\text{threat}}$ < 1.966. In the final model design in Chapter 5 a further investigation will be done to these regions.

4.4.2. Variable security operator performance, d'

Experiments with a variable $thres_{\text{threat}}$ were performed, because it was expected that varying $thres_{\text{threat}}$ would be of key influence on the trade-off between security and efficiency. This does however not mean that $thres_{\text{threat}}$ is the only parameter that influence the security and efficiency performance. As was explained in Section 4.1, the hit rate (TPR) and false positive rate (FPR) are also largely dependent on the performance of the security operator, denoted with d'. Therefore, the next preliminary experiment will be carried out with setting d' to be a variable.

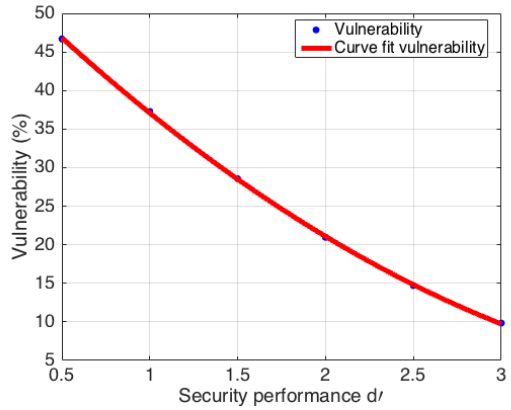
In Section 4.1 it was presented that the ROC-curve is used to model the performance of the security operators and security equipment. The ROC-curve describes how well distinction can be made between agents that contain forbidden items and agents that do not (allowed passengers). The parameter d' is a indicator of the performance of the security system. A good ability to distinguish between forbidden and allowed passengers results in a high d', a bad ability in a low one. For airport systems it can be assumed that approximately $0.5 < d' < 3$. [5].

As was done in the previous experiment, the influence of varying this parameter on security (vulnerability) and efficiency (security time) has been investigated. The results are visualized in Figure 4.5.

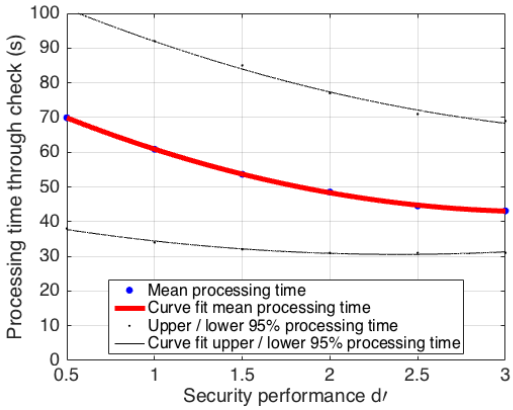
In Figure 4.5a it can be viewed that, ceteris paribus, increasing the security performance results in a lower vulnerability of the system. The difference between the worst performing system (d' = 0.5) and the best performing system (d' = 3.0) is significant: in the first case the vulnerability of the system is 47%, in the second case the vulnerability is only 10 %. Furthermore it can be noticed that, although an increase in d' always results in a lower vulnerability, the function is convex, meaning that for a larger d', an increase in d' will yield a relatively smaller decrease in vulnerability.

Figure 4.5b demonstrates that an increase in security performance also yields a decrease in security time. This can be explained by the fact that a higher d' results in a lower FPR, meaning fewer extra checks of passengers, and thus a smaller average waiting time.

The security and efficiency outputs are plotted against each other for varying values of security performance d' in Figure 4.6.



(a) Vulnerability of airport system for different security performance d'



(b) Efficiency of airport system for different security performance d'

Figure 4.5: Security and efficiency for different security performance d'

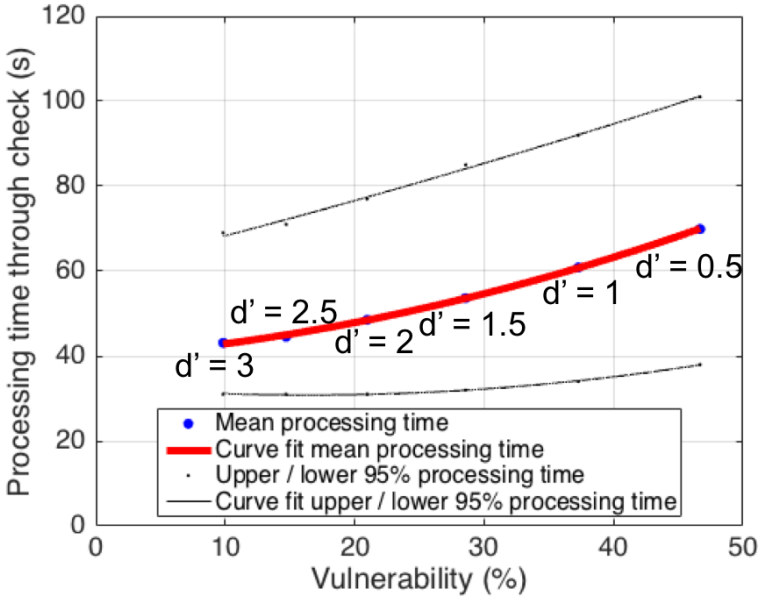


Figure 4.6: Efficiency of airport system vs. vulnerability for different security performance d'

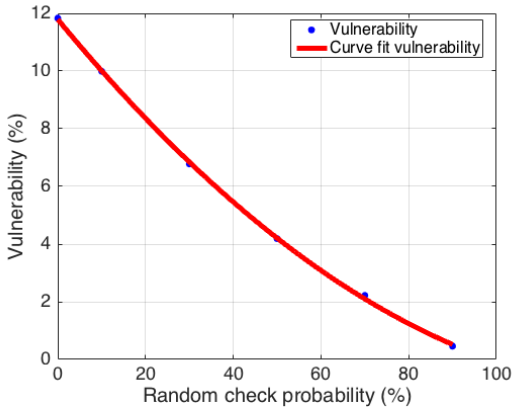
When regarding Figure 4.6, it is desirable for airports to have both security time and vulnerability as small as possible. In other words, the better the system is functioning, the closer it will get to the bottom left corner of the graph. It can be viewed that the larger d', the closer it gets to the bottom left corner of the graph. Furthermore it can be seen that although d' is increased with a constant step of 0.5, the space between the datapoints in Figure 4.6 decreases. This means that improving performance can be done relatively effectively when the performance is low. Lastly, a slight upward concavity can be noticed in Figure 4.6. This means that for high values of d', an extra increase in d will benefit more in terms of security performance than in terms of efficiency performance.

4.4.3. Variable random check probability, $p_{\text{randomcheck}}$

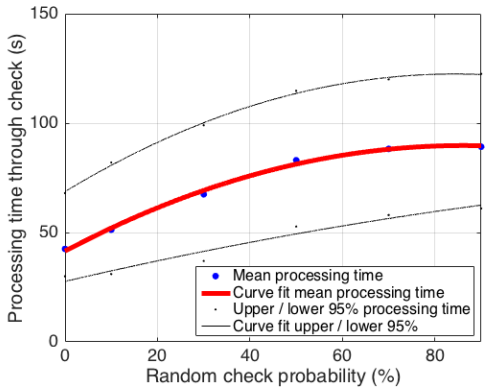
The final variable that will be investigated on its relation to security and efficiency is the random check probability $p_{\text{randomcheck}}$. The rationale behind the choice to investigate this parameter is the following. Security managers might not be able to change the threshold for threat, which was investigated in the first preliminary experiment. This might be the case because managers are not allowed to change the setting of the equipment, or because requesting a security operator to change his threshold is not possible in reality. Fur-

thermore, changing the security performance parameter d' is not something that can be done from one day to another: this requires resources, for example for new equipment or training sessions for personnel. Then next to $thres_{\text{threat}}$ and d', there is always one parameter that security managers and operators can vary: the probability of a random check.

Increasing $p_{\text{randomcheck}}$ will result in a higher FPR as more allowed passengers are checked. But it may also result in a higher TPR, since the random checks may intercept forbidden passengers that might have been missed by the regular checks. Simulations have been performed for six different scenarios for the random check probability, namely: $p_{\text{randomcheck}}$ = 0, 0.1, 0.3, 0.5, 0.7 and 0.9 Figure 4.7 shows the results of the simulations with varying $p_{\text{randomcheck}}$.



(a) Vulnerability of airport system for random check probabilities $p_{randomcheck}$



(b) Efficiency of airport system for different random check probabilities $p_{\text{randomcheck}}$

Figure 4.7: Security and efficiency for different random check probabilities $p_{\text{randomcheck}}$

In Figure 4.7a it can be seen that an increase in $p_{\text{randomcheck}}$ results in an almost linear decrease of vulnerability. Increasing $p_{\text{randomcheck}}$ from 0 to 0.9 reduces the vulnerability with a factor of 18. This is however at a cost. As was to be expected, Figure 4.7b shows that, ceteris paribus, an increase in random check probability results in a higher average security time. This line is however not linear, but concave: for high values of $p_{\text{randomcheck}}$, an extra increase of $p_{\text{randomcheck}}$ will hardly result in any extra expected security time.

The security time is plotted against vulnerability in Figure 4.8.

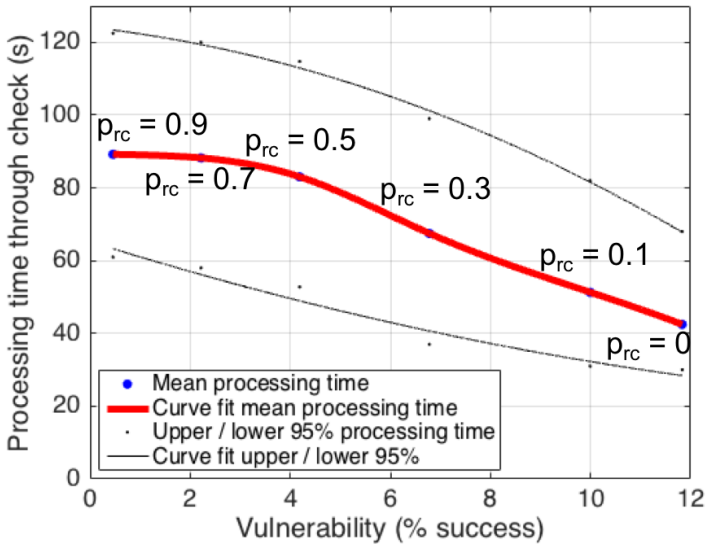


Figure 4.8: Efficiency of airport system vs. vulnerability for different random check probabilities $p_{\text{randomcheck}}$

In Figure 4.8 the relation between vulnerability and security time is visualized for varying values of $p_{\text{randomcheck}}$. It can be viewed that for low random check probabilities, vulnerability is traded-off against security time: every reduction in vulnerability is at the cost of extra security time. But a very interesting result is found for high values of $p_{\text{randomcheck}}$. If $p_{\text{randomcheck}} = 0.7$, a reduction in vulnerability can be achieved with increasing $p_{\text{randomcheck}}$ to 0.9, while hardly increasing the security time. Following this reasoning, an advice for airports that are suffering from high threat probabilities can thus be that if they are already performing a lot of random checks, they can increase the number of random checks at low costs: extra random checks will result in a further decreased vulnerability and hardly in extra times spent in security.

4.5. Generalization of preliminary results

When the conceptualization, formalization, implementation, simulation and first analysis of the model have been performed, the next step is to generalize the results. In this step the useful results from the experiments with the preliminary model are identified and further options for studying the security efficiency trade-off are suggested.

The generalization is split up into three parts. First the results of the preliminary experiments are summarized and some first conclusions are drawn. Next a Global Sensitivity Analysis (GSA) is performed, to analyze which of the parameters is of highest influence in the model when more parameters are set to be variable at the same time. The parameter of highest influence may namely be of interest to study into more detail in a next iteration. The final part of the generalization is the proposal for the next modeling iteration, in which a bridge is formed to the next section, Chapter 5, concerning the final model design.

4.5.1. Summarizing and concluding preliminary model experiments

In Chapter 3 it was presented that the research would focus on the analysis of the trade-offs and relations between security and efficiency. Using signal detection theory a model was formed in which the security system could focus either on security, on efficiency, or somewhere in between by varying a threshold for threat, $thres_{\text{threat}}$. Experiments using the AATOM simulator showed the influence of varying $thres_{\text{threat}}$ on the security and efficiency performance of the airport security system. It turned out that varying $thres_{\text{threat}}$ would trade security performance off against efficiency performance. The rate in which this happened however varied for different regions. In the one region a little decrease in security performance (vulnerability) directly resulted in a large improvement in efficiency performance, while in the other region further decreasing the security performance only yielded tenths of seconds of efficiency performance improvement. In between there is a region in which the trade-off is the strongest: significant amounts of security performance are traded off against significant amounts of efficiency performance. This is probably the region in which most airports want to find themselves: there is no quick win possible in terms of either security or efficiency performance. The airport is thus performing in an optimal region, and deliberate choices need to be made regarding the trade-off. Summarized, this experiment has given two interesting insights: (i) signal detection theory can be applied to airport security operations such that security performance is traded for efficiency performance and (ii) this trade-off happens at a varying rate, which could be divided into three typical regions. These insights will be used in the design of the final model in Chapter 5.

Next to modifying $thres_{\text{threat}}$ it was investigated what the influence of d' was on the security and efficiency performance. The results of this experiment can be very interesting for airport security managers: they may be curious to how a variable operator performance influences the security and efficiency performance of airport terminals. But for this research, to the trade-offs between security and efficiency, this is a less interesting area, since there is not a real trade-off identified.

Varying the random check probability was also shown to result in a trade-off between security and efficiency. Certainly for small values of $p_{\text{randomcheck}}$ there was a clear trade-off between security and efficiency performance. Increasing the number of random checks is a security measure that very easily can be implemented on airports. Therefore it can be very interesting to investigate the influence that this measure has in terms of security and efficiency performance, and how it trades off these two performance. But when looking into the results for the variation of $p_{\text{randomcheck}}$, one sees that at $p_{\text{randomcheck}} = 0$, the vulnerability is 0.5%, while at the highest value ($p_{\text{randomcheck}} = 0.9$), the vulnerability equals 11.8%. This is a factor difference of less than 25, between the lowest and highest value of the $p_{\text{randomcheck}}$ domain. This is a low number when compared to the experiments with variable $thres_{\text{threat}}$, where this factor is equal to 182. This indicates that varying $p_{\text{randomcheck}}$ is mainly of influence on the efficiency performance, and less on security performance. This therefore shows that $p_{\text{randomcheck}}$ might be a less interesting variable to investigate in the final model.

This will however further be clarified by the GSA, further in this section.

There is an important note that should be placed to the current investigation with variable $thres_{\text{threat}}$. Figure 4.4 basically is an ROC-curve, turned 90 degrees anti-clockwise and translated to the security and efficiency performance. It does not yet give much more information than only how TPR and FPR translate to security and efficiency performance. When making the model more detailed however, it is expected that this relation will be less trivial. This will be the case when autonomous and diverse agents are implemented, if the environment is made dynamic, and if there are more non-linear interactions between the agents and between the agents and the environment. How this can be improved in the final model will be explained in the proposal for the next modelling iteration, further in this section.

4.5.2. Global sensitivity analysis of results

In the three experiments, the three investigated parameters are all varied one at the time while all other parameters are kept constant. This is called local sensitivity analysis. Performing these experiments can say something about the role of the investigated parameter. But the result can be very dependent on the values of the other parameters in the model during the experiment. For example, varying d' can have a different effect when the threshold level is held constant at $thres_{\text{threat}} = 0.842$ instead of what was currently done ($thres_{\text{threat}} = 1.966$), then the behaviour of the system might be very different. To investigate the global role and importance of parameters in the system in every possible circumstance, Global Sensitivity Analysis (GSA) can be applied

The model for security time computation (Equation 4.3) is subject to a lot of parameters which all add uncertainty to the output of the model. With GSA one can investigate the uncertainty that parameters propagate to the output. In GSA one analyses which parameters propagate a relatively large portion of the total uncertainty to the output. Those parameters that propagate the largest portion of uncertainty are interesting for further investigation because of two reasons. (i) The propagated uncertainty can be reduced in order to make the eventual output less uncertain and thus more valuable. But more importantly (ii): the larger the uncertainty propagation of a parameter, the higher the importance and influence of that parameter on the output of the model. In this way, GSA can indicate parameters that are interesting to explore. For more information on GSA, one is referred to Saltelli et al.[20]. For understanding the results of GSA below, it is only important to know that the relative magnitude of S_{Tx} is a measure for the proportion of uncertainty that has been propagated to the output by a particular input parameter. The measure takes into account possible dependencies between parameters. The absolute value of S_{Tx} is not of importance; the insight is in the relative value of S_{Tx} of a certain variable compared to that of another variable. By ranking the outcomes of the GSA from high to low values for S_{Tx}, it can be stated that this is a ranking of global importance of the parameter.

In the current preliminary model, the security performance (measured in vulnerability) is dependent on three factors: $thres_{\text{threat}}$, d' and $p_{\text{randomcheck}}$. When the vulnerability model is implemented in SimLab2.2 and GSA is performed the ranking is as in Table 4.3.

Table 4.3: Result of GSA for vulnerability influencing parameters in expanded AATOM model

Parameter	S_{Tx}	Distribution
$thres_{\text{threat}}$	0.610	$U(0.842, 3.09)$
d'	0.321	$U(0.5, 3)$
$p_{\text{randomcheck}}$	0.198	$U(0, 1)$

In Table 4.3 it can be viewed that $thres_{\text{threat}}$ by far is the most influencing parameter in the model. The reason for this is clear: varying $thres_{\text{threat}}$ is of large influence to the overall vulnerability. It determines on the one hand if a passenger or his luggage is proceeded for an extra check, but then also determines if the passenger or luggage is identified as forbidden luggage in this extra check. The second most important parameter is d'. The reason for this result is clearly that it makes a large difference if an operator is of high quality (hence can identify forbidden items, and thus reduce vulnerability very well), or if he is of low quality. $p_{\text{randomcheck}}$ is of small importance: of course it contributes in reducing vulnerability when increased, but the influence is much lower than that of $thres_{\text{threat}}$ and d'.

When GSA is applied to efficiency performance in the model, there are more parameters of influence. The parameters of influence are all time distributions, the probability for the number of chosen boxes for the X-ray scan, and again the three parameters that are set to be variable: $thres_{\text{threat}}$, d' and $p_{\text{randomcheck}}$. The model was again implemented in SimLab2.2 and GSA was performed. The result of the performed GSA of the processing time in the security check of the model can be found in Table 4.4. For reference, the distributions of the parameters used in the model are also given in this table.

Table 4.4: Result of GSA for efficiency influencing parameters in expanded AATOM model

Parameter	S_{Tx}	Distribution
$t_{lugcheck}$	0,373	$N(104.67, 80.86)$
$t_{lugcollect}$	0,312	$N(71.5, 54.95)$
$p_{\text{randomcheck}}$	0,193	$U(0,1)$
$t_{lugdrop}$	0,141	$N(54.6, 36.09)$
d'	0,034	$U(0.5,3)$
$t_{phycheck}$	0,030	$N(43.00, 20.96)$
$thres_{\text{threat}}$	0,014	$U(0.842, 3.09)$
p^3_{box}	0,007	$p = 0.188$
p^2_{box}	0,007	$p = 0.485$
p^4_{box}	0,007	$p = 0.01$
$t^3_{\text{x-ray}}$	0,007	$N(20.82, 11.04)$
p^1_{box}	0,007	$p = 0.317$
$t^2_{\text{x-ray}}$	0,006	$N(16.44, 9.08)$
$t^1_{\text{x-ray}}$	0,006	$N(10.28, 5.06)$
$t^4_{\text{x-ray}}$	0,006	$N(21.00, 11.98)$

In Table 4.4 surprisingly the most influencing factors turn out to be not one of the variable parameters, but the parameters related to distribution of time. The parameters $t_{lugcheck}$ and $t_{lugcollect}$ receive the highest value for S_{Tx}. This is probably due to the large uncertainty in the normal distributions: the σ is rather high, which introduces a lot of uncertainty to the output: the expected processing time of security. The fact that these parameters are identified as very influencing, means that in further research it might be a good idea to focus on calibrating these parameters better.

Concerning the variable parameters, $p_{\text{randomcheck}}$ clearly is of largest influence. This is according to expectation: increasing $p_{\text{randomcheck}}$ proportionally increases the number of checks, which introduces much extra security time. The parameter d' is also of significant influence. This is clearly for the reason that operator performance largely influences the FPR. Surprisingly $thres_{\text{threat}}$ is of relatively smallest influence on the efficiency performance. An explanation for be can be the fact that for high values of $thres_{\text{threat}}$, varying the value of $thres_{\text{threat}}$ does not make a large difference. For example, at $d' = 2.778$, changing $thres_{\text{threat}}$ from 1.966 to 3.09 (hence shifting from 50% focus on security to 0 % focus on security) only decreases the FPR with 2 percent point. These low levels of focus on security will however hardly happen in reality. Figure 4.3a shows that for focus on security < 20 %, the vulnerability rises > 40%, which will be considered undesirable for most airports.

Finally, it can be deduced from Table 4.4 that the time distributions of X-ray processing and the probabilities for different boxes are of low influence on the total output. This is not very surprising, as this parameter is split up into many parts, and for that reason the distributions are simply less often used than for example $t_{lugcollect}$. This result can also be seen as a proof that the time distribution for X-ray is calibrated a relatively detailed level, as it does not contribute much uncertainty.

From the GSA can be concluded that of the variable parameters, $thres_{\text{threat}}$ is most influencing on security and less on efficiency, while the inverse is true for $p_{\text{randomcheck}}$. It can be defended that if the range of $thres_{\text{threat}}$ is set to more realistic values, the relative importance becomes larger. It is therefore concluded that $thres_{\text{threat}}$ is an interesting parameter to further investigate.

The other conclusion from GSA is that the distributions of $t_{lugcheck}$ and $t_{lugcollect}$ should be further investigated, while the distributions of $t_{\text{x-ray}}$ is probably calibrated well. This will be included in the recommendations in Chapter 7.

4.5.3. Proposal for next modelling iteration

In the current simulations, the security operators all function in the same way. First they compare threat levels with their assigned $thres_{\text{threat}}$, of which the value is equal for all operators. Consequently, based on the outcome they give a waiting instruction or not. As was discussed in Chapter 3, the strength of the used agent-based modelling paradigm is the opportunity to implement autonomous and diverse agents, and let them interact in a dynamic environment. Of these possibilities, there is only made use of the possibility to interact. It may be interesting for the next modelling iteration to focus on other opportunities that agent-based modelling offers. This will contribute in making the model more realistic.

Both the conclusion of the experiments and the GSA indicated that $thres_{\text{threat}}$ is a parameter that is of interest to investigate further. Investigating further means modelling in a different, possibly more detailed way. Together with d', $thres_{\text{threat}}$ is mainly determining the decision making process of the operators in the preliminary model. Therefore it is proposed for the final model, to investigate the possibilities to model the decision maker differently and on a lower abstract level.

It is of interest to investigate how to model the operators as autonomous agents who make own decisions. Next to that, diverse operators can be modelled: e.g. slow, good, tired, or differently focused operators. But not only the operators can be modelled into more detail, also the targeted people that the operators make decisions about can be investigated: the passengers. Further study and review is needed to find the way to model diverse, different types of passengers. And this all should be placed in a dynamic environment: an environment in which the circumstances change over time.

The above proposed improvements are all improvements to investigate the model to a *deeper* level, of lower abstraction. One could also choose to *widen* the investigation to the security efficiency trade-off. This can for example be done by looking into other threat scenarios, or into consequence assessments. It was however decided that it is a better idea to dive deeper into one aspect and come up with a useful detailed result, rather than investigating different things on a high abstract level. Therefore, the above described ideas will be worked out in the next section, in which the modelling of the final design is treated.

5

Model design

Using the results from the generalization, the last step of the first iteration of modelling in Chapter 4, the next iteration can be started. This iteration is performed in the same way: first a conceptual model will be designed (Section 5.1), followed by the specification of a formal model (Section 5.2) and the implementation of the model in the AATOM simulator (Section 5.3). The simulation and analysis using the final model will be performed using case studies, later in Chapter 6.

5.1. Conceptual model design

As was discussed in Chapter 2 and 2, the strength of the used agent-based modelling paradigm is the opportunity to implement autonomous and diverse agents, and let them interact in a dynamic environment. Therefore, as was described in the proposal for the next modelling iteration in Section 4.5, the aim is to make use of these possibilities and adapt the preliminary model. To do so, use will be made of the theories, models and data that were treated in the last section of Chapter 2.

First the diffusion model will be implemented as a cognitive model for decision making of the security operators. This introduces autonomy to the agents and allows for creating diverse agents. The diffusion model allows for incorporating a biomathematical fatigue model, as described in Chapter 2. This makes the behaviour of the operator agents different over time and thus incorporates the influence of the dynamic nature of the airport terminal environment. The final addition that will be introduced in this section is the diversity of the passengers. As Kirschenbaum described, passengers should not be seen as passive elements, and therefore an example of how diverse passengers can be modelled is given in this section. To keep this section structured and clear, the calibration of the different implemented models is treated directly after the introduction of the respective model. An overview of all calibrated parameters that are introduced in the model is given in Appendix D.

For the preliminary model, a graphical representation of the conceptual model was made. Because, with respect to interactions between agents in the airport environment, not many changes have been applied in the final model, it was decided that it would not contribute to design a new graphical representation for the final model.

In the final conceptual model, most elements of the preliminary model are still in place. Some elements however are removed or adapted. The major changes with respect to the preliminary model are:

- The removal of the random check probability $p_{randomcheck}$ of luggage and passengers;
- The observation of threat levels of luggage and passengers (l_t and p_t by security operators, and thereby also the comparison of threat levels with the threshold for threat (thres_{threat}). This is replaced by the diffusion model;
- The removal of the measurement of performance of operators in terms of d';
- The decoupling of forbidden body and luggage. Agents can now have a forbidden luggage and an allowed body and vice-versa.

Note that the WTMD still has a thres_{threat} and still observes p_t, as this is not a security operator, but security *equipment*.

5.1.1. Diffusion model

In Chapter 2 the Ratcliff Diffusion Model (RDM) was introduced. In this model the decisions of operators and the time they take for a decision are dependent on the drift rate v, a threshold a and the bias z of the operator. The values of these variables can vary per decision, and can be influenced by external factors. Drift rate v for example is usually dependent on the stimulus that the decision maker receives as an input. Before being able to use RDM it is important to calibrate it very well. For this purpose a diffusion process calibration algorithm called DMAT was introduced[62]. In this section it is treated how the diffusion model will be used for more detailed modelling of the security system and how the free parameters of the diffusion model are calibrated to data about security operator's behavior.

The diffusion model will be used for the decision process that is carried out by the security operators. It can be used for all three types of security operators (X-ray operator, luggage check operator and physical check operator), since all need to make a decision on whether an object or passenger is considered to be forbidden or not. The X-ray operator decides if a piece of luggage should proceed for an extra check, and the luggage and physical check operator determine if an object is found and considered to be forbidden. If the result of their diffusion decision process is positive (crosses threshold a), then they will decide that the item should proceed for extra check (X-ray operator) or should be confiscated (luggage & physical check). If the result of their diffusion decision process is negative (crosses 0), then they will decide that the item can pass through the security check.

The drift rate v, or the speed at which the operators accumulate their evidence and in what direction, is determined by the stimulus they receive from the luggage or from the passenger. In the model, two types of stimuli are considered. Stimulus A is the stimulus of an allowed item / passenger, and stimulus F is the stimulus of a forbidden item / passenger. Clearly, stimulus F from a forbidden item will result in a more positive drift rate v than stimulus A. But this does not necessarily mean that stimulus A will always result in a positive result of the diffusion process. Depending on the magnitude of drift rate v, the threshold a and the bias z, the chances on a positive decision are largely influenced. How exactly these parameters play a role in the probability of a positive result of the decision process will be explained below.

For the explanation of the role of the bias z, consider all other parameters (including intertrial and non-decision parameters) to be fixed. A positive bias, or high number for z will thus always result in more false positives and a higher hit rate, if all other parameters stay the same. This characteristic of modifying bias can be compared to the threshold that was used in the preliminary model. But modifying z implies more: a high bias (large z) will result in shorter response times for positive decisions and longer response times for negative decisions. Modifying the bias z in the diffusion decision process of the security operators will be the main input for case study 1 in Chapter 6: *the acute-chronic goal responsibility trade-off*.

Except for varying the bias z, one could also take a look at the threshold a. An important reason why the diffusion model is chosen to represent the autonomous decisions of the operator agents, is because of its ability to model the *speed-accuracy trade-off* that was described in Chapter 2. Moving the threshold a is considered to be trading speed for accuracy. This works as follows. Set as assumption all parameters constant and assume z dependent on a (e.g. $z = a/2$, no bias). If threshold a is increased and drift rate v remains constant, more evidence is required for a positive result. But also more evidence is required for a negative result. The decision is thus based on more evidence. In other words, the decision is more *accurate*. This comes however at a cost: it takes longer to reach either of the boundaries, so the reaction time will be longer, or the *speed* will be lower. The inverse is true when threshold a is decreased: the decisions will become less accurate, but at a higher reaction speed. In Figure 5.1 a graphical representation of the consequences of a smaller a is given.

In Figure 5.1 a high value for a is depicted by the black small horizontal lines as boundaries, and a low value for a is depicted by the thick red line. It can be viewed that for Decision 1 the difference is only in terms of speed. With the lower value of a, the same decision is made but the response time (RT) is smaller. But for Decision 2 something else is true. If the line of Decision 2 is followed from the starting point, one sees that the line crosses the bottom boundary of the small red boundaries earlier than it crosses the upper boundary. While at a higher accuracy (larger a), the upper boundary would have been crossed after having accumulated more evidence. And since the latter is chosen after more evidence, this was probably the right decision. The above described method is how the speed-accuracy trade-off is implemented in the final model.

There is also the possibility to vary drift rate v based on more inputs than only a forbidden stimulus (F) and an allowed stimulus (A). For this research however, it is chosen to simplify this as varying a and z and analyzing the results is already quite extensive. It is something however that will be treated in the discussions, implications and recommendations chapter, Chapter 7.

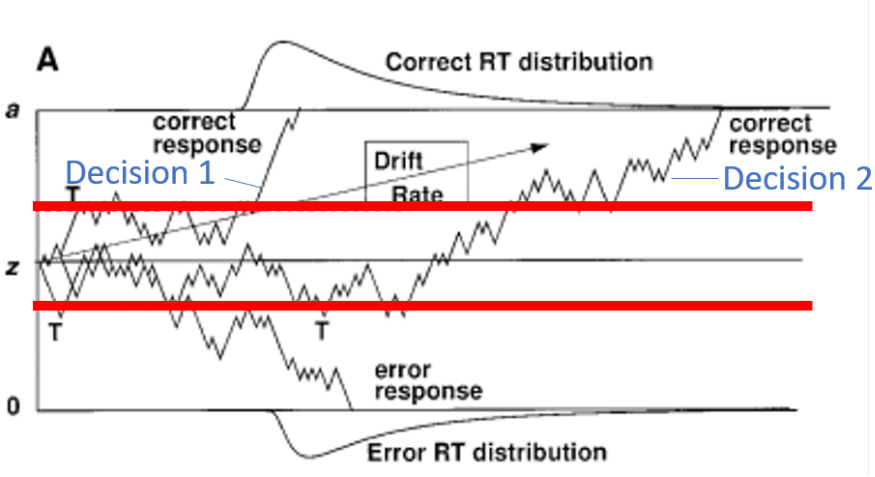


Figure 5.1: Graphical representation of speed-accuracy trade-off in diffusion process

Having explained the possibilities with the diffusion model, it is important to calibrate it correctly. The calibration is of large influence on the model, since the operators' decisions will largely influence the security and efficiency performance of the system. For calibration of the model, the calibration algorithm DMAT, created Vandekerckhove and Tuerlinckx is used[62]. This algorithm requires as input data about the response of an operator and the time in which he made the response.

The latter, data about response times, is retrieved from RTHA: the data on which the preliminary model was calibrated could also be used for this purpose. Data about the response of the operator however was more difficult to retrieve, as RTHA did not provide data on FPR and TPR. Therefore, as was done for calibrating the preliminary model, data from Table 4.1 was used. Also in this case, the model was calibrated on two set-ups: (i) security focused airport (TPR = 97%, FPR = 20%) and (ii) efficiency focused airport (TPR = 45%, FPR = 0.1%). The response times were taken from the physical check, as the data on physical check response times from RTHA had the highest number of observations.

Next an important assumption was made. Clearly, the response time of the different types of operator (X-ray operator, luggage check operator and physical check operator) is distributed differently. However, since every operator needs to make a comparable choice (is the luggage / passenger forbidden or not?) it is assumed that for every operator the decision time is equally distributed. The difference in response time between the different operators is thus made in a different *non-decision time*, T_{er}, and its intertrial variable range s_t. As was done in the preliminary research, different non-decision times T_{er} and s_t were specified for different number of boxes over which the luggage is spread. This was necessary to make the fits of the diffusion response times and the obtained data significant. This will be further explained in Section 5.3, where verification and validation are treated.

The result of the calibration using the DMAT algorithm can be found in Table 5.1.

As can be viewed in Table 5.1, $0.0910 \leq z \leq 0.8143$, where $z = 0.0910$ models an efficiency focused bias and $z = 0.8143$ models a security focused bias. The standard value for z lies exactly between these two values. The lower boundary b has a standard value of 0, as is usual in diffusion models. The variable range is to account for the change in threshold a: modifying a should not have an effect on the bias, and hence b needs to adapt accordingly. Note that the maximum value of b will be used when the minimal value of a is used and vice-versa. Another notion is that the absolute value of v_a is larger than the absolute value of v_f. This means that operators are quicker in collecting evidence that an item is allowed, than in collecting evidence that an item is forbidden. A final notion is put to the fact that since the two stimuli A and F both trigger a different drift rate, they also have a different value for intertrial variation of drift rate, namely η_a and η_f. It can be viewed in the in Table 5.1 that $\eta_a > \eta_f$. This means that the variation in v_a is larger than the variation in v_f. This implies that the following. For some allowed items, evidence that the item is allowed is obtained very quickly, while for others it is not. Forbidden items have more constant rates for collecting evidence.

5.1.2. Fatigue model

As has been described in the literature review in Chapter 2, the number of decisions that already have been made on a day influences the quality of the later made decisions. If there is any job in which many decisions

Table 5.1: Calibration of parameters in diffusion model

Parameter	Symbol	Standard value	Variable range
Upper boundary	a	0.9585	(0.5325, 4.7925)
Lower boundary	b	0	(-3.834, 0.426)
Starting point (bias)	z	0.4522	(0.0910, 0.8134)
Drift rate (allowed)	v_a	-0.0859	-
Drift rate (forbidden)	v_f	0.0392	-
Non-decision time (physical check)	$T_{er,phy}$	32.713	-
Non-decision time (luggage check)	$T_{er,lug}$	61.819	-
Non-decision time (X-ray check, 1 box)	$T_{er,x-ray^1}$	3.670	-
Non-decision time (X-ray check, 2 box)	$T_{er,x-ray^2}$	7.861	-
Non-decision time (X-ray check, 3 box)	$T_{er,x-ray^3}$	12.242	-
Non-decision time (X-ray check, 4 box)	$T_{er,x-ray^4}$	15.359	-
Bias variability	s_z	0.158	-
Non-dec. time var. (physical check)	$s_{t,phy}$	25.000	-
Non-dec. time var. (luggage check)	$s_{t,lug}$	47.243	-
Non-dec. time var. (X-ray check, 1 box)	$s_{t,x-ray^1}$	1.299	-
Non-dec. time var. (X-ray check, 2 box)	$s_{t,x-ray^2}$	12.007	-
Non-dec. time var. (X-ray check, 3 box)	$s_{t,x-ray^3}$	15.356	-
Non-dec. time var. (X-ray check, 4 box)	$s_{t,x-ray^4}$	16.738	-
Drift rate variability (allowed)	η_a	0.04988	-
Drift rate variability (forbidden)	η_f	0.02364	-

need to be made, it is the job of a security operator. It is therefore not realistic to assume that an operator is able to make decisions of the same quality at the beginning as at the end of his working day. Agent-based modelling provides the opportunity to account for this dynamic behaviour of security operators.

The diffusion model that was previously introduced, can be combined with McCauley's biomathematical fatigue model, as was described in Chapter 2. Walsh et al. investigated this combination and found that the resulting model fitted experimental choice data very well[39]. Therefore, a combination of the Ratcliff Diffusion Model and McCauley's biomathematical fatigue model will be used to model the state of fatigue of an operator.

In the research of Walsh et al., it is assumed that the fatigue score resulting from McCauley's fatigue model has an influence on the drift rate of the decision maker. The influence is a linear relation, represented in equation 5.1.

$$v_{dynamic} = a_v * F(t) + b_v \tag{5.1}$$

In equation 5.1, $v_{dynamic}$ is the dynamic drift rate, resulting from the fatigue relation in which the fatigue number F (a real, positive number) is variable over time. a and b are constants: a is the rate at which the drift rate is adjusted per fatigue point, and b is the diffusion constant. If b is positive, a is negative, and vice-versa. This means that for any person with either a positive or negative drift rate constant b, a higher fatigue number F results in a $v_{dynamic}$ that is closer to zero. In other words, for all decision makers, tiredness results in a smaller absolute drift rate. The result is that tiredness has two consequences. The first is that the stimulus is received worse and the possibility of a good choice being made is reduced. The second consequence is that fewer evidence is collected per unit of time, and thus that the decision maker will have a larger response time.

No literature was found about the influence of a work shift on the fatigue (score) of people. But a model concerning the influence of the time on the day on the fatigue score of an average person that needs to perform a vigilance task has been found in literature. This relation is plotted in Figure 5.2.

The fatigue score F in Figure 5.2 is based on parameters with predefined standard values that are used in the fatigue score differential equations computation by McCauley et al.[17]. The result is thus a graph of the fatigue score of an average person during the day, who takes an average number of decisions. It is seen that the fatigue score increases gradually during the day. This increase in fatigue score declines around 17:00, probably at the end of an average working day and/or with a recovering meal. The fatigue score increases most just before bed time.

In the flight schedule that was implemented in the AATOM simulator (see Appendix B), it can be viewed that there are three departure flight peaks during the day at RTHA: around 07:00, around 14:00 and around

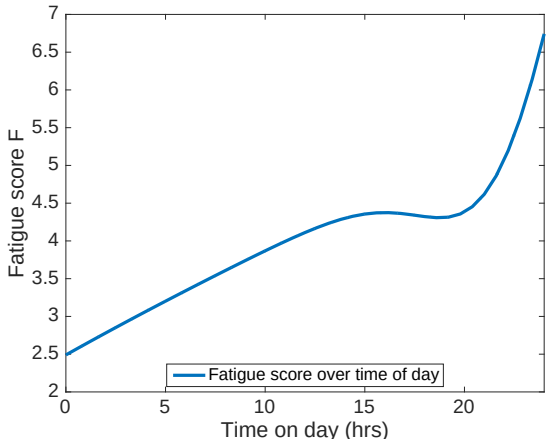


Figure 5.2: Fatigue score over day for average person

19:00. The passengers boarding these flights are handled averagely one hour before departure. If the fatigue score of an average person during the time of the early passenger peak at 06:00 (3.336) is compared to the fatigue score during the latest peak at 18:00 (4.323), there is a significant difference. It must be noted that these fatigue scores may not necessarily match perfectly with the fatigue scores of real operators at these peak times on the day, because of various reasons [1]. But apart from this, it can be assumed that an operator at 07:00 suffers less from decision fatigue than an operator working at the 19:00 peak. The goal of this research is to identify relations and trade-offs between security and efficiency in different circumstances. Combining the last two arguments, it can be concluded that the fatigue over time per day relation can be used to model the relative effect of fatigue on operator performance. How exactly this will be done will be explained in Chapter 6.

The calibration of the fatigue model has been performed based on the findings of Walsh et al.[39]. In this research, the values of a and b of the dynamic fatigue drift rate equation (Equation 5.1) have been calibrated on experimental data of people performing a PVT. The influence of fatigue on the PVT can be compared to influence that fatigue has on a two-choice decision, as explained in Chapter 2. Therefore, the relative influence of one fatigue point on the drift rate in the cited research is adopted for this thesis research. In mathematical terms, the same ratio a_v/b_v has been used. This ratio was found to be $a_v/b_v = 0.0172$.

For calibration of the model it is assumed that the calibrated values of v_a and v_f in Table 5.1 are valid for the middle of the day: 12:00. Knowing a_v/b_v and using the just stated assumptions, the values for a_v and b_v can be calibrated for the model used in this thesis research. The result is shown in Table 5.2.

Table 5.2: Calibration of parameters in diffusion model

Parameter	Standard value	Variable range
a_{v_a}	1.591E-03	(1.424E-03, 1.766E-03)
a_{v_f}	-7.259E-03	(-8.057E-04, -6.496E-04)
b_{v_a}	-9.244E-02	-
b_{v_f}	4.218E-02	-

In Table 5.2 distinction is made between a_{v_a} and a_{v_f}, and b_{v_a} and b_{v_f}, because there are two different drift rates for different stimuli. Note that a_{v_a} is positive and a_{v_f} is negative, resulting in smaller absolute drift rates for higher fatigue scores. Furthermore, also based on the calibration of Walsh et al., a range has been specified for the values of a_{v_a} and a_{v_f} in the following way[39]. In their calibrated values of a_v and b_v, Walsh et al. give standard deviations of the values. A standard deviation of 1 resulted in a variable range of $0.0154 \leq a_v/b_v \leq 0.0191$. This range has been used to calibrate the range of the parameters in Table 5.2. The specified range can be used to model diverse operators: operators who suffer more or less from diffusion fatigue during the day.

Using the described fatigue model, one can research how the system performance changes during the day, which incorporates the dynamic aspect of the system.

[1] Example reasons: (i) operators working at 07:00 peak started working relatively early on the day so they may have a higher fatigue score, (ii) operators working at 19:00 have started with working later so they may have a lower than average fatigue score

5.1.3. Diverse passengers model

In the literature review in Chapter 2 a citation of Kirschenbaum was given where he states that passengers should not be modelled as passive elements that are handled through a security check. There are different types of passengers with different behaviour, and they should be modelled accordingly. This is what will be done in the diverse passenger model.

In the same paper Kirschenbaum provides data on how often passengers are checked, conditional on their type of flight. It is stated that for charter flights an average of 1 in 2-3 passengers require an extra check, while this value is equal to 1 in 7-9 passengers for business flights. Therefore different types of flight days are modelled, in which only flights of one type will be simulated. One is a charter flight day, in which the probability is 41.67% ($= \frac{\frac{1}{2}+\frac{1}{3}}{2}$) that a passenger or a luggage piece carries a forbidden item. For a business flight day this number is equal to 12.70% ($= \frac{\frac{1}{7}+\frac{1}{9}}{2}$). For an average flight day containing both types of flight, a proportion of these two fractions was taken based on data from Kirschenbaum. This results in a probability of 20.67% on a forbidden item for a luggage piece of a passenger.

Other than was done in the preliminary model, these probabilities are drawn separately for the luggage and for the body of passengers. In this way, most passengers only have allowed items, some passengers have only forbidden luggage, the same number of (but different) passengers have forbidden objects on their body, and only few passengers have both.

But another distinguishing factor between passengers is the type of carry-on luggage they bring along. As this is of large influence to the processing time of the X-ray scan, different passengers get assigned a different number of required boxes for the X-ray scan. Based on data of RTHA on the number of required boxes per passenger check, a distribution has been found for the number of required boxes for passengers at the X-ray scan. This distribution, together with the above described forbidden item distribution, is presented in Table 5.3.

Table 5.3: Calibration of parameters in diffusion model

Parameter	Standard value	Variable range
$p^{average}_{\text{forbidden}}$	0.2067	-
$p^{business}_{\text{forbidden}}$	0.1270	-
$p^{charter}_{\text{forbidden}}$	0.4167	-
p^{1}_{box}	0.317	-
p^{2}_{box}	0.485	-
p^{3}_{box}	0.188	-
p^{4}_{box}	0.010	-

Using this model, it is possible to simulate the performance of airports for different type of flight days. The distribution of the number of required boxes per passenger can also be used to model different type of flight days, other than charter or business days. For example, the model could be further calibrated on the difference in box distribution between summer and winter: in the winter people wear big jackets which probably requires more boxes.

5.2. Formal model

The formal model of the final model is comparable to the preliminary formal model as described in Section 4.2. It is written in the same language, LEADSTO, uses some of the same sorts, predicates and constants, and the relations are comparable. But there are some differences. Following from the conceptual model specified in 5.1, the introduced differences are:

- No random checks are included in the model, since $p_{\text{randomcheck}} = 0$ in the model;
- Operators do not observe a threat level from passengers and luggage, but observe a stimulus if the passenger / luggage has a forbidden item or not;
- Operators do not compare with their internal state $thres_{\text{threat}}$, but make a decision as a consequence of the diffusion process;
- Inputs for the diffusion process, diffusion constants, are added.

Note that the WTMD sensor still makes use of the threat level to base it's decisions on.

The sorts are changed marginally when compared to the explained sorts of the preliminary formal model in Section 4.2. The additional sorts, added to the sorts in the preliminary model, are:

Sort	Explanation	Elements
AGENTPROPERTY	a set of all properties of agents	{p$_f$, diff}
OBJECTPROPERTY	a set of all properties of objects	{l$_f$}
DIFF	a set of all diffusion constants	{diff$_a$, diff$_f$}
TIME	a set of linearly ordered time points	{tdiff$_{xray}$, tdiff$_{lug}$, tdiff$_{WTMD}$, tdiff$_{phy}$

The additionally added predicates are the following:

Predicate	Description
initiate_diffusion_process(operator, constants)	The agent *operator* initiates the diffusion process with stated *constants* as input
output_diff_process(operator)	The output of the diffusion process of the agent *operator*

Compared to the preliminary model, some constants need to be added. These are the following:

Constant	Description	Range
diff$_a$	Diffusion constants for allowed item	*see Table 5.1*
diff$_f$	Diffusion constants for forbidden item	*see Table 5.1*
l_f	If luggage contains forbidden item or not	boolean
p_f	If passenger contains forbidden item or not	boolean
tdiff$_{lug}$	Luggage check duration	*Consequence of diffusion process*
tdiff$_{phy}$	Physical check duration	*Consequence of diffusion process*
tdiff$_{WTMD}$	WTMD check duration	*Consequence of diffusion process*
tdiff$_{xray}$	X-ray check duration	*Consequence of diffusion process*

M1: X-ray sensor, X-ray operator and luggage check operator performing luggage check

M 1.1: Observing forbidden property of luggage in X-ray
observed_own_location(op$_{xray}$, $l_{op_{xray}}$) & observed_location_of(op$_{xray}$, l_{xray}, lug) →
$_{[0,0,1,1]}$ observed_property_of(op$_{xray}$, l_f, lug)

M 1.2: Initiating X-ray operator diffusion process for forbidden item
observed_property_of(op$_{xray}$, l_f, lug) & l_f = true → $_{[0,0,1,tdiff_{xray}]}$ initiate_diffusion_process(op$_{xray}$, diff$_f$)

M 1.3: Initiating X-ray operator diffusion process for allowed item
observed_property_of(op$_{xray}$, l_f, lug) & l_f = false → $_{[0,0,1,tdiff_{xray}]}$ initiate_diffusion_process(op$_{xray}$, diff$_a$)

M 1.4: Output of X-ray operator diffusion process
initiate_diffusion_process(op$_{xray}$, diff) → $_{[0,0,tdiff_{xray},1]}$ output_diff_process(op$_{xray}$)

M 1.5: Information transfer between X-ray operator and luggage check operator
output_diff_process(op$_{xray}$) = true → $_{[0,0,1,1]}$ communicated_from_to(op$_{xray}$, op$_{lug}$, perform_check(lug))

M 1.6: Observing forbidden property of luggage in luggage check
communicated_from_to(op$_{xray}$, op$_{lug}$, perform_check(lug)) & observed_own_location(op$_{lug}$, $l_{op_{lug}}$) &
observed_location_of(op$_{lug}$, $l_{op_{lug}}$, lug) → $_{[0,0,1,1]}$ observed_property_of(op$_{lug}$, l_f, lug)

M 1.7: Initiating luggage check operator diffusion process for forbidden item
observed_property_of(op$_{lug}$, l_f, lug) & l_f = true → $_{[0,0,1,tdiff_{lug}]}$ initiate_diffusion_process(op$_{lug}$, diff$_f$)

M 1.8: Initiating luggage check operator diffusion process for allowed item
observed_property_of(op$_{lug}$, l_f, lug) & l_f = false → $_{[0,0,1,tdiff_{lug}]}$ initiate_diffusion_process(op$_{lug}$, diff$_a$)

M 1.9: Output of luggage check operator diffusion process
initiate_diffusion_process(op$_{lug}$, diff) $\rightarrow_{[0,0,tdiff_{lug},1]}$ output_diff_process(op$_{xray}$)

M 1.10: Arrest luggage with positive result
output_diff_process(op$_{xray}$) = true $\rightarrow_{[0,0,1,1]}$ arrested(op$_{lug}$, lug)

M2: WTMD sensor and physical check operator performing physical check
M 2.1: Observing threat level of passenger in WTMD
observed_own_location(WTMD, l_{WTMD}) & observed_location_of(WTMD, l_{WTMD}, pax)
$\rightarrow_{[0,0,1,t_{WTMD}]}$ observed_property_of(WTMD, p_t, pax)

M 2.2: Information transfer between WTMD and physical check operator
observed_property_of(WTMD, p_t, pax) & property_of(thres$_{threat}$, WTMD) & $p_t \geq$ thres$_{threat}$ $\rightarrow$
$_{[0,0,t_{WTMD},1]}$ communicated_from_to(WTMD, op$_{phy}$, perform_check(pax))

M 2.3: Observing forbidden property of passenger in physical check
communicated_from_to(WTMD, op$_{phy}$, perform_check(pax)) & observed_own_location(op$_{phy}$, $l_{op_{phy}}$) &
observed_location_of(op$_{phy}$, $l_{op_{phy}}$, pax) $\rightarrow_{[0,0,1,1]}$ observed_property_of(op$_{phy}$, p_f, lug)

M 2.4: Initiating physical check operator diffusion process for forbidden item
observed_property_of(op$_{phy}$, p_f, pax) & p_f = true $\rightarrow_{[0,0,1,tdiff_{phy}]}$ initiate_diffusion_process(op$_{phy}$, diff$_f$)

M 2.5: Initiating physical check operator diffusion process for allowed item
observed_property_of(op$_{phy}$, p_f, lug) & p_f = false $\rightarrow_{[0,0,1,tdiff_{phy}]}$ initiate_diffusion_process(op$_{phy}$, diff$_a$)

M 2.6: Output of physical check operator diffusion process
initiate_diffusion_process(op$_{phy}$, diff) $\rightarrow_{[0,0,tdiff_{phy},1]}$ output_diff_process(op$_{phy}$)

M 2.7: Arrest luggage with positive result
output_diff_process(op$_{phy}$) = true $\rightarrow_{[0,0,1,1]}$ arrested(op$_{phy}$, pax)

5.3. Implementation, verification and validation of model

In this section the implementation, verification and validation of the final model are preformed.

5.3.1. Implementation of model

Having specified the final conceptual and formal model, the model is ready to be implemented in the AATOM simulator. The implementation consisted of the following parts:

- Rewriting a MATLAB algorithm of the drift diffusion process to JAVA;[7]
- Specifying a variable fatigue score over time, following the line of Figure 5.2;
- Coupling the fatigue score F to the drift rate v by using F as input;
- Linking forbidden and allowed items using different stimuli to respectively drift rate v_f and v_a;
- Split forbidden item probability: separate probability for forbidden luggage and body
- Implement variable forbidden item probability for different flights.

5.3.2. Verification of model

In order to be sure that the diffusion process algorithm was working properly, small experiments with different parameters have been performed. These different parameters were given as an input to the downloaded MATLAB algorithm and to the self-created JAVA algorithm and the results were compared. If the number of simulations was increased sufficiently, the results converged to the same values. Hereby the rewritten diffusion process code in JAVA was verified.

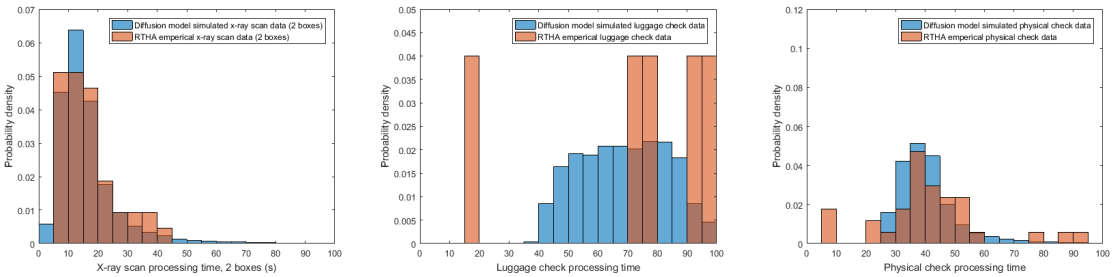
The fatigue model was then verified as follows: first different minutes on the day were set as input to the diffusion model in the JAVA algorithm. Then an algebraically determined drift rate after fatigue correction was given as input to the MATLAB algorithm. Consequently, the outputs for large n were compared again, and were found to be similar. Next outputs of diffusion decision process for different passengers (allowed and forbidden) were compared and a significant difference was found in the outputs. This meant that the distinction in drift rates caused by different stimuli from passenger types was working properly.

The variable forbidden item probability was verified by again giving the agents different colors. Passengers with allowed luggage and body were given the color *red*, passengers with allowed luggage and forbidden body *blue*, passengers with forbidden luggage and allowed body *green*, and passengers with both forbidden luggage and forbidden body were given the color *pink*. Using sample counting the implementation was verified.

A simulation of one flight morning takes approximately 7-8 minutes. Per core on a computer one simulation can be performed, so on a quad-core processor 4 simulations can be performed in parallel, resulting in an average waiting time of 2 minutes per simulation.

5.3.3. Validation of model

Having performed this, the new implementations in the model were verified and validation of the model was performed. A first step was the validation of the calibrated diffusion decision processes. To do so, outputs and response times of the diffusion processes of X-ray operators, luggage check operators and physical check operators were compared to data from RTHA, on which the model was calibrated. This is done using the MATLAB diffusion process algorithm. The calibrated data was implemented in the MATLAB algorithm, and 10,000 data points of decisions were generated. The output of this algorithm is the decision (positive or negative) and the response time. The decision output of the diffusion process is compared to the calibration data for the highest security focus (97% TPR, 20% FPR) and the lowest security focus (45 % TPR, 0.1 % FPR). For both the highest and the lowest security focus, the output of the decisions matched the literature data within a margin of +/- 0.5%. Next the response times were compared to data from RTHA about response times of different operators. The results of this validation step are graphically depicted in Figure 5.3.



n = 10,000
p = 0.8993

(a) X-ray check response time (2 boxes) validation

n = 10,000
p = 0.6604

(b) Luggage check response time validation

n = 10,000
p = 0.6213

(c) Physical check response time validation

Figure 5.3: Validation of diffusion decision response times with RTHA data

In Figure 5.3 the probability distribution of the diffusion decision response times are compared with the empirical data from RTHA about the response times. On the two distributions Kolmogorov-Smirnov tests were performed to test for the hypothesis if the distributions could identified to be different ones. For all distributions (the depicted in Figure 5.3, but also X-ray check with 1, 3 and 4 boxes), the p-value was significantly higher than 0.05. Also when one takes a look at the distributions in Figure 5.3, it is seen that the distributions are indeed similar. Only for the luggage check time, the distributions are difficult to match since there are

only five date points from RTHA to compare to. Given the similarity of the distributions and the responses with data and literature, it can be concluded that the diffusion decision model is validated.

After having validated the diffusion decision model, it is interesting how the diffusion model behaves between the maximum and minimum value of security focus. To get an insight in this, the diffusion model for varying bias z is plotted together with the ROC-curve from the preliminary model in Figure 5.4.

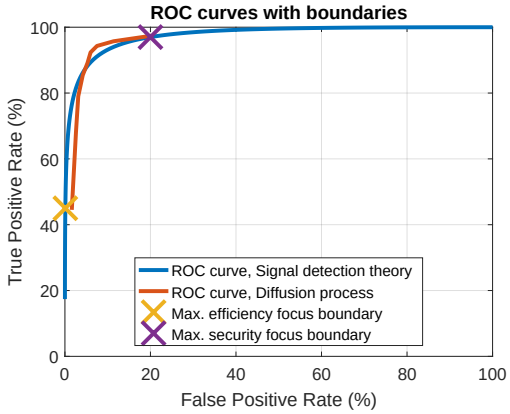


Figure 5.4: ROC-curves of signal detection theory and diffusion process

Both models are calibrated on the same maximum and minimum security focus boundary. The area in between is determined by the calibration of the signal detection model and the diffusion model. Although the two models are calibrated on the same boundaries, it can be viewed in Figure 5.4 that the diffusion process ROC curve behaves differently from the signal detection theory ROC curve. For high security focus, the diffusion decision model performs better: a high TPR is obtained at a lower FPR. Also the gradient of the diffusion decision process is steep for low FPR. This indicates a region in which TPR can be increased at low cost of FPR. For low TPR, the signal detection ROC curve performs better than the diffusion process curve.

Validation of the biomathematical model used within the current diffusion decision model can only be done by performing empirical research. It is however true that the used biomathematical model is a validated one, and the incorporation of the biomathematical model within the diffusion decision model is validated as well in a recent research[39]. Therefore it can be stated that the used biomathematical diffusion decision model is validated.

As was described in the validation of the preliminary research, the given inputs need to lead to an output that matches reality. Specifically, this means that the throughput rate in the security check during the simulations should be comparable to the realistic security throughput rate (2.6 pax per lane per minute) when implementing the calibrated diffusion decision model. First validating simulations however turned out that this was not the case: again a throughput rate between 1 and 2 passengers per lane per minute was the output of the security system. Therefore, again a correction factor CF was applied. To the normally distributed processing times such as luggage drop and luggage collect, CF was applied in the same way as in the preliminary model. For the diffusion decision process, the correction factor was applied by multiplying CF with the non-decision time T_{er}. By performing iterative simulations, it was found that $CF = 0.65$ led to a throughput rate of 2.6 pax per lane per minute, just like it did in the preliminary model. Therefore, CF was set to be 0.65 in the final implementation

The last form of validation is performed in the evaluation of the case studies in Chapter 6. In this evaluation different types of operators are placed in the trade-off between security and efficiency. By reviewing if the placement of these operators is according to expectation, an empirical form of validation is performed.

6

Experiments and Results

In this chapter the experiments will be performed using the model that was completed in Chapter 5. To test the newly added elements in the model, four case studies and one empirical evaluation will be performed. All studies are designed to help answering the research question that was specified in Chapter 3: "How could the dynamic relations between security and efficiency in airport terminal operations be identified and analyzed?"

To answer this question, firstly the model will be tested by applying two trade-off theories to the implemented diffusion model. In the first case study this is the acute-chronic goal responsibility trade-off, and in the second case study this is the speed-accuracy trade-off. Next, the influence of tiredness on operators is tested by placing them in the acute-chronic goal responsibility trade-off in case study 3. In case study 4 the influence of different types of passenger on the security and efficiency performance will be tested. Case study 5 is different from all case studies. In this case study an empirical research performed by another MSc student is used to determine the place of real operators in the trade-off. This last case study can also be used as validation of the research.

In case study 1-4 sensitivity analysis is performed on the model. This is done by performing simulations with different set-ups. Across different set-ups within a case study, one or more parameters are varied and the influence of these variable parameters on the system performance is analyzed. In all experiments 250 simulations are performed for all different set-ups within the experiment. This results in a total of 61 set-ups with each 250 simulations, representing a total of 15,250 simulated flight days.

Performance in terms of efficiency is measured in all case studies as mean processing time through the security check and mean queuing time before the security check. This means that the output of every simulation is a value for the mean processing and queuing time for the simulated day. This results in 250 x 2 efficiency performance outputs for every set-up.

The security performance measure is vulnerability. This vulnerability is the total vulnerability of the system (except in case study 2). This means that it is a combination of the vulnerability of the luggage check (performed by X-ray operators and luggage check operators) and the vulnerability of the passenger check (performed by the WTMD and the physical check operator). The luggage check vulnerability is a combination of two diffusion decision processes, the passenger check vulnerability a combination of a diffusion decision process and the signal detection of the WTMD. The difference in computed vulnerability for luggage or passenger bodies can be investigated.

In this chapter the set-up of the case studies will be introduced, results will be graphically shown and the characteristics of the results will be analyzed. This analysis will be performed for every case study separately: Section 6.1 - 6.4 will respectively treat case study 1 - 4. In Section 6.5 an empirical evaluation of the case studies is performed. A synthesis on the overall results of the different case studies and a discussion about their implications will be presented in Chapter 7.

For completeness a collection of all resulting figures from the experiments is presented in Appendix C.

6.1. Case study 1: Acute - chronic goal responsibility trade-off

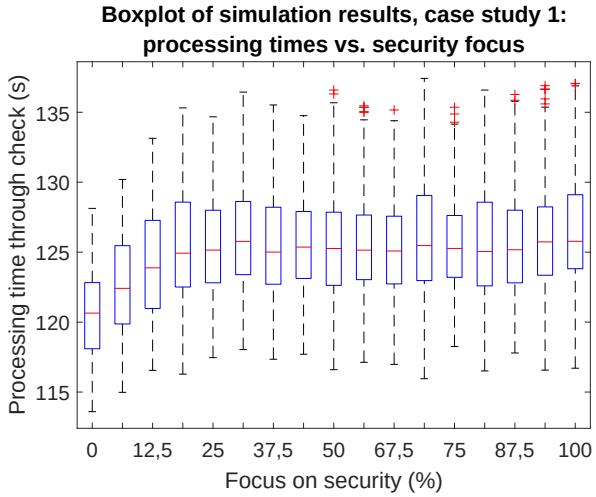
In the first case study the acute-chronic goal responsibility trade-off will be analyzed using the diffusion decision process of the operators and the signal detection of the WTMD. Both the diffusion process and the signal detection allow for a varying focus in security. For the WTMD (signal detection) this is the thres_{threat}, as was done in the preliminary model. For the operators (diffusion decision process), this is done by varying the bias z. The boundaries within thres_{threat} and z can be varied and are determined by the calibration of the model. The boundary that leads to the highest security performance (low thres_{threat}, high z) is regarded as a 100% focus on security. The boundary that leads to the highest efficiency performance (high thres_{threat}, low z) is regarded as 0% focus on security. Between these boundaries, 15 other set-ups are defined, which are steps of 6.25% in security focus. The values of thres_{threat} and bias z for different security focus steps are linear increments of equal step size between the two calibrated boundaries. The security focus of 50% corresponds to the "standard value", of $\text{thres}_{threat} = 1.966$ and $z = 0.4522$.

Table 6.1: Variable parameters for case study 1

Variable parameter	0 % security focus	100 % security focus	Step size (6.25 %)
thres_{threat}	3.090	0.842	-0.1405
z	0.091	0.8134	0.0452

Result of case study 1

The result of the experiment in terms of efficiency performance is presented in a boxplot in Figure 6.1.



$n = 250$, 17 set-ups

Figure 6.1: Boxplot of processing times distribution for different security focus

In the boxplot in Figure 6.1 the resulting mean processing times are plotted. As holds for every case study in this research, for every set-up $n = 250$. A box in the boxplot shows the range of the processing times of the 250 simulation results for that set-up. The small horizontal red lines represent the median of the processing times. The blue box shows the range within which 50% of the results fall. The red crosses represent outliers.

A first observation is a positive correlation between focus on security and processing time. For a low focus on security, the processing times is low and the highest blue box and median is found at a 100 % focus on security. But the increase is not linear. For low focus on security, the gradient is high. Then when the focus on security is increased to > 30%, the processing time becomes almost constant, or even slightly decreasing. The processing times follow this slight decrease up to a security focus of around 75 %, after which the processing time increases again when focus on security is further raised.

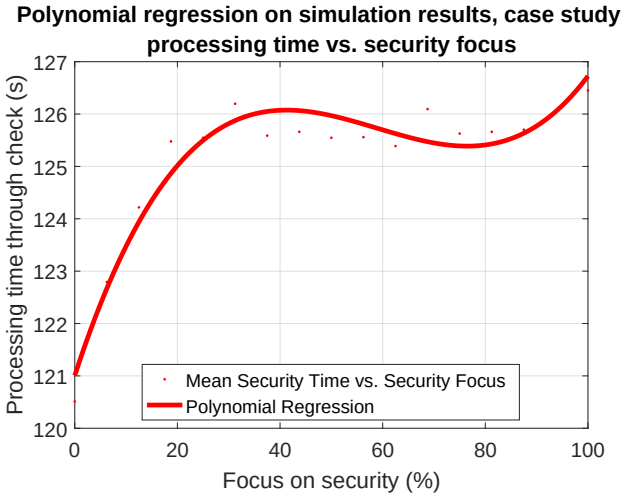
Another observation is that for all set-ups of different focus of security, the spread of the results is large. The minima and the maxima of all set-ups have a range of approximately 10 - 15 seconds. This is 8 - 12 % of

the median value of the processing times. This uncertainty is probably due to the large number of non-linear interactions that occur in the agent-based model. Because of the interactions of the passengers with each other and with the operators, some simulation days the mean processing time can emerge to be higher than others. Therefore it will be chosen to make use of the mean of the simulation results: all 250 resulting mean processing times will be averaged, and this mean will be taken as the result of the set-up to fit a regression on.

The outliers are all on the upper side of the boxplot. This means that the outliers are simulations with very high processing times. These outliers occur mostly above 50% of security focus. Apparently a high security focus implies situations that need extra attention, which result in a very high security processing time. This is an interesting topic for further research. As the mean of the average processing time is of interest, outliers that are outside of 2 standard deviations of the mean of the result are of a set-up are filtered.

The results in Figure 6.1 only show the behavior for 17 set-ups with a different focus on security. When one wants to know how the system exactly behaves for different focus on security, one can estimate the regions between these 17 set-ups. Because of the large uncertainty of 8 - 12 % of the results, it is not preferred to perform simple linear or third order interpolation of the means[1] of the results. Instead, a regression will be performed to model the relation between processing time and focus on security.

As was shortly described above, the boxplots can be roughly subdivided into three regions: 0 - 30% where the processing time increases strongly with focus on security, 30 - 75 % where the processing times are nearly constant or slightly decreasing with higher focus on security, and 75 - 100 % focus on security where the processing time increases slightly again. When a regression is performed, the line must take into account this behavior. Furthermore, the fit should be good, hence the value for Adjusted R^2 should be high and the root mean squared error (RMSE) should be low. Lastly, the model should not overfit, meaning the regression models behavior that actually is not present in the model. Different regression models, linear and non-linear ((a)symmetrical sigmoid, polynomial with different orders) have been investigated taking these requirements into account. The best fit to the mean processing times is a polynomial regression to the power 3. Higher orders showed overfitting behavior and other models had a worse fit in terms of adjusted R^2 and RMSE. The result of the third order polynomial regression through the means of the results is shown in Figure 6.2.



Adjusted R^2: 0.9257
$RMSE$: 0.3993

Figure 6.2: Polynomial regression on means of processing times distribution for different security focus

In Figure 6.2 the dots are the means of the average security times of the 250 simulations. The red line is the cube polynomial regression through these data points. With an adjusted R^2 of 0.9257, the fit is good. Therefore, the polynomial regression to the power 3 is accepted as a regression fit for the data.

In this figure the three regions are depicted even clearer. There is a strong increase in processing time between 0 and 30 % focus on efficiency. After this, the processing time clearly decreases with an increasing

[1] note the difference between the medians, displayed in Figure 6.1, and the means, displayed in Figure 6.2

focus on security, to later increase again, but less strongly. This is new evidence of the three distinguished regions that were discussed earlier.

It is interesting to analyze the different behavior of the system in these three regions. A difference in behavior is caused by differentiating parameters. The parameter on which sensitivity analysis is performed is the focus on security of the operators, so it must have something to do with the decision making of the operators. Decisions made by the operators can increase the mean processing time in two ways: (i) many positives result in many extra checks, which increase the mean processing time, and (ii) a longer response time results in longer processing times. It is therefore interesting to give an overview of the number of positives per bias and the length of the response times for different bias z. Such an overview is presented in Table 6.2.

Table 6.2: Average TPR, FPR and corresponding response times for 10,000 X-ray check operator (3 boxes) decisions for varying security focus

Security focus	$RT_{phy,forbid.}(t)$	TPR_{phy}	$RT_{phy,allowed}(t)$	$FPRphy$
0%	21.20	44.5%	14.06	1.6%
12.5%	25.37	68.2%	15.62	2.6%
25%	26.20	79.0%	17.11	3.1%
37.5%	26.49	85.4%	18.19	4.2%
50%	25.47	89.2%	19.41	5.3%
62.5%	23.75	92.4%	20.31	6.0%
75%	21.84	94.3%	21.05	7.5%
87.5%	19.49	95.7%	21.41	11.4%
100%	16.91	97.0%	20.69	20.1%

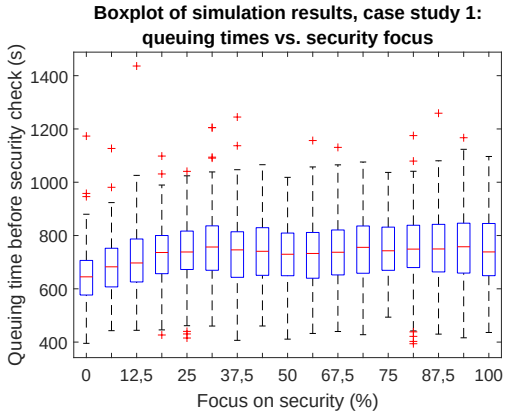
In Table 6.2, the average TPR and FPR and corresponding response times for allowed and forbidden items, are given for for 10,000 X-ray check operator (3 boxes)[2] decisions for nine different security focus set-ups. The diffusion decision process changes when the focus on security is changed: the bias z is varied. One sees that for low focus on security (between 0 and 30 %), an increase in security focus results in a large increase in hit rate. This is the explanation of the large increase in mean processing time for low focus on security: there number of hits increase strongly.

The medium focus on security (between 30 and 75 %) in Figure 6.2 is a very interesting region: while focus on security increases, the mean processing times decrease. A possible mathematical cause for this behavior can be found in Table 6.2. In the region of medium security focus, it is seen that TPR and FPR do not increase much with increasing focus on security. What changes however, is the speed at which the hits are found. Response times for forbidden items decrease strongly when the security focus (bias z) is increased. The mathematical explanation for this in the diffusion model is the following: a forbidden item stimulus has a relatively small variability v_f, so an increasing bias reduces the response time to a large extent. These shorter response times for forbidden items lower the average processing time, also because the major part of the objects that are checked twice are forbidden items. This can be a cause of the decreasing processing times in the medium focus region. But next to this theoretical explanation, this behavior can also be seen as global emergent behavior of the system: the non-linear interactions within the system cause this behavior, and further research can be performed to what exactly causes the system behavior in this region. When the security focus is further increased above 75%, the total amount of extra checks raises strongly again. Many passengers are checked. Although these decisions are made quickly, the large number of checked passengers causes the mean processing times to rise.

Next to mean processing times, the mean queuing times have also been measured. An insight to the influence of security focus on mean queuing times can be retrieved from Figure 6.3.

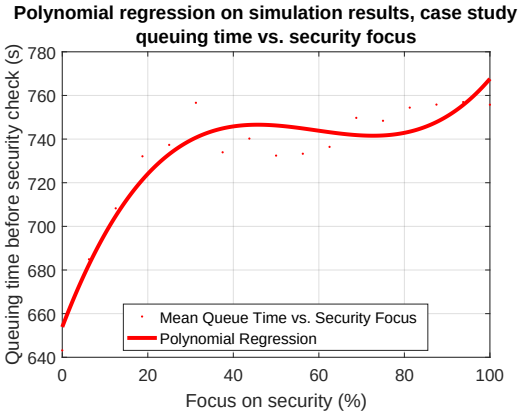
For the queue times the three similar regions can be identified. Again, clearly three regions can be identified: (i) strong increase, (ii) slight decrease and (iii) again increase in queue times with increasing focus on security. The shape of the queuing time vs. security focus is comparable to the processing time vs. security focus. This is as expected: the queue forms as a result of the security processes, so longer security processes result in longer queues with larger waiting times. It is interesting to see that an increase of 5 seconds in mean processing time already results in an extra mean queuing time of more than a minute. This indicates how sensitive the efficiency performance is to small increases. It can thus be concluded that even the smallest changes in mean processing times are important to investigate.

[2]Since the only difference in diffusion process calibration with other operators (luggage and X-ray check) is the non-decision time, comparable tables would be retrieved when decisions of these operators would be analyzed.



n = 250, 17 set-ups

(a) Boxplot of queue times distribution

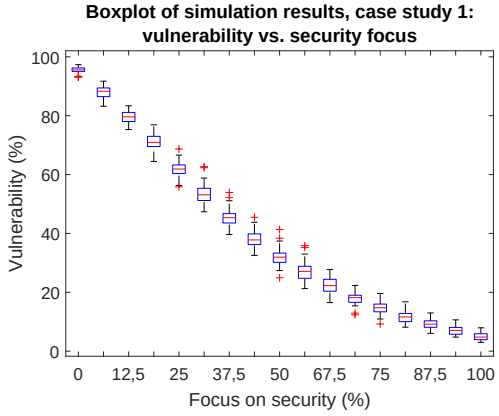


Adjusted R^2: 0.8714
$RMSE$: 10.6930

(b) Polynomial regression on mean queue times

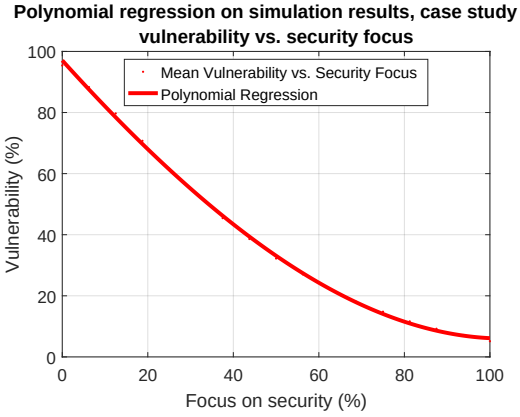
Figure 6.3: Relation between queuing time and security focus

But next to the processing and queuing time, the increase in focus on security is also of major influence on the security performance. The boxplot and the polynomial fit are shown in Figure 6.4.



n = 250, 17 set-ups

(a) Boxplot of vulnerability distribution



Adjusted R^2: 0.9991
$RMSE$: 0.9121

(b) Polynomial regression on vulnerability

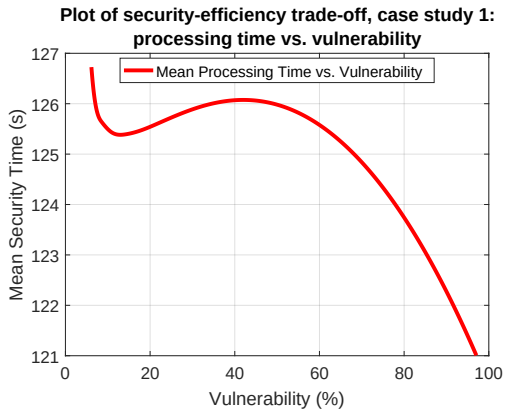
Figure 6.4: Relation between vulnerability and security focus

The result in Figure 6.4 confirms the expectations about the relation between focus on security and security performance. The higher the focus on security, the smaller the vulnerability. The cause is the varying bias z: the closer the starting point of the decision process to the upper boundary, the higher the probability on a hit. In Figure 6.4a it can be seen that for the extreme values of security focus, the uncertainty becomes smaller; the largest blue boxes are found in the area around 50 % focus on security. Another insight from Figure 6.4a is that two regions can be identified. For low focus on security, the gradient is larger. The turning point is around 70% focus on security. For high focus on security, an increase in focus on security contributes less in security performance than for low focus on security. This second region (high focus on security) even shows a little convexity. This is not strange: it is logic that 0% vulnerability is an asymptote.

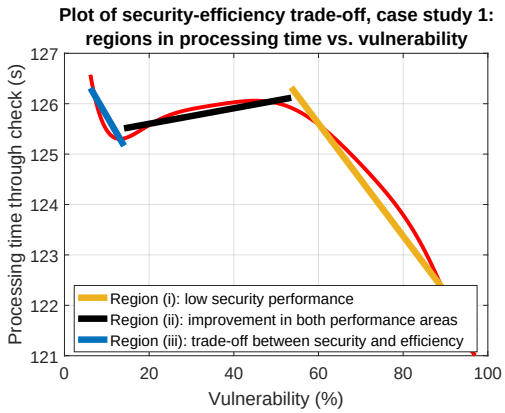
The model is calibrated based on literature that says that the worst performing operator in terms of security will have a hit rate with a minimum value of 45 % (read vulnerability of 55%). The reason that for a very low focus on security, the vulnerability crosses this boundary is because the total vulnerability is measured. This is the combination of the vulnerability of the luggage check (X-ray operator + luggage check operator,

meaning 2 times a diffusion decision) and the physical check (WTMD + physical check operator, meaning a combination of diffusion decision and signal detection). Because all four operators types are calibrated on the above described data, vulnerability can approach 100 % for small focus on security. It can however be assumed that airports will not choose to be in the region below 30 - 40 % of focus on security, as they will take their task of ensuring security seriously.

As was done in the analysis of the preliminary model, efficiency performance can be plotted vs. vulnerability, since both are varied for different focus on security. The result is shown in Figure 6.5.



(a) Security-efficiency trade-off for varying focus on security



(b) Regions in security-efficiency trade-off for varying focus on security

Figure 6.5: Security-efficiency trade-off graphs for varying focus on security

In Figure 6.5 the security-efficiency trade-off for varying focus on security is plotted. The first insight is that it has a different shape than the security-efficiency trade-offs that were identified in the preliminary research. Region (ii) and (iii), of 0 % < vulnerability < 40 %, together can be compared in shape to the shape that was found in the preliminary research, but a new insight is found for the region > 40 % vulnerability. Here the mean processing time increases exponentially when decreasing vulnerability. This negative relation between security and efficiency performance is the consequence of the behavior of the system at low security focus. When the security focus is low, both processing time increases and vulnerability decrease with increasing security focus. This combined effect causes a strong negative relation between security and efficiency.

The division in three regions is made graphical in Figure 6.5b. The middle region, between a vulnerability of 12 % and 40 % provides an interesting new insight. In this region, both security performance and efficiency performance improve with increasing security focus. The particular security and efficiency behavior of the system in this region causes this remarkable gradient of the slope. On the one hand, vulnerability decreases with increasing security focus. But because the efficiency performance also increases as shown in figure 6.2, increasing security focus in this reason is a win-win situation. A possible mathematical explanation for this behavior was given in the analysis of Figure 6.2, namely the slight increase of FPR and TPR, and the decrease of response times for forbidden items. But apart from the outcome of these decisions and their response times, the system is also dependent on non-linear interactions between the agents within the security check. The cause of this global emergent behavior of the middle region is an interesting one for further research. In the following case studies, special attention will be paid to how the system behaves in this region (ii).

The final region, region (i) is the exponential increase of security time when further decreasing vulnerability. Here there is a clear trade-off, where security performance can be traded against efficiency performance. The rate in which this is changed is however not constant. As was the case in the preliminary model, the closer one gets to a vulnerability of 0 %, the higher the cost is in terms of efficiency performance. This is the most ideal place for airports to find themselves in: if an airport performs in region (i), the security performance is not sufficient and if an airport performs in region (ii) it can perform better both in terms of security and efficiency by focusing more on security. Where the airport exactly wants to be in region (iii) depends on the chosen focus on security. Region (iii) is coherent with the increase in processing time in Figure 6.2: it is caused by a security focus of > 75 %.

Because this region is the most interesting one for airports to focus on, a zoom on this region is visualized in Figure 6.6. In this figure it is seen that the gradient of the slope increases strongly when vulnerability is further decreased from 13%.

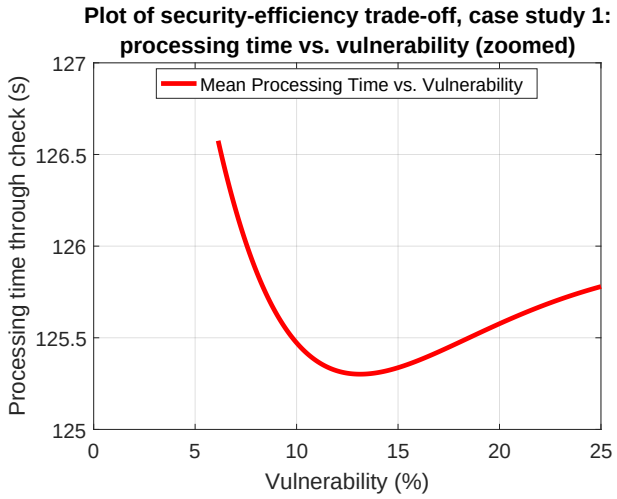


Figure 6.6: Security-efficiency trade-off graphs for varying focus on security, zoomed

6.2. Case study 2: Speed - accuracy trade-off

The second case study analyzes the speed-accuracy trade-off using the diffusion decision process of the operators. The $thres_{threat}$ and bias z are fixed for the different set-ups to the standard value of respectively 1.966 and 0.4522. In this case study, in which 9 set-ups are made, the decision boundary a is adjusted. Because a local sensitivity analysis is performed to measure the effects of the shift in boundary, this boundary shift may not imply a shift in bias at the same time. Therefore, also the lower decision boundary b needs to be varied for the different set-ups. In the set-ups, how much a is moved upward is equal to how much b is moved downward, and vice-versa.

It can be expected that for small boundaries (speed focus) a small increase in width between the boundaries has a larger effect on the trade-off than for wide boundaries. This means that the step size with which the boundaries are adjusted in different set-ups should be variable. No literature or data was to be found about minimum or maximum values for a in the diffusion process for the speed-accuracy trade-off. The values for a thus need to be determined arbitrarily. This has been set-up as follows. The standard value of a is 0.9585. Instead of increasing and decreasing this value by adding/subtracting a fixed step size, this value is multiplied with a fixed factor for different step sizes. In the first iteration of this case study, the multiplication chosen multiplication factor was 3. The results of the first iteration showed that this step size was too large, and therefore set-ups were placed between the set-ups of the first iteration. This resulted in an eventual step size of $\sqrt{3}$.

The lowest value of a was chosen to be 0.5325, which is four steps away from the standard value of a = 0.9585. One step further would result in a upper boundary of lower than the standard value for bias z, so this was not possible. To keep the proportions balanced, the maximum value for a is also four steps away from the standard value, resulting in a maximum upper boundary a of 4.7925. As b needs to be adjusted accordingly to $-3.834 \geq b \geq 0.4260$.

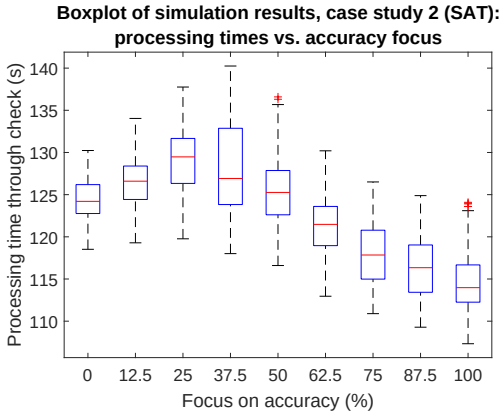
Table 6.3: Variable parameters for case study 2

Variable parameter	0 % accuracy focus	100 % accuracy focus	Step size (12.5 %)
a	0.5325	4.7925	$\sqrt{3}$
b	0.4260	-3.834	$-\sqrt{3}$

Result of case study 2

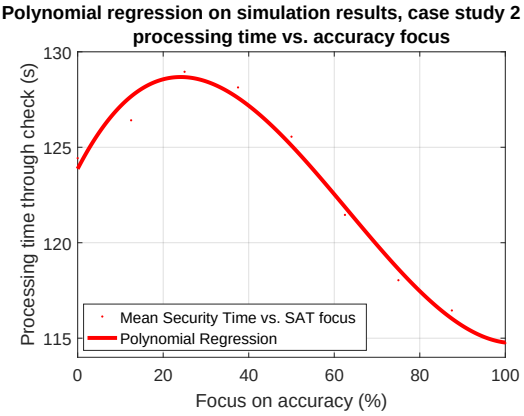
First the efficiency performance will be analyzed. As was done in the first case study, the boxplot and the polynomial fit to the mean of the average processing times will be presented in Figure 6.7.

For analysis of the shapes of Figure 6.7, let's start with a system in which every operator is 50-50 focused on accuracy and on speed. If the operators choose to focus more on speed, hence less on accuracy, the mean processing time of the entire system will go up. The reason for this is that a focus on speed might result in quicker decisions, but also in less accurate decisions. Hence, although the operator decides more quickly,



n = 250, 9 set-ups

(a) Boxplot of processing times per accuracy focus



Adjusted R^2: 0.9855
$RMSE$: 0.7963

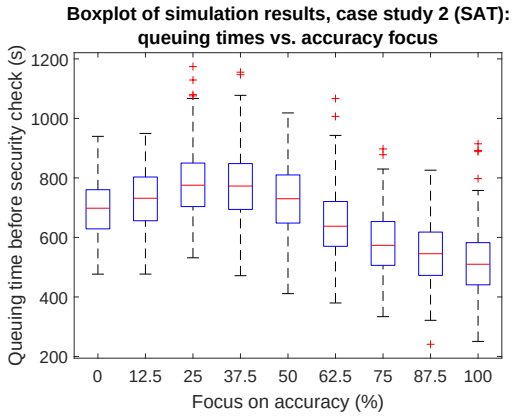
(b) Polynomial regression on mean processing times

Figure 6.7: Relation between mean processing times and accuracy focus

he also decides wrongly and sends more allowed luggage and passengers for extra checks than necessary. If the operators focus more and more on speed, after a while (accuracy focus < 25 %) the mean processing time drops. This is caused by the extreme decrease in response times. Every operator, even the operators who perform the extra checks, decide very quickly. Therefore, processing time can be saved when focusing on speed extremely.

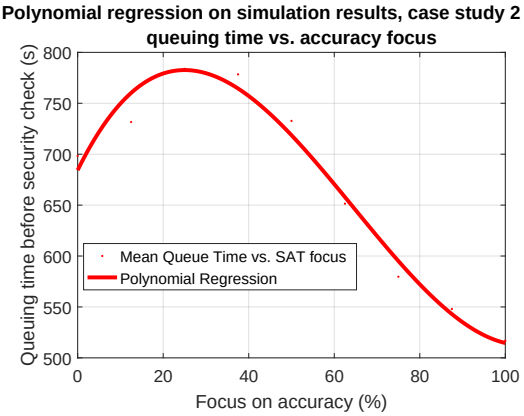
For the analysis, let's go back to the situation of the system in which every operator is 50-50 focused on accuracy and speed. If the operators choose to focus more on accuracy, this will benefit in terms of efficiency performance. The reason for this is that it will decrease the number of false alarm rates, hence the number of extra checks. This relation is linear up until a focus on accuracy of 85 %. An extra focus on accuracy will not benefit more in terms of efficiency: hardly any wrong decisions are made at a complete focus on accuracy and thus there is no more time to win. This can again be seen as an asymptote.

The influence of focus on accuracy on the queuing times is visualized in Figure 6.8



n = 250, 9 set-ups

(a) Boxplot of queuing times per accuracy focus



Adjusted R^2: 0.9657
$RMSE$: 18.5347

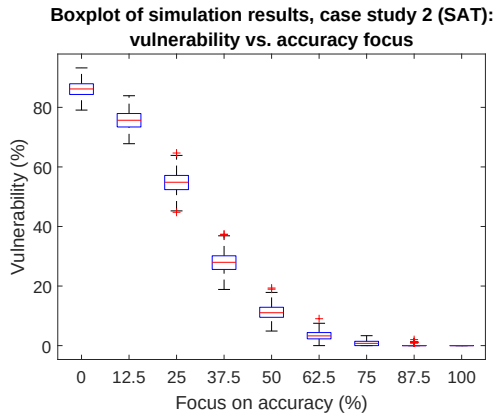
(b) Polynomial regression on mean queue times

Figure 6.8: Relation between queuing times and accuracy focus

As was the situation in case study 1, the relation between the mean processing times and focus on accuracy, and the queuing times and focus on accuracy is identical. The only difference is the values on the y-axis: again it is seen how sensitive the queuing system is to even only seconds increase of processing time.

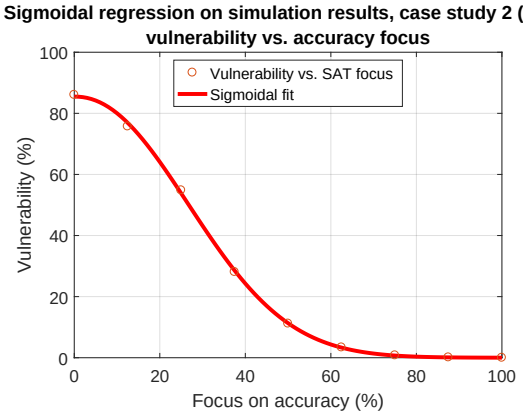
An improvement of more than four minutes average waiting time is obtained when increasing the focus on accuracy from 28 % to 100 %.

The influence of the focus on accuracy on the security performance (vulnerability) is shown in Figure 6.9. Note that in this case study the measure for vulnerability is only the vulnerability of the luggage check. The reason for this is that only the luggage check consists of two diffusion modelled decision makers (X-ray and luggage check operator), while the passenger body check consists of a signal detection device (WTMD) combined with a diffusion decision maker (physical checker). Also considering the passenger body check would thus have not given extra information.



n = 250, 9 set-ups

(a) Boxplot of vulnerability distribution



Adjusted R^2: 0.9993
$RMSE$: 0.8621

(b) Polynomial regression on vulnerability

Figure 6.9: Relation between vulnerability and accuracy focus

In Figure 6.9a it can be seen that an increase in focus on accuracy ensures a lower vulnerability. This follows from the definition of the diffusion model, in which widely spread decision boundaries a and b lead to fewer false decisions. Furthermore it is seen that the spread of the results becomes smaller for higher accuracy focus. It is viewed in Figure 6.9a that for an extreme focus on accuracy (87.5 % and 100 %), almost all results show a vulnerability of 0% of the luggage check system. This is not a very realistic situation, as obtaining a vulnerability of exactly 0 % is hardly possible. It could therefore be concluded that the specified boundaries a and b for a 100 % accuracy focus are chosen too widely. It does however give an indication about the importance of the focus on accuracy.

It was not possible to fit a linear or polynomial regression to the data points mean vulnerability for different accuracy focus. Therefore, a non-linear regression was performed to fit the data, namely an asymmetrical sigmoidal function. It is seen in Figure 6.9b that this function gives a very good fit with an adjusted R^2 of 0.9993.

Vulnerability is plotted against mean processing time for varying accuracy focus in Figure 6.10

Except from the high region of vulnerability > 75%, one sees that it is generally true that a decrease in vulnerability also leads to a decrease in processing times in the speed-accuracy trade-off. This is caused by the increase in focus on accuracy. The more the operators focus on accuracy, the less vulnerable the system will be, and the smaller the mean processing time will be because of the fewer false positives. This is an interesting insight: when one wants to increase efficiency performance, one should not work at a higher speed, but more accurately. The results show that, although response times will be longer, this weighs against the reduced time due to fewer false positives.

Figure 6.10 shows that the more the vulnerability is decreased, the more the processing time is decreased. When looking at the shape, it suggests that the processing time decreases more and more when vulnerability is decreased further. It seems to imply as if processing time can be brought to zero if vulnerability is brought to zero. In reality, this situation will never occur. People need time to make decisions, and people make wrong decisions. To show that this point of perfect efficiency and security performance will not be reached by the model, the data points have been plotted extra bold in Figure 6.10. In this way it can be seen that, although the steps between different speed-accuracy set-ups are equal, the data points come closer to each

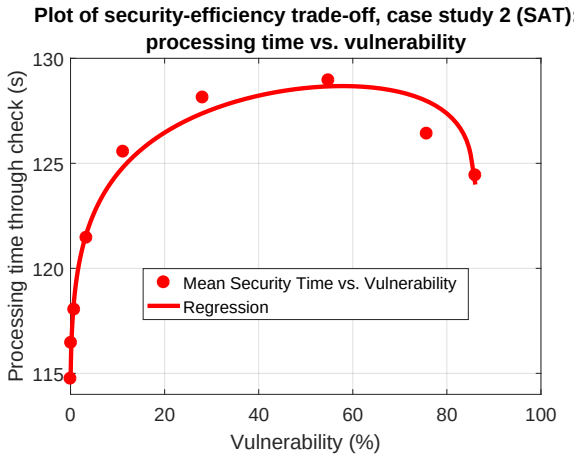


Figure 6.10: Security-efficiency trade-off for varying accuracy focus

other when 0 % vulnerability is approached. This shows that further increasing the accuracy focus when this focus is already high will have fewer effect than it had when the accuracy focus was lower.

Although the perfect efficiency and security performance will not be reached in reality, it is not strange that the line tends towards this point for extreme high accuracy focus. A thought experiment can be done to the perfect operator. The perfect operator makes no mistakes and needs no time to perform the check. The perfect operator is thus an infinitely accurate operator: he makes no mistakes, and since he never makes a mistake in collecting wrong evidence, he also can make the decision in no-time. This is the explanation of the direction of the curve towards perfect security and efficiency performance.

The diffusion decision process

From the speed-accuracy trade-off case study it can be concluded that it will always be in the benefit of the system to focus on accuracy and not on speed. Although focusing on speed might increase the efficiency performance of the system, it is at a very high cost of security performance. This is an interesting insight that will further be discussed in Chapter 7.

6.3. Case study 3: Impact of fatigue on operators

The third case study repeats the first case study, but then compares fit operators with tired operators. Fit operators are operators that start working at 05:00 and work at a flight peak at 07:00. Tired operators are operators that have to handle the same number of flights and passengers, but who start at 17:00 and handle the peak at 19:00.

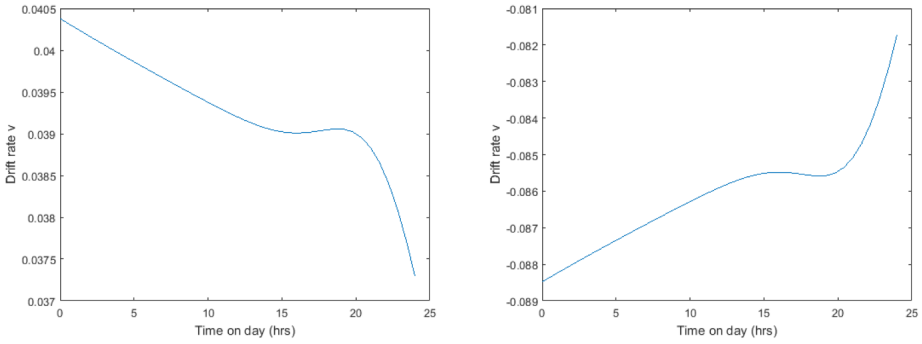
These operators are modeled to be more tired by adapting their drift rate with a fatigue scaling factor, as explained in Section 5.1. When operators make a decision in the simulations, they first determine their drift rate v. For determining their drift rate v, first the current fatigue score F should be determined. The fatigue score is obtained by providing the time at which the decision needs to be made to a fatigue score determining algorithm, that follows the shape of Figure 5.2. Knowing the fatigue score, the drift rate can be determined based on the stimulus (forbidden or allowed item) and the dynamic drift rate equation (Equation 5.1).

The resulting graphs for the drift rate of tired and fit operators for different stimuli are plotted against the time of the day in Figure 6.11.

Having configured the above, again 17 set-ups with varying bias z and thres_{threat} are constructed with the same values for z and thres_{threat} as in case study 1. The same base conditions are true for both the 250 simulations for fit and for tired operators. In this way the behavior of the system with tired operators can be compared very well to the behavior of the system with fit operators.

Table 6.4: Differentiating parameters for case study 3

Operator type	Start time	Peak handling
Fit operator	05:00	07:00
Tired operator	17:00	19:00



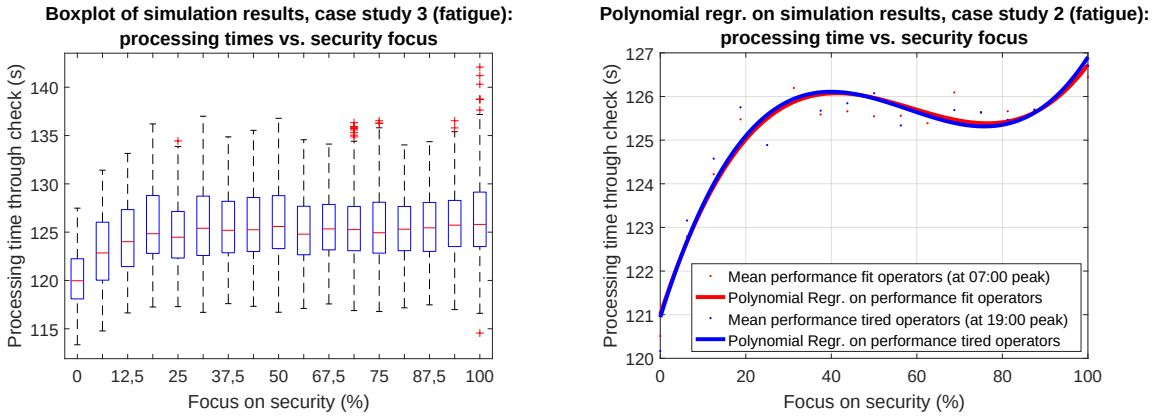
(a) Dynamic drift rate over day for forbidden item stimulus

(b) Dynamic drift rate over day for forbidden item stimulus

Figure 6.11: Dynamic drift rates for different stimuli

Result of case study 3

The results in terms of efficiency performance can be found in Figure 6.12.



n = 250, 17 set-ups

Adjusted R^2: 0.8891
$RMSE$: 0.4782

(a) Boxplot of processing times per security focus for tired operators

(b) Polynomial regression on mean processing times for tired operators

Figure 6.12: Relation between mean processing times and security focus for fatigued operators

In both graphs of figure Figure 6.12 it can be seen that the shape of the efficiency performance for varying focus on security does not differ very much for tired operators compared to fit operators. There are however some differences: for low focus on security, the efficiency performance is worse for tired operators. The gradient in which focus on security decreases efficiency performance is stronger for tired operators than for fit operators. The next region also shows more extreme behavior: between 30 and 75% focus on security, the negative gradient is also stronger. It is even so negative that for a security focus between 60 and 85%, tired operators perform better in terms of efficiency performance than fit operators. Then for high focus on security, again, the graph is steeper. It can thus be concluded from Figure 6.12 that the effect of focus on security on the efficiency performance is more extreme for tired operators than for fit operators.

What is the cause of this more extreme behavior of the system? The first region, 0 - 30% focus on security, was identified in case study 1 to have a steep gradient because of the large increase in hits when increasing security focus in this region. A larger portion of the forbidden items proceeds for extra check, so the efficiency performance decreases. The cause for the steeper slope for tired operators is their relatively large required RT: because the number of checks is increased, there are more checks and these checks take longer.

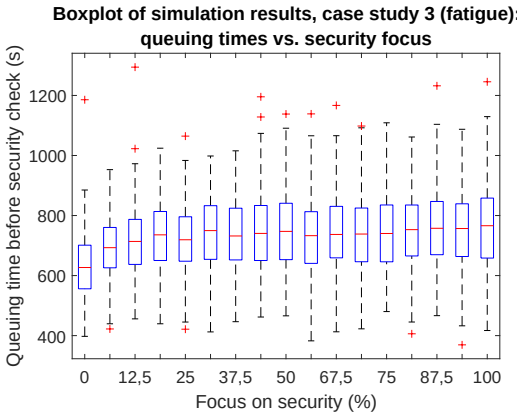
The next region between 30 and 75% is also very interesting. In this region, the decrease in processing time for increasing focus on security is stronger for tired operators than for fit operators. In case study 1

it was identified that this increase in performance is caused by the lower response time of the decisions of the operators, caused by the increased bias z. Tired operators generally have lower response times than fit operators because of their lowered absolute drift rate v. Therefore, if the response time of these operators is decreased due to increasing bias z, this has a large effect on the operators. In this region it means a more extreme effect: a more negative slope for processing time with increasing security focus.

For the last region the same is true as for the first region. At security > 75%, an increase of security focus leads to a large increase in checks. As these checks have a larger response time, the processing time increases strongly.

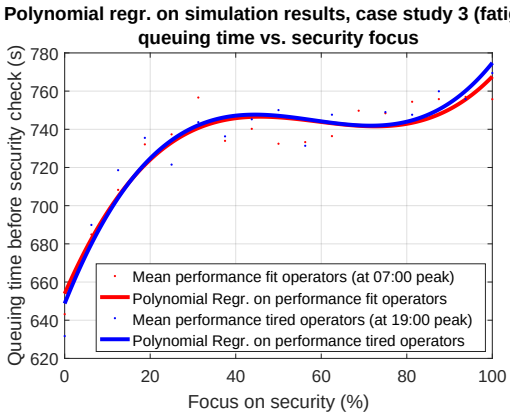
The conclusion is that tired operators are more sensitive to changes that affect their response time. The response time of tired operators is larger because of a lower absolute value for drift rate v. This makes the response time more sensitive to changes of the bias z, which causes more extreme behavior in terms of efficiency performance when compared to fit operators.

To investigate how these security times propagate in queue times, average queue times are taken as an efficiency performance parameter and plotted against security focus in Figure 6.13.



n = 250, 17 set-ups

(a) Boxplot of queuing times per security focus for fatigued operators



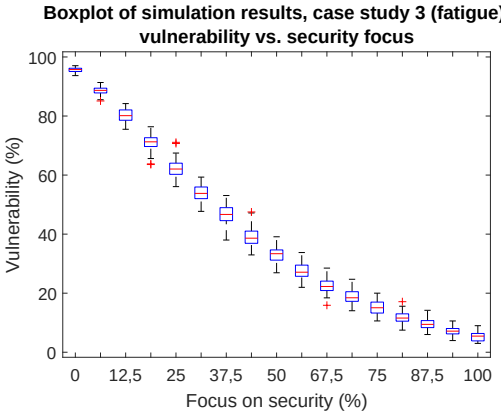
Adjusted R^2: 0.8891
$RMSE$: 10.6506

(b) Polynomial regression on mean queuing times for fatigued operators

Figure 6.13: Relation between mean queuing times and security focus for fatigued operators

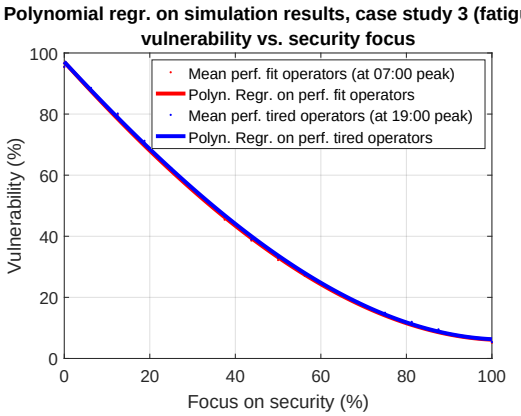
In Figure 6.13 the same behavior is seen as was identified in Figure 6.12 and case study 1. Again three regions are visible. The difference with respect to Figure 6.12 is that the queuing times for tired operators never become lower than that of fit operators for security focus > 15%. Apparently higher response times have a larger influence on the queues than on the processing time. For a 100% focus on security, the difference in queue time is 5 seconds. This means that all passengers experience on average a queuing time of 5 seconds more when the operators are tired.

The effect of the fatigue of operators in terms of security performance is plotted in Figure 6.14 on the next page.



n = 250, 9 set-ups

(a) Boxplot of vulnerability distribution for fatigued operators



Adjusted R^2: 0.9991
$RMSE$: 0.9018

(b) Polynomial regression on vulnerability for fatigued operators

Figure 6.14: Relation between vulnerability and security focus for fatigued operators

It can be viewed in Figure 6.14 that for all focuses on security, the system is more vulnerable with tired operators than with fit operators. The difference in result is however very small, as it can be seen that the red line is almost completely covered by the blue line. There is however a difference: for a more clear illustration, the values of the vulnerability for different security focus for fit and tired operators are given in Table 6.5

Table 6.5: Vulnerability difference fit and tired operators

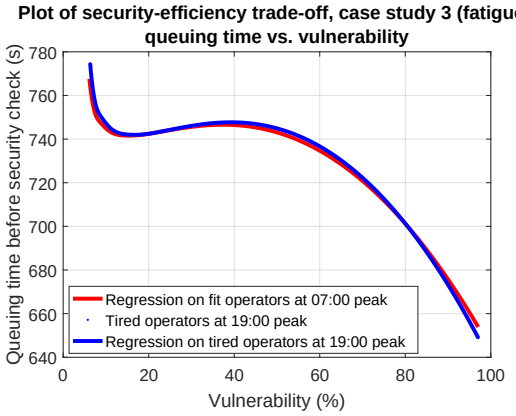
Security focus (%)	Vul. fit op. (%)	Vul tired op. (%)	Diff. (perc. point)
0	97.06	97.11	0.05
12.5	78.42	78.86	0.43
25	61.40	62.02	0.62
37.5	46.21	46.88	0.66
50	33.11	33.71	0.60
62.5	22.32	22.79	0.47
75	14.07	14.40	0.33
87.5	8.60	8.81	0.22
100	6.14	6.31	0.17

From Table 6.5 can be concluded that the largest difference in vulnerability emerges for medium values of security focus. The reason for this lies in the properties of the diffusion process for decision making. For high or low focus on security, the bias is very close to either boundaries. Evidence is thus collected for a relatively short time interval. For medium focus, evidence is collected for a large time interval. Therefore, a change in drift rate v has a large influence on the collected evidence, and thus on the outcome of the decision.

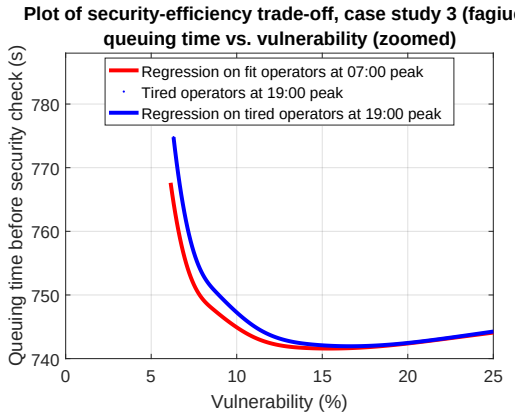
The efficiency performance is plotted against the vulnerability performance in Figure 6.15. Because the effect of fatigue on queuing time was most significant, average queuing time is chosen as the efficiency parameter.

In Figure 6.15 it can be viewed that the shape of the security-efficiency trade-off is the same for tired operators as for fit operators. The difference is that for the most important part, the curve with the tired operators is shifted right and up with respect to the fit operator curve. This is due to the worse performance in both efficiency and security of tired operators.

An interesting insight is obtained when zoomed in on the region with low vulnerability, as is done in Figure 6.15b. Here it can be viewed that for a vulnerability of > 15% the system performance hardly differs for fit and tired operators. Although the tired operators curve has been shifted right and up, the two curves intersect in this region. This means that for this vulnerability region, if operators choose to focus more on security, they can obtain the same security and efficiency performance as they could when they were fit, but focusing less on security.



(a) Security-efficiency trade-off for fatigued operators



(b) Security-efficiency trade-off for fatigued operators, zoomed

Figure 6.15: Security-efficiency trade-off graphs for fatigued operators

A difference in the two curves emerges for vulnerability < 15%. Here both the security and efficiency performance becomes worse when operators are tired. This is an important insight, as most airports will try to have their operations perform in this research. Example implications are the following. A fit operator that is performing at a security focus of 93.7%, his security performance will be a vulnerability of 7%. If the same operator wants to maintain this security performance at 19:00, he will have to shift his security focus to 94.6%. This is however at a cost: his efficiency performance will decrease because of (i) his fatigue leading to larger response times and (ii) his shift to a higher focus on efficiency. The result is that the average waiting time for the passengers increases with 6.73 seconds. During a regular flight day on RTHA in which almost 1,000 passengers travel during a peak, this means that there is more than 6,000 seconds of total extra waiting time: this is a bad hospitality, and this is time that passengers could also spend in shops.

This result provides insight on the varying performance of operators during the day. The results can be used to anticipate on different performance over different periods on the day, for example by using more resources to cope with the worse performance. Another example implication is the possibility to investigate the effect of an extra break for security operators. These implications and recommendations will further be discussed in Chapter 7.

6.4. Case study 4: Diverse passengers

In the fourth case study not the operators but the passengers are modelled to be diverse. As was discussed in Chapter 5, passengers who are on a business flight have a much smaller probability of requiring an extra check than passengers on a charter flight. This will be of large influence on the performance of the system: more checks means longer security processing times and thus longer queues.

To find the implication of these different types of flights to the performance of the security system, three different types of flight days have been set-up. The first is an average flight day, which has been used for case study 1 - 3. On the average flight day, the probability that a piece of luggage or a passenger's body contains a forbidden item ($p^{average}_{forbidden}$) = 0.2067. For charter flights, $p^{charter}_{forbidden}$ = 0.4167 and for business flights $p^{business}_{forbidden}$ = 0.1270[35]. The behavior of these different types of flight days for varying security focus will be modeled, just as in case study 1 and 3 for varying $thres_{threat}$ and bias z. The difference of a charter and business flight day with respect to an average flight day will be investigated and made clear.

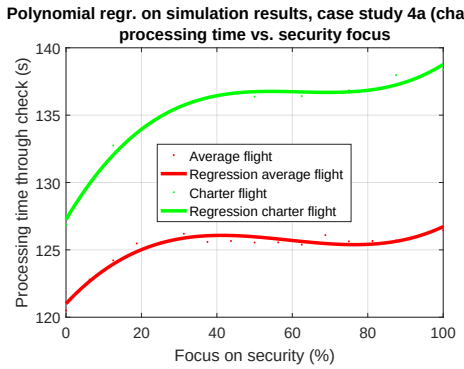
Note that these probabilities of a forbidden item are both for a piece of luggage and for a passenger's body. In the preliminary model, a passenger was either a forbidden passenger (having both forbidden luggage and forbidden body) or an allowed passenger (having allowed luggage and body). But in this case study, the random probabilities are considered separately for forbidden luggage and bodies. This was necessary because the decision about the body of the passenger is made based on threat level (by the WTMD) and influenced by a stimulus from the type of passenger (at the physical check), while the decision about the luggage is only influenced by the passenger's type. This results for example for average flight days, that 4.3% of the passengers contain both a forbidden luggage and a forbidden item on his body, 16.4% carries only forbidden luggage, also 16.4% only has a forbidden item on his body, and 62.9% carry no forbidden item.

Table 6.6: Differentiating parameters for case study 4

Flight day type	$p_{forbidden}$
Average	0.2067
Charter	0.4167
Business	0.1270

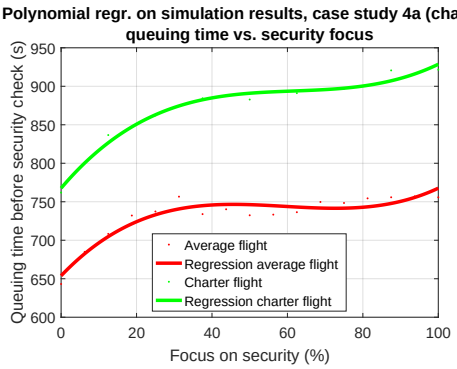
Result of case study 4

In Figure 6.16 the results of the efficiency performance for charter flights are compared with those of an average flight.



Adjusted R^2: 0.9743
$RMSE$: 0.5665

(a) Polynomial regression on mean processing times for charter passengers



Adjusted R^2: 0.9635
$RMSE$: 9.4386

(b) Polynomial regression on mean queue times for charter passengers

Figure 6.16: Relation between efficiency performance s and security focus for charter passengers

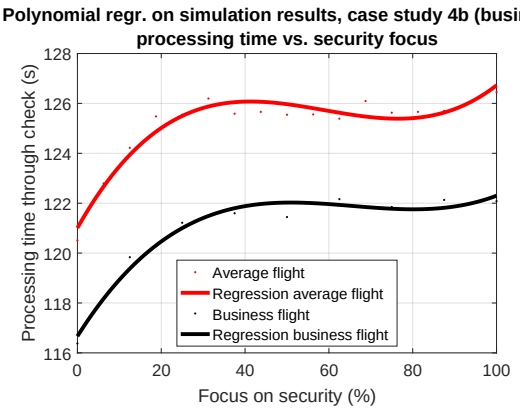
The first obvious observation is that for both the security processing time as the queuing time, charter flights will perform worse in terms of efficiency. The difference is significant: the processing time increases with 8% on average. This has an even larger effect on the queuing time, which increases on average with almost 20%. The type of flight is thus of large influence on the queues before the security check.

The three regions that were earlier identified in case study 1 and 3 are also present in the charter flight curve for efficiency performance. Region (i) and (iii) for low and high focus on security have a comparable shape. Region (ii) however, for medium focus on security between 40 and 80%, shows different behavior on a charter flight day. In this region the mean processing time remains constant with increasing focus on security, while on an average day, the slope was negative. For queuing times, the relation between focus on security and queuing time is even positive for charter flight days in this region, while an average day shows a negative line here. Apparently the increased number of forbidden items on charter flight days makes that also in this region the efficiency performance decreases when increasing focus on security.

The efficiency performance of business flights is compared to that of average flight days in Figure 6.17.

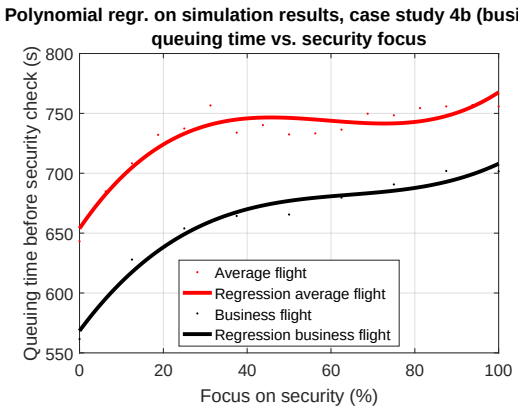
The business flight days result in quicker operation times: the average processing times are 3% lower, while the average queuing times decrease with 10%. This difference is smaller than the difference was between charter flights and average flights. The reason for this is that the $p_{forbidden}$ only decreased from 0.2067 in average flights to 0.1270 in business flights, while it was 0.4167 for charter flights. Changing $p_{forbidden}$ from 0.2067 to 0.1270 (60 % of the average value), thus has smaller influence than changing $p_{forbidden}$ from 0.2067 to 0.4167 (202 % of the average value). It can thus be concluded that the efficiency performance decreases proportionally with $p_{forbidden}$.

Again the curves in both graphs have comparable shapes. Again three regions can be identified. The shape of the regions with very high or very low focus on security for business flights are almost identical to the same regions for average fight days. But as was the case for the charter flights, the medium focus region behaves differently for business flight days, certainly when measured in average queuing time. Apparently the number of forbidden items is of large influence to the behavior of the system in this region of security focus:



Adjusted R^2: 0.9469
$RMSE$: 0.4299

(a) Polynomial regression on mean processing times for business passengers



Adjusted R^2: 0.9696
$RMSE$: 9.7379

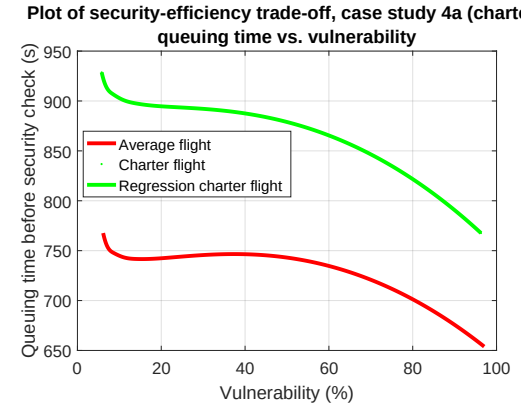
(b) Polynomial regression on mean queue times for business passengers

Figure 6.17: Relation between efficiency performance s and security focus for business passengers

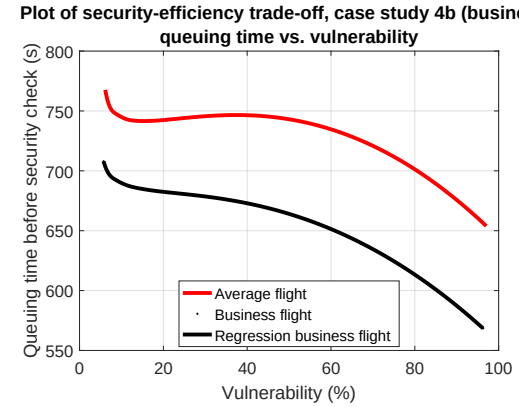
both for a small and a large number of forbidden items, the curve becomes more positive in this region. It is interesting to further investigate what exactly triggers this behavior.

The security performance of the system for different flight days is also investigated. The result was that hit rates did not differ significantly over different flight days. This was according to expectation, because the only variable parameter in this case study is $p_{forbidden}$. Where the efficiency of the system is largely dependent on this parameter, security performance is not: security performance is measured in vulnerability, and the % of hits does not change if more forbidden items are checked. The reason for this is that the operator decision performance is not influenced by the number of forbidden items the operator gets presented. It is however interesting to investigate how security performance depends on the number of forbidden items checked. This could very well be combined with case study 3, to the (decision) fatigue of operators. This will be proposed in Chapter 7.

Although the security performance did not change, it is interesting to take a look at how the security-efficiency trade-off holds for different type of flight days. The results for both charter and business days are plotted in Figure 6.18. The efficiency performance measure is average queuing time. This has been done because average queuing time can be an important measure that security managers can use to anticipate their planning.



(a) Security efficiency trade-off for charter passengers



(b) Security-efficiency trade-off for business passengers

Figure 6.18: Security-efficiency trade-off for flights with types of passengers

As explained, the vulnerability is comparable on the different flight days. This results in the security-efficiency trade-off graphs to be of comparable shape, except for the segments in which the efficiency performance differs from the average flight days. For both charter and business flight days, it can be seen that also at medium to high vulnerability (20 - 70%) the efficiency performance decreases when security increases. There is thus no region in which both security and efficiency performance increase when the security focus is increased. This is interesting emergent behavior, which will further be treated in Chapter 7.

6.5. Empirical evaluation of case studies: position of real airport security operators in trade-off

The final case study is different from the first four case studies. In the final case study, the result of an empirical research performed by another MSc student is combined with the result of this thesis research[34]. In the empirical research an investigation was performed to how real security operators at Rotterdam The Hague Airport make trade-offs in terms of security and efficiency. The results are compared with the found trade-offs in this thesis research. This evaluation thereby serves as a validation of the results and can give useful insights for practice.

In the empirical research, firstly the four most important security and efficiency performance measures are distinguished. These showed to be:

- Hit rate (TPR, equals (1 - vulnerability));
- False positive rate (FPR);
- Missed flights;
- Waiting time.

Next discrete choice modelling and utility theory is applied to investigate how security operators make trade-offs between these performance measures. 70 security operators at RTHA filled out an inquiry that was set-up in such a way that these trade-offs became apparent. The result of the empirical research is a quantification of the six trade-offs, between each of the performance areas.

Of these six trade-offs, two are of particular interest to combine with this thesis research, because these trade-offs are investigated as well:

- The trade-off between hit rate and false positive rate (TPR - FPR);
- The trade-off between hit rate and waiting time (TPR - waiting time).

The identified trade-offs in the empirical research are defined as follows. Based on the used discrete choice model (mixed logit, or latent class model are used), the outcome of the TPR-FPR trade-off is that operators are willing to trade X percentage point of hit rate (TPR) for Y percentage point of false positive rate (FPR). For the TPR - waiting time trade-off, the result is that operators are willing to trade X percentage point of hit rate (TPR) for Z minutes of average waiting time for a passenger.

The use of the empirical research in this thesis research will be the following. In the ROC curves that were found in Section 5.3, the gradient of the ROC curve can be seen as the magnitude of the TPR-FPR trade-off of the operators. A steep slope means that operators want to trade 1 percentage point of TPR for a small number of FPR percentage points: an increase in TPR is obtained by a small increase in FPR. When the slope is more flat, this indicates that operators value TPR very highly, and that TPR is traded for a large number of FPR percentage points. Knowing the number of TPR percentage points that an operator wants to trade for FPR percentage points, it can be assumed that the operators will converge to that value in te ROC-curve where the trade-off is exactly made in the way they want to make it. The location of an operator on the ROC-curve is thus the location where the trade-off of the operator coincides with the slope of the ROC-curve.

The same is true for for the TPR-waiting time trade-off. The gradient of the security-efficiency trade-off as found in case study 1 can be seen as the magnitude of the trade-off of the operators. If the slope is very steep, operators want to change TPR only for a large amount of waiting time. If the slope is rather flat, operators are already willing to change a percentage point of TPR for only small decrease in waiting time. The position of an operator on the security-efficiency trade-off curve is in this case thus the position where the trade-off of the operator coincides with the slope of the security-efficiency trade-off curve.

As stated, both a mixed logit model and a latent class model was used in the empirical research. The mixed logit model was used to determine how the average operator on RTHA makes his trade-offs. Using this result, the location of the average RTHA operator on the ROC-curve and on the security-efficiency trade-off curve is determined. The result is presented in Section 6.5.1.

The latent class model was used to distinguish between different types of operators at RTHA. Three different operator types were distinguished:

- Passenger level of service sensitive operators (28% of RTHA operators);
- Highly security focused operators (59% of RTHA operators);
- Highly efficiency focused operators (13% of RTHA operators).

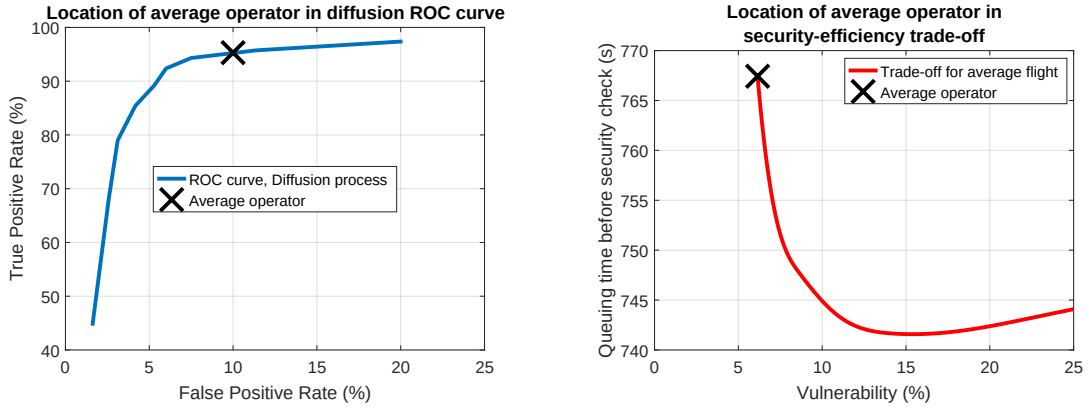
As was done for the average operator, the three different operator types will also be located on the ROC-curve and on the security-efficiency trade-off curve. An analysis is performed on the how the behavior of the operators influences their placement, and the differences in placement between the ROC-curve and the security-efficiency trade-off curve are elaborated. The results and analysis are presented in Section 6.5.2.

Combining the empirical research with this thesis research serves as an example on how future empirical researches can be used in combination with this modelling type of research. Furthermore, it serves as a validation in the sense that it is checked if real security operators can actually be located in the trade-offs that were identified by modelling.

6.5.1. Position of average security operators in trade-offs

The outcome of the empirical research to average operators was that the average operator is willing to trade 1 percentage point of TPR (read: vulnerability) against 3.71 percentage point of FPR. This means, that if they are willing to decrease TPR with 1 percentage point, if this results in an improvement of 3.71 percentage point in FPR. It is assumed that because operators are willing to make this trade-off, they will converge their behavior up to a point in which this trade-off is made.

The ROC-curve is the curve of trade-offs that are made between TPR and FPR. The average operator is thus assumed to take place in the ROC-curve at the position where the above described trade-off is exactly made. The proper gradient thus needs to be found. This is done in Figure 6.19a.



(a) Diffusion ROC curve with position of average RTHA security operator

(b) Security-efficiency trade-off with position of average RTHA security operator

Figure 6.19: Positions of average RTHA security operators in trade-offs

It can be viewed in Figure 6.19a that when 1 percentage point TPR is traded against 3.71 percentage point FPR, the average operator will converge to a hit rate of 95%, while having a FPR of 10%. This corresponds to an security focus of 85%. The average operator is thus relatively focused on security, resulting in fairly good results in terms of security performance. The consequence is a FPR of 10%, which will lead to extra processing times and queuing times, but the average operator is willing to accept this.

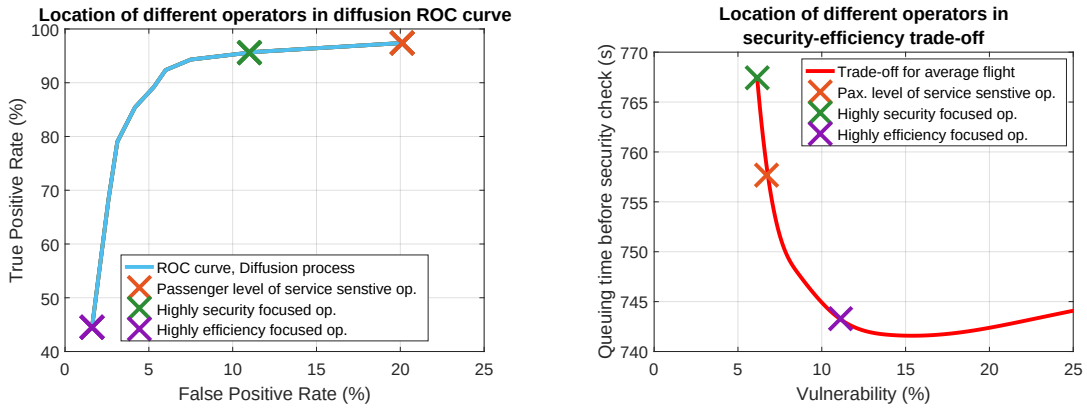
The next trade-off identified in the empirical research is the trade-off between TPR and waiting time. The result of the empirical research is that average operators are willing to trade 1 percentage point of TPR against 2.76 minutes of waiting time (166 seconds). The same method has been applied: in the trade-off curve of queuing time vs. vulnerability, the point is found in which this trade-off is made. The result can be found in Figure 6.19b.

When the average operator is plotted in the queuing time vs. vulnerability curve, it turns out that the average operator tries to find the lowest vulnerability possible. TPR is valued very high by the average operator, so much that he is willing to give up a large amount of efficiency performance. This corresponds to a security focus of 100%.

This result is not in line with how the operators responded to the TPR - FPR trade-off. In the trade off between vulnerability and false positive rate, operators showed to have a security focus of 85%, while in the result in Figure 6.19b, the same average operator has a 100% security focus. These results contradict. Apparently, operators on average have the aim to be 100% security focused when this is compared to waiting time. But at the same time, they are willing to trade TPR with FPR at a lower rate. The real thing that operators influence, is the value of TPR and FPR; queuing time is only a result of this. Since the operators' aim concerning waiting time is different from their aim concerning TPR and FPR, they will not obtain both desired results. If they actually choose their position in the TPR - FPR trade-off the way that became apparent from the research (depicted in Figure 6.19a), the result in terms of waiting time becomes different compared to their aim for the waiting time trade-off (depicted in Figure 6.19b). This is an insight that should be made clear to operators: if they want to reach their vulnerability goal that they set to themselves, they should choose their position in the TPR - FPR trade-off more to the right top of the ROC-curve in Figure 6.19a. This means that, when in doubt, they should send an item for an extra check more frequently than they currently do.

6.5.2. Position of three types of security operators in trade-offs

In the same empirical research, a latent class model was used to distinguish between different types of operators. Three different operators were identified at RTHA: Passenger level of service sensitive operators, highly security focused operators and highly efficiency focused operators. The place of these different operators on the ROC curve in the TPR-FPR trade-off is shown in Figure 6.20a.



(a) Diffusion ROC curve with position of three different RTHA security operator types

(b) Security-efficiency trade-off with position of three different RTHA security operator types

Figure 6.20: Positions of three different RTHA security operator types in trade-offs

From Figure 6.20a it is obvious that three different operator types all have a different location in the ROC curve. Firstly it is seen that highly efficiency focused operators, truly are highly focused on efficiency. The trade-off they make between TPR and FPR (1 TPR percentage point for 0.018 FPR percentage point) places them in the lowest regions of true positive rate.[3] This has as a consequence that they will have a small number of false positives. They will thus be very effective in terms of processing times.

The most remarkable point in Figure 6.20a is the point of passenger level of service sensitive operators. The result of the empirical research is that both FPR and TPR are valued positively by these operators. These

[3] In this region 1 pp TPR is traded for 0.04 pp FPR

operators thus want to increase FPR and TPR as much as possible (within the specified ranges of the empirical researches). This places the operators clearly at the most top-right point of the graph. This remarkable response of these operators should however be further investigated. Probably they value FPR positively when compared with other parameters than only TPR. A new empirical research would thus be necessary, to define this point in the TPR - FPR trade-off better; no operator really wants to have false positives.

The final point is that of the highly security focused operators. These are willing to trade 1 percentage point of TPR to 4.10 percentage point of FPR. This places them at a TPR of 96% and a FPR of 11 %, slightly higher than the average operator. This corresponds to a security focus of 87%.

These three different operators could also be placed in the trade-of between queuing time and vulnerability. This is done in Figure 6.20b.

For placement of the highly efficiency focused operators in Figure 6.20b, an assumption must be declared. As was found in case study 1, there are two regions in the security-efficiency trade-off in which security performance is traded against efficiency performance (region (i) and (iii)). Based on the trade-off made by the highly efficiency operators, the highly efficiency focused operators could be positioned on two locations, namely at a vulnerability of 11.12% or at a vulnerability of 60%. It is assumed that when an operator performs the trade-off in the way the efficiency focused operator does, he will choose the lower vulnerability. This is substantiated by the fact that operators have to do tests every now and then, and operators performing at a vulnerability of 60% would be picked out for performing too badly.

The highly efficiency focused operators are willing to trade 1 TPR percentage point for only 1.21 seconds. This positions them on a point of a high efficiency performance with relatively short queuing times and a low security performance with a vulnerability of 11.12%. According to expectation, this operator type is located most bottom right of the three operator types.

An interesting insight is obtained comparing Figure 6.20b to Figure 6.20a. It is viewed that highly security focused operators turn out to be the most security focused in Figure 6.20b, while in the passenger level of service sensitive operators were the most security focused in Figure 6.20a. This again points to the mismatch between what operators aim in terms of waiting times, or what operators actually do when trading TPR against FPR. To further illustrate this difference, Table 6.7 has been set-up. In this table the focus on security of the different operators in the different trade-offs is given. It can be viewed that for all operators it is true that there is a mismatch between their TPR-FPR trade-off focus, and their TPR-waiting time trade-off focus.

Table 6.7: Security focus for TPR-FPR and TPR-waiting time trade-offs

Operator type	TPR-FPR	TPR-waiting time
Average operator	85%	100%
Passenger level of service sensitive operator	100%	95%
Highly security focused operator	86%	100%
Highly efficiency focused operator	0%	81%

Why passenger level of service sensitive operators are high up in the ROC-curve and not in the security-efficiency trade-off is can be understood as follows. The most important parameter for passenger level of service sensitive operators is the number of missed flights. Everything else is not too important: this results in a low value for FPR, and also in a lower value for waiting times: these operators only want to trade 1 percentage point TPR for 13 seconds. This places them at a vulnerability of 6.71%: better performing in terms of security than efficiency focused operators, but worse than security focused operators.

The final observation is that highly security focused operators are at the highest point of the security-efficiency trade-off. These operators are only willing to trade 1 percentage point of TPR for 9.93 minutes of waiting time (596 seconds). This makes them focused on security as much as possible: they are committed to do almost anything to make the vulnerability as large as possible. As 59% of the total operators consists of these type of operators, these are also the main cause for the average operator to be in the highest point of the graph in Figure 6.19b. The highly security focused operators are thus the main cause of the overall high security focus of the complete security system.

7

Discussion, implications and recommendations

In this chapter the result of the thesis research will be examined. After the research questions from Chapter 3 approached by creating a model in Chapter 4 and 5, Chapter 6 presented case studies in which experiments were performed with the model. The results in Chapter 6 are treated case study per case study. In this chapter, a synthesis of the results will be presented. What conclusions can be drawn based on the different case studies? What are similarities between different results and what are differences? How can the results be compared with theory from literature? What are the implications for related researches or how can airports use the results? And what are recommendations for further studies? These questions are answered in this chapter.

This chapter is divided into three parts. The first part is the discussion in Section 7.1, in which the results of the case studies are compared and reviewed and linked with trade-off theories. The next section is Section 7.2, in which the implications for both related research and practice are treated. The final section of this chapter is Section 7.3, in which recommendations for further research are given.

An important input for this chapter has been a meeting at RTHA on January 19th 2018, where the results of this thesis research were presented to a public for the first time. The response of RTHA was positive: after the presentation there was a long discussion about the results, also treating possible implications for practice and recommendations for further research.

7.1. Discussion of results

In this section the results presented in Chapter 6 will be discussed. This is done by looking at the overall picture of the results, rather than analyzing what occurred in one specific case study, as was done in the previous chapter. This section is subdivided into three subsections. In the first two sections an answer will be given to research subquestion 3: how can trade-off theories be used to analyze and identify trade-offs between security and efficiency? First in Section 7.1.1 the identified acute-chronic goal responsibility trade-off will be treated and compared to literature that was reviewed in Chapter 2. The same is done in Section 7.1.2, but then for the speed-accuracy trade-off. The final section is about assumptions. In almost every research assumptions need to be made in order to be able to perform the research within a certain scope. It is however important that these assumptions are well-supported and the influence of these assumptions on the result of the research should be regarded. That is what is treated in the final section. Section 7.1.3.

7.1.1. Acute-chronic goal responsibility trade-off

In Chapter 2 it was found that the acute-chronic goal responsibility trade-off must be handled by management, which are responsible for placing the focus on chronic goals. The concerned chronic goal in airport terminal operations is the security of all humans and valuable objects and items. Chronic goals should always be put first, but at the same time this can not be done at all cost. There is not a golden recipe for the trade-off between "faster-better-cheaper" and chronic goals, as the situation in which these trade-offs take place can differ very much in nature. It is the responsibility of managers to make proper choices in this trade-off, which is done by making fact-based judgments in an iterative process.

For airport managers there is thus a challenging and important task to make objective decisions within this trade-off. Managers make these decisions effective by instructing their personnel. It is therefore of high importance that managers know how their personnel make trade-offs under different conditions and what the consequences of the choices of their personnel are. This is what is addressed in this thesis research. It is shown how the acute-chronic goal responsibility trade-off is made within the security system of the airport, and it is investigated what the influence of different circumstances is on the trade-off behavior.

For every investigated circumstance for the acute-chronic goal responsibility trade-off (case study 1, 3 and 4) it was found that there are three regions related to security focus in which operators can perform. Of these three regions, operators will most probably always attempt to be in the second or third region, as the security performance in terms of vulnerability is too bad when operators perform in the first region (with security focus < 30%). It is interesting to see that in the second region, for a security focus between 30 and 75%, the security performance of the system can be increased at very low cost. Vulnerability can be decreased in this region while maintaining a constant (or even increasing) efficiency performance. It was investigated what causes this behavior of the system in the second region. When one looks at a probable mathematical cause the small increase in FPR together with an increasing hit rate and response time is an explanation. But the formation of these regions are also emergent behavior, as a consequence of non-linear interactions within the agent-based model. Further research to this interesting behavior of the system is recommended. The described behavior in the medium security focused region is the reason that it is most beneficial for airport security systems to have a high focus on security (> 75%). This is in line with the literature on the acute-chronic goal responsibility trade-off, which prescribes that chronic goals should always be put first.

Within this third region of high security focus, empirical research showed that when operators are asked about their trade-off between waiting time and TPR, they indicate that they focus almost completely on increasing TPR. But when further investigated how they make their trade-off between TPR and FPR, it is seen that on average they do not focus completely on TPR at all cost of FPR. The explanation for this is the following. The security operators are probably aware of the fact that their manager demands them to have a very high focus on security, despite of the consequences in efficiency (waiting time). But these operators probably do not realize that the trade-off they make between TPR and FPR is exactly what influences this performance. Security operators could thus be trained better in realization of the consequence of their decisions on overall security performance.

Although the average operator shows to aim for a very high security performance, there is a category of the operators who focus more on efficiency. Some operators are more sensitive to increased waiting times then others: next to highly security focused operators, there are also highly efficiency focused and passenger level of service sensitive operators. The implication of the different style of making decisions of these operators is made clear in the case study evaluation and provides very useful information to airport terminal managers.

The above described trade-off is a trade-off that is valid for particular chosen fixed parameters. To explore further, specific properties of the system are changed. The agent-based modelling paradigm allows for implementing dynamic behavior of agents. This is done by introducing a dynamic fatigue parameter, which reflects the performance change caused by tiredness of the operator and emerges to influence the overall performance of the system. Tiredness was shown to have a higher impact in terms of security and efficiency when the focus on security is high. But also when tired operators were introduced to the system, the behavior of the acute-chronic goal responsibility remained equal: the same three discriminating regions could be identified, only the gradient within the identified three regions varied for the tired operators situation.

Operators are not the only agents in the model influencing the performance of the security system. As was introduced by Kirschenbaum, passengers can not be seen as passive elements that only just need to be proceeded through the security check. By modelling agents representing diverse passengers with a different probability of forbidden items, it is investigated what the influence is of different type of passengers on the performance of the system. Although this adaption to the model only influenced the efficiency performance, again the same three typical regions were identified.

As a conclusion, the focus on security of operators is what predominantly determines the acute-chronic goal responsibility trade-off. One could ask: is there really something like a difference in security focus? Are not all operators performing at their best in terms of security, since this is what the manager asks them and why they do their job? In discussions with team leaders of security operators at RTHA it turned out that indeed there is something as a difference in security focus. An example is a "black hole" that X-ray operators can view on images, caused by an undetermined, possibly forbidden object. Regulation prescribes that such a black hole should always be investigated. The situation occurs that if this black hole is small enough, some operators choose to not send this "black hole" luggage through for an extra check, but rather let it proceed. Fur-

thermore, research supports that "10% of security personnel exceeds or bends rules when the situation calls for it", and "12% of security personnel states that breaking (security) protocol is sometimes necessary to get the job done"[21]. This shows the importance of the incorporation of a security focus. In this research, bias z and thres_{threat} are chosen as discriminating parameters in the acute-chronic goal responsibility. These parameters leave room for many other possible implementations, such as a causal relation between the length of the queue and the bias of the operators. But the exact causal relations of circumstances on thres_{threat} and bias z is something that can be investigated in further research; in this research the importance of these parameters is proven by showing how varying security focus is a way of making the acute-cronic goal decisions of operators explicit.

7.1.2. Speed-accuracy trade-off

In the literature review in Chapter 2 the speed-accuracy trade-off was introduced. It was stated that this is a trade-off that is mapped in literature relatively well when compared to the trade-off discussed in the previous section. It is generally assumed that within many tasks that need to be performed within a certain amount of time, increasing speed in the performed tasks results in decreasing accuracy and vice-versa. It was found that many aspects influence the performance of a person within the speed-accuracy trade-off, like basic quality to perform the task or arousal rate. But independent of these properties and conditions, the best position in the speed-accuracy trade-off is to perform the task as fast as possible, without sacrificing accuracy.

This last statement is in line with the findings of case study 2, which investigates the speed-accuracy trade-off. The result from case study 2 is that it is most beneficial for the performance of the security system in terms of both security and efficiency to have a high focus on accuracy. The reason for this is that, although response times might take longer for accurate decisions, this effect will be compensated for by the decrease in the number of false alarms that cause an extra check. As long as the average extra response time does not exceed the average extra time induced by the extra checks, a focus on accuracy is beneficial. This underwrites the statement in literature, that "the best position is to perform the task (..) without sacrificing accuracy".

Another interesting insight from case study 2 was the effect of increasing speed in terms of efficiency performance. When the focus on accuracy was decreased from 40 to 20% (read: the focus on efficiency was increased from 60 to 80 %), actually the average processing speed of the system became lower. Only if operators decrease the focus on accuracy from 20 to 0 %, a real benefit in terms of processing speed is obtained. However, the security performance in this region is so bad, that such a low focus on accuracy should not be considered. This result provides the following insight: operators might perceive that if they really focus on speed, that indeed the system will perform better in terms of efficiency. The cost in terms of security performance is however too high to consider this amount of focus on speed. For other regions, with acceptable levels of security performance, a lower focus on accuracy will only lead to a worse efficiency performance. Therefore it is concluded that for operators a focus on accuracy is most beneficial, up to a level that the total extra induced response times cost less time than the total number of extra false checks. What exactly is this boundary, is something that could be retrieved with simulations in further research.

7.1.3. Implications of assumptions on research

To be able to perform a research within a certain scope, assumptions need to be made about uncertainties that are beyond the scope of the research. It is however important to consider the effect of this assumptions. The most important assumptions are explicated and analyzed in this section.

First of all, the results discussed in the previous two subsections are based on local sensitivity analysis: analysis in which one parameter is changed and the effects of the parameters are checked. It may however be true that it is not possible to only vary one parameter at the time, because of interactions between different parameters within the system. It may be true that if one parameter is varied, other parameters should be varied at the same time. In this research it is assumed that it is possible to vary the security focus and the accuracy focus of the security system without modifying other parameters. To investigate if this is possible, global sensitivity analysis can be applied to find what the relative influence is of a certain parameter on the global properties of a system. This was done in the preliminary research. Global sensitivity analysis however only shows the importance of a parameter, and does not necessarily quantify the influence of a parameter on the performance of a system. For determining the influence of a parameter on the performance of the system, local sensitivity analysis is applied. To generalize the results as much as possible, the local sensitivity analysis of varying bias z and thres_{threat} has been performed for different circumstances. But it is impossible to model all possible circumstances in one research. The obtained results should thus be interpreted as a representation of the behavior of the model, relatively for different situations. No direct conclusions can be

drawn about absolute numbers that were obtained during the research. It is therefore that the emphasis was put on describing the relative behavior of the system in certain regions, rather than coming up with absolute numbers indicating the performance of the security system.

For the placement of the operators within the TPR-FPR trade-off and within the security-efficiency trade-off, a comparable assumption is made. The empirical research only provided information about how much operators want to give up efficiency performance for a better security performance and vice versa. This empirical research thus only provided information about the gradient of the slope within either trade-offs, and not about the absolute position. To determine the absolute position of the operators, a very well calibrated model is necessary and further empirical research to this absolute position is required. The value of the applied methodology in this research however, is that it is a first implementation of a combination between the modelling of and the empirical research to security-efficiency trade-offs on airports.

The varying system performance is dependent on multiple parameters as described above, but also on how well these parameters resemble reality. The goal of this research is to investigate the security-efficiency trade-offs in airport terminal, rather than to come up with a perfectly calibrated model. Therefore assumptions have been made in order to come to a calibrated model for signal detection theory and diffusion decision processes. These assumptions are chosen deliberately, because the chosen parameters are of large influence on the performance of the model. It can not be stressed enough that a well calibrated model is very important to be able to draw relevant conclusions. An example is the information obtained in the meeting at RTHA: a teamleader of security operators at RTHA explained that the peak of tiredness is merely at 13:00 around lunch: early in the morning and late in the evening operators are more fit. To account for this kind of real properties of the airport security system, a calibration proposal is written as a recommendation, further in this chapter. The extensions that were implemented in this model (signal detection, diffusion process) allow for such an accurate calibration of the model. When such a very well calibrated model is designed, the proposed trade-off assessment methodology in this research becomes of an even higher value.

The final assumption is the third order polynomial fit that was used to describe the relation between security focus and efficiency performance. The fit of this polynomial regression is not perfect. It does follow the shape of the three different regions in the system, so it takes into account the different behavior within these regions. It could also have been chosen to take the three linear fits to the different regions, as was done in Figure 6.5b. When this had been done however, information about the increasing gradient in region (iii) and the decreasing gradient in region (i) would have been lost. The choice of a polynomial fit is therefore preferred over multiple linear fits. But clearly, the necessity of a polynomial fit is undesired. One can get rid of the necessity of a polynomial fit by specifying more set-ups per case study, and by performing more simulations per set-up. Doing this, the eventual aim is that the behavior of the system can be described exactly by simulation results, rather than by a fit on these results. But as a simulation of one flight day takes seven minutes of computation time, the total number of simulations for this research was limited, leading to the spread data that required a fit.

7.2. Implications of results

In this section the implications of the result of this thesis research are set out. Two different implications of the research are distinguished. First the implications of the research for practical purposes on airports is treated in Section 7.2.1. This section also serves as an answer to research subquestion 4, about the development of a tool. This section is followed by the implications of the results for further and related research in Section 7.2.2.

7.2.1. Practical implications of results

The practical implications of the results are the implications that this thesis research can have on real airports. As earlier discussed, the results provide an insight about how airport terminal security operators make decisions regarding the security-efficiency trade-off. As was described in the discussion on the acute-chronic goal responsibility trade-off, the responsibility for the security is in the hands of the airport managers. When airport managers can obtain a better insight in how their operators make trade-offs and where the operators are located within the trade-offs, this helps the managers in making better decisions.

But for (busy) airport managers handing over a report containing a research methodology and static results in graphs is not useful. Airport managers are more pragmatic: they want something easy to understand in which the results are made explicit and insightful. For this reason, as an out-of-the-box contribution to this research, a tool is designed that can be used by airport terminal managers to get insight in the security-

efficiency trade-offs. This tool can be found by scanning the following QR-code, or by surfing to the link below on a computer or a mobile device:

sqn.nl/tradeoff

In the trade-off tool all results of this thesis research are implemented. In the tool a manager, security operator or other user can experience the consequences of security decisions. All case studies are in the tool, meaning that the user can choose an average, charter or business flight day and whether operators are tired or not. By adjusting the focus on security between 0 and 100%, the security performance (vulnerability) and efficiency performance (processing time) are presented to the user. In this way the user can "play" with the trade-off, without having to perform simulations with the AATOM simulator. This is a quick method in which the results can be made useful for airport managers, for example at RTHA. The insights can be used for example to anticipate on different performance periods on the day, for example by using more resources when the average performance of operators is the lowest. Another example implication is the possibility to investigate the effect of an extra break for security operators. The tool can be opened on computer and on mobile devices, so that it can be used anywhere.

During the meeting at RTHA this tool was presented and the response was very positive. RTHA supports the research by providing TU Delft with information and date. They do this because the research is important, but they value it even more if they really can use the results, which is what they can do using this tool. Next to this, for obtaining the data, often time is asked from the security operators. If these security operators always just need to cooperate and never see what happens with the tests they make, they will be less willing to participate in later experiments. This is also a reason why RTHA valued the tool: at first hand they were thinking about presenting a flyer with the experiment results to the participating operators, but a tool like this is probably a more appealing way of presenting the results. A comment that was made during the meeting was that, to really use this tool, it is important to have a well-calibrated model. Furthermore, during this meeting also some feedback was obtained about the interface of the app, which was implemented in the tool afterwards.

7.2.2. Implication of results for related research

The results of this research will not just be used for practical implementations. This research was part of a cluster at TU Delft in which under the supervision of Alexei Sharpanskykh and PhD candidate Stef Janssen, six MSc students worked on different topics within the subject of agent-based modelling in airport terminal operations. This cluster will proceed with this subject after the completion of this thesis research. The results of this thesis research are thus of major implication for further researches within this research group.

The major contribution of the research is in the introduction and calibration of the diffusion model. The diffusion model is a very promising, elaborate model in which many parameters can be adapted to model for different (dynamic) behavior and autonomous decision making of security operators. If more experiments are performed with security operators, the diffusion model can be calibrated more realistically, which will only contribute to the way the AATOM model resembles reality.

In this research the use of the diffusion model in AATOM has been explored thoroughly. It is investigated how the diffusion model can be used to model dynamic behavior of the operators, by introducing dynamic

parameters regarding fatigue of operators to the model. But the purpose of introducing such dynamic parameters to the model was not only to show what how fatigue affects operator performance. The investigation to tired operators was just an example on how (dynamic) circumstances can influence the performance of operators and thereby of the entire system. The implemented diffusion model allows for many more possible influences on operator performance. In the meeting with RTHA many other example influences on operator performance were named: morning persons vs. evening persons, temperature in the security check or other environmental circumstances, different shift lengths, etc. Provided that the influence of these properties are modelled well by extensive calibration, the proposed research methodology can make the consequences in terms of performance of these different dynamic circumstances and properties explicit. And when security operators are modelled up to a detailed level, this will help in performing security risk assessment and airport efficiency analysis better.

But not only the operator performance is of influence on the trade-offs and the performance of the system within this trade-off. As was introduced by Kirschenbaum, passengers can not be seen as passive elements that need to be proceeded through the security check. By modelling diverse passengers with a different probability of forbidden items, it is investigated what the influence is of different type of passengers on the performance of the system. But modelling different probabilities for forbidden items is also just an example implementation of the possibilities with the model. In the meeting with RTHA, many other different passenger properties that possibly influence trade-offs and the performance of the system were mentioned. Examples provided are stress of passengers which arrive just before their flight, or the fact that one should not differentiate between business and charter flights, but rather between destinations.

This research provided a first example of modelling diverse passengers. Any type of passenger can be modelled, as long as validated models and data are used to do so properly. An very interesting interaction can be found when the behaviour of these passengers also has implications on the performance of operators. An example could be that operators become more tired from stressed passengers than from normal passengers, which influences the operator's performance. If such interactions are taken into account, the possibilities provided by the agent-based model paradigm are exploited well.

Another implication is the contribution of this research to the PhD research of co-supervisor of the cluster, Stef Janssen. The subject of his PhD research is to analyze trade-offs within airport terminal operations. The results of this thesis research can be used as inspiration for his PhD research.

The final implications are for empirical researches that will help further develop AATOM. It was found when applying the results from a survey in the evaluation of the case studies, that there are inconsistencies how people approach different trade-offs. Knowing the results from the case studies and the evaluation of the case study, a new survey could be set out that takes into account these inconsistencies. For example, operators should be made aware how the choice in one trade-off affects the result in the other trade-off. Another option is to investigate revealed preference instead of stated preference: rather observe what is happening, instead of asking the operators what is happening.

Another implication for empirical research is the suggestion for new empirical researches. These researches should be configured such that the DMAT calibration algorithm can be applied. Doing this, a large variety of experiments could be performed with operators under different circumstances. Using the results of these experiments, the respective circumstances can be modelled within AATOM, resulting in a more detailed model.

But next to this specific implication to related research, this research is one of the first investigations that explicitly models trade-offs within an agent-based model. Performance areas like security, safety, efficiency and resilience can not be investigated apart from each other. When assessing the performance on one area, other areas should be taken into account as well. Agent-based modelling has proven to be a very suitable paradigm to model autonomous, diverse operators in a dynamic environment and to assess how local properties of these agents can emerge to global behavior of a system. The combination of trade-off analysis and agent-based modelling is an interesting field of research, and this research sets first steps in this area.

7.3. Recommendations for further research

In this section recommendations are given for further research. The recommendations are grouped within different subjects. Within these groups, a short explanation is given and the recommendations are summarized in bullet points for clear overview.

Calibration of the model

Useful results can only be obtained with a proper calibrated model. An important part of this research was the calibration of the model in order to be able to draw conclusions about the security-efficiency trade-off. But this was not the main focus of the research. It is therefore recommended in further research to put effort in the calibration of the model, especially for the diffusion process model. Using DMAT experimental results can be given as an input to an algorithm, which computes the drift rate parameters. The only required input from experimental results is the response of the operator, the response time of the operator, and the type of operator. By setting out proper experiments that evaluate the TPR and FPR performance of operators in various circumstances, the operators can be modeled more detailed in AATOM, leading to more useful results. But not only the performance of the operators can be calibrated better. As was identified earlier in this chapter, there are a many different passenger properties that influence the behavior of the system. Measurements on the airport can be performed to assess the security and efficiency performance in different passenger circumstances. For example, the difference in processing times between summer and winter could be examined. The more passenger properties are taken into account, the more realistic the model is, and the more useful the results will be.

- Perform experiments to investigate operator performance and use result for calibrating diffusion decision model;
- Measure performance at airports for different types of distinguished passengers.

Improvement of diffusion decision process

Not only the calibration of the diffusion model can be performed in a better way. Currently the only input that triggers the rate of collection of evidence of the operator (drift rate v), is the fact if an item is forbidden or not. But in reality not every forbidden item induces the same stimulus to the security operator. Some items are more easily detectable than others. Easily detectable items will have a higher TPR and lower response times. This is currently not implemented in the model. Instead of only conditioning v on a forbidden item or not, the drift rate can be made dependent on the threat level of a certain item. When the threat level distribution is than calibrated realistically, the results of the decision process will also be more comparable to real decisions made by security operators. Also, currently the diffusion process parameters are the same for all operators. This means that they collect evidence in the same way and make the decision after the same amount of accumulated evidence. This is not necessarily a wrong assumption based on the fact that all operators on RTHA need to be able to perform all different tasks during a work shift. But on the other hand, X-ray checking, luggage checking and physical checking are three different jobs, in which evidence is probably collected in different ways. Therefore, calibration of the diffusion decision process for the three different jobs may be in place. Another improvement to the diffusion decision process is the correct specification of maximum accuracy focus of the operators. As a very high focus on accuracy is beneficial for security operators, it is interesting to determine exactly where this boundary lies, and how an operator can get as close to this boundary as possible.

- Make drift rate in diffusion decision process dependent on threat level (not only forbidden item);
- Calibrate diffusion decision parameters for the three different operator jobs (X-ray, luggage and physical check);
- Find exact and validated boundaries of speed-accuracy trade-off.

Improvement of AATOM simulator

As was explained both in the implementation of the preliminary model and the final model, the current interactions in the security check do not (yet) resemble reality completely. When distributions from experimental data about values for the processing times are implemented in the model, the resulting throughput rate of the security check is much lower than the throughput rate in reality. The proposed solution is to implement a correction factor CF with which the calibrated processing times are reduced in order to obtain a realistic throughput rate for the security. But this is not an ideal situation. The recommendation is therefore to find new solutions to ensure that the throughput rate of the security check in the simulator matches the real throughput rate of security checks. Current example shortcomings are that passengers can not move their luggage to another table during luggage collection in order to make space for a waiting passenger, or the fact that only three people can drop and collect their luggage at the same time. In short, the recommendation is to find new implementation solutions to get closer to full validation of the model.

Further research within built model

Within the expanded model that is proposed in Chapter 5, four different case studies have been performed. But these case studies are clearly not the only case studies that can be performed within the current existing expanded version of AATOM. Other example case studies include:

- Perform case studies with more different circumstances, such as fatigue sensitive vs. non-fatigue sensitive operators, high-skilled vs. low-skilled operators, winter vs. summer days (which influences the use of the number of boxes);
- Perform sensitivity analysis in which two parameters are varied at the same time;
- Investigate the model behavior during different phases of passenger accumulation in security check (accumulation phase, equilibrium and saturation phase);
- Change of speed-accuracy focus during the day (e.g. due to fatigue);
- Define a causal relation between the queue size and the security focus of operators;[21]
- Perform simulations in which different operators are working at the same time (e.g. different operators identified in evaluation of case studies);
- Investigate the effect of an extra break for operators;
- Preform more simulations for current set-ups to obtain results with less noise;
- Define more set-ups within case studies to better model the behavior in the region between the current set-ups.

Inspiration for further model extension

Although the current model already allows for much more further research, the model is far from finished. Currently only a part of the security risk assessment method that was described in Chapter 2 is implemented, namely the vulnerability of the system. This is not the only security measure there is. One could also look at consequence of the threat scenario, for example by linking the size of the forbidden item that passes through the security check to a consequence number. Furthermore, in this research only one threat scenario is analyzed. In the literature review report "Analysing Security and Efficiency of Airport Terminal Operations: Literature review" many more threat scenarios are identified[36]. Implementing more threat scenarios in order to obtain a complete picture of the total risk imposed to an airport is one of the ultimate goals of AATOM. Other possible extensions of the model are to implement more different types of passengers, for example passengers who travel in groups, or make the processing time distribution of the passengers conditional to the destination of their flights. Another recommendation is to establish an interaction between the diversity of passengers and the performance of operators. For example, it is probable that physical and luggage operators become more tired from a day in which they need to perform an extra check on many passengers an items (e.g. charter flight day), then on a day with only few positives (e.g. business flight day). Currently the behavior of passengers is not of influence on the security operators, while there might be an interesting link here. A final recommendation is to investigate other parameters that are of influence on the security-efficiency trade-off. Currently some determining parameters are distinguished (thres_{threat} and bias z) but there might be more. Global sensitivity analysis can be performed to find more influencing parameters, which then can be further investigated.

- Perform other steps of the security risk assessment framework, such as consequence assessment;
- Define and investigate other threat scenarios;
- Implement more different types of passengers to better resemble the diversity of passengers in real airports;
- Investigate the influence that different passenger types have on the performance of operators;
- Perform global sensitivity analysis to find more parameters that influence the security-efficiency trade-off within airport terminal operations.

8

Conclusion

This research sets the first steps to modelling dynamic relations and trade-offs between the airport terminal operations performance areas of security and efficiency. This modelling is done using an agent-based modelling approach. As a modelling and simulation environment, the model AATOM and the appurtenant AATOM simulator are used. Because the field of trade-off modelling within agent-based models is relatively unexplored, firstly a mapping of the possibilities of trade-off modelling within AATOM is made, by performing a preliminary exploration. The results of this preliminary exploration are used to build an expanded version of AATOM, that can be used to analyze trade-offs between security and efficiency within airport terminal operations.

In the literature study preceding to this research, a trade-off called the *acute-chronic goal responsibility trade-off* was found to be described extensively. In short this trade-off concerns the focus on chronic goals within a system, against "faster-better-cheaper" goals. The chronic goal within the security check, is the security detection performance. To model for this trade-off, a variable focus on security was defined in the AATOM model. Inspiration for this is the ROC-curve of security scanning equipment. By varying a certain threshold, the focus can be shifted from security to efficiency. Using signal detection theory, security equipment (a WTMD) with a variable focus on security is implemented in the model. To model for the people that perform security checks (X-ray operators, luggage check operators and physical check operators), signal detection theory does not suffice as this leaves out (dynamical) human aspects. Therefore, a diffusion decision process model is implemented in the AATOM model, in which autonomous behavior of humans can be modeled op to a detailed level. Varying the parameter for bias, z, is found to be the discriminating factor in the focus on security of autonomous operators.

Interesting global emergent behavior of the system was found when plotting the efficiency performance (processing time in the security check) against the security performance (vulnerability of the security system). Three different regions were identified: (i) bad security performance, (ii) improvement in both performance areas and (iii) security-efficiency trade-off at acceptable vulnerability levels. Most airports will perform in region (iii), as making trade-offs in this region is most optimal and an acceptable security performance is reached. Where the airports are placed exactly depends on the choices made by the operators and on the types of passengers present on the airport. The second region, improvement of both performance areas is an interesting one to further investigate.

The identified trade-off has been tested for different circumstances. Dynamic operators of which the performance is dependent on their tiredness during the day are implemented using a biomathematical fatigue model, and diverse passengers are modelled. Results show similar behavior as described above, with slight adjustment caused by the different properties of the system. But the most important contribution of this research are not the absolute results of these case studies. The largest contribution is in the fact that a methodology is developed in which operators and passengers with different (dynamic) properties can be modelled within the AATOM model. If, using empirical researches on airports, the (dynamic) properties of operators and passengers are calibrated appropriately, interesting results with useful implications for airports in practice can be obtained.

Another trade-off that was identified in literature is the *speed-accuracy trade-off*. The implemented diffusion decision model allowed for modelling this trade-off by varying the decision boundary a. Result of the experiments was that a focus on accuracy is most beneficial for airport operations both in terms of effi-

ciency and security performance. This is in line with the literature about the speed-accuracy trade-off, which states that optimal performance is to focus on optimal accuracy, while trying to perform the task as quickly as possible. When empirical research is performed to find "optimal accuracy", and to find what the maximum amount of time is that operators can take for their decisions, this can be used again to better calibrate the model and to be able to present more insightful results.

An example of how this research could be combined with empirical researches is given by implementing the results of a parallel performed thesis research by another MSc student. The combination of both researches showed what the consequences are of the trade-offs that different types of operators make. Furthermore, it was identified that there is a mismatch between how operators focus on security when making different types of trade-offs.

As mentioned, the result of this thesis research can help airport terminal managers in better judging the consequence of decisions and circumstances of security operators. The resulting graphs of the case studies can serve as an instrument for security managers to make better decisions regarding security-efficiency trade-offs. To already provide such an instrument, an online tool is designed in which the results of this research are made insightful for airport managers.

Concluding, this research addressed the trade-offs made between security and efficiency within airport terminal operations. These two performance areas can not be analyzed apart from each other, since influencing the one has a large impact on the other. This research provided a methodology on how these two performance areas can be analyzed, when modelled together in an agent-based model. When thorough calibration of the agents within the model is performed, this methodology can be used to provide a very useful insight in the consequences of decisions made within the security-efficiency trade-off in different circumstances .

A

Graphical representation of conceptual model

In this appendix, the graphical visualization of the preliminary conceptual models is presented. The graphical visualization is given in Figure A.2, and Figure A.1 serves for the explanation of specific parts of the visualization.

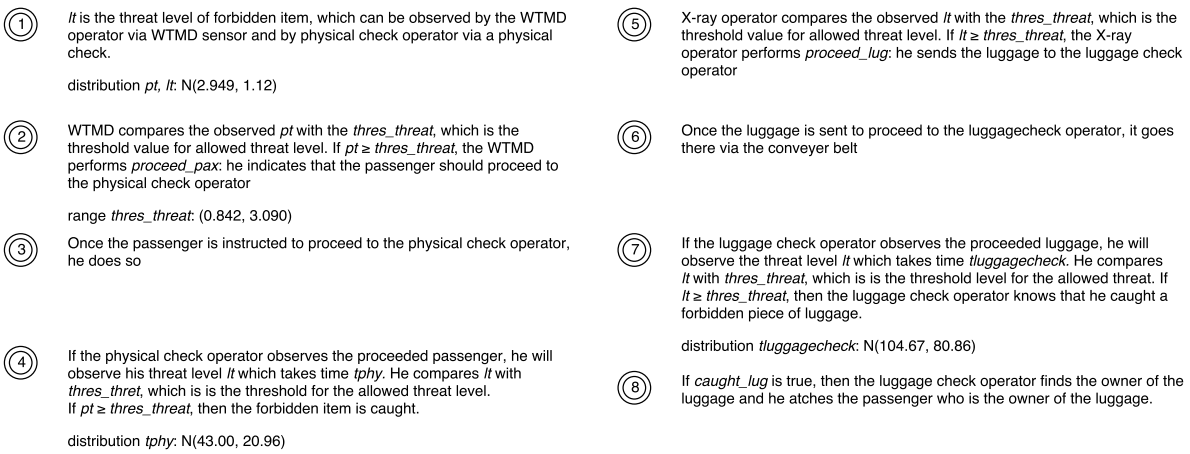


Figure A.1: Conceptual model for implementation of TSI

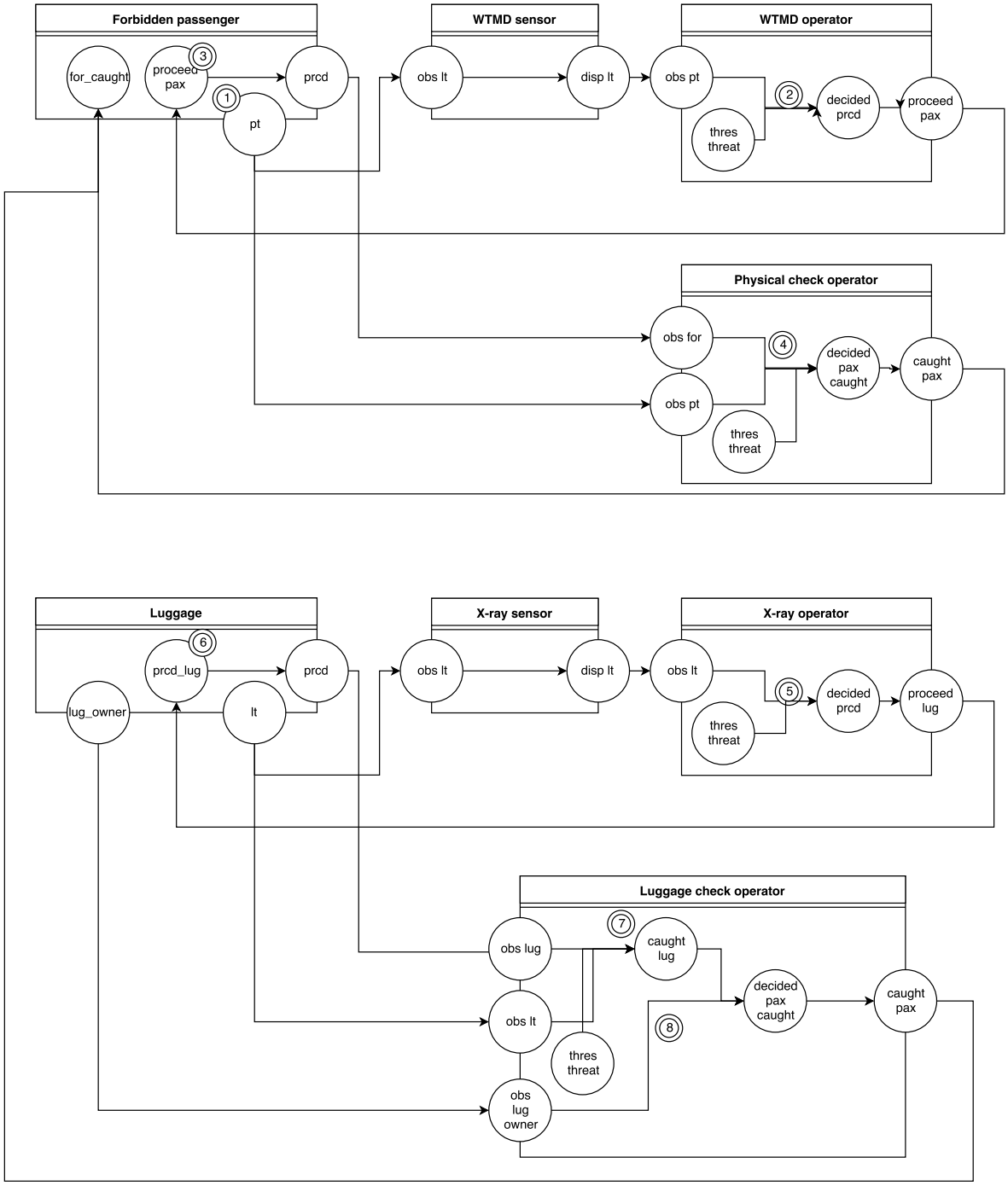


Figure A.2: Conceptual model for implementation of TSI

B

Flight schedule

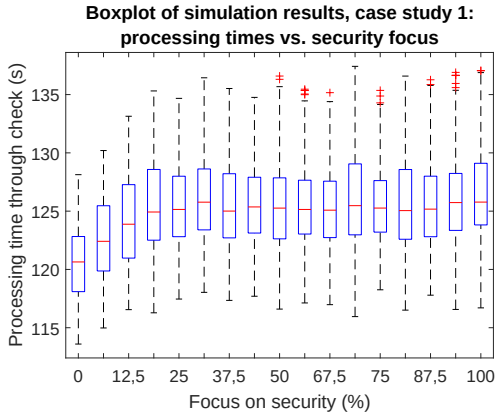
Table B.1: Flight schedule implemented in AATOM

Departure time	Flight	Destination	Type	Capacity	Pax	Passengers arriving during block			
						02:00	01:30	01:00	00:30
				Load Factor:	0.9	10%	40%	40%	10%
06:55	HV6035	Rome - Fiumicino	B737	192	173	17	69	69	17
07:00	HV6301	Gran Canaria	B738	148	133	13	53	53	13
07:00	HV6493	Venetie	B737	192	173	17	69	69	17
07:05	BA4450	London City	E190	98	88	9	35	35	9
07:25	HV6771	Budapest	B737	192	173	17	69	69	17
08:00	HV5021	Malaga	B738	148	133	13	53	53	13
09:55	BA4452	London City	E190	98	88	9	35	35	9

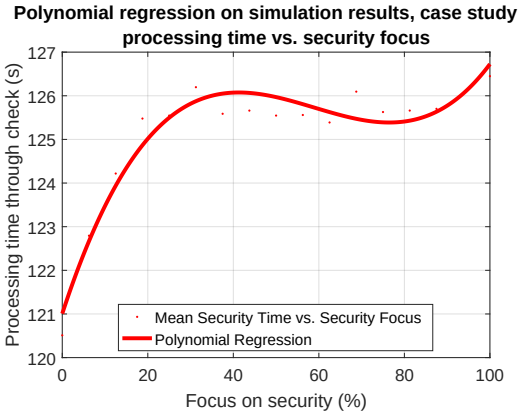
C

Resulting graphs from simulations

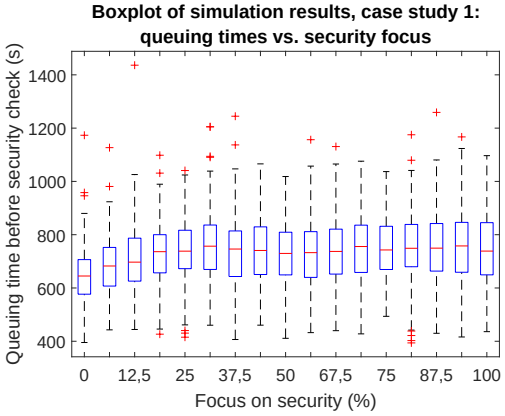
C.1. Efficiency performance simulation results and fits



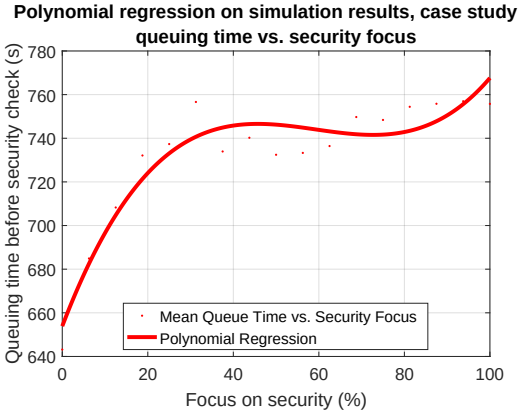
n = 250, 17 set-ups



Adjusted R^2: 0.9257
$RMSE$: 0.3993

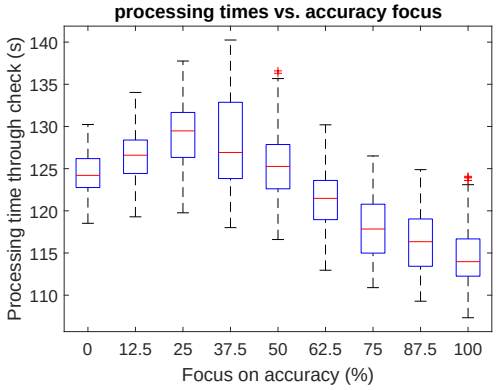


n = 250, 17 set-ups



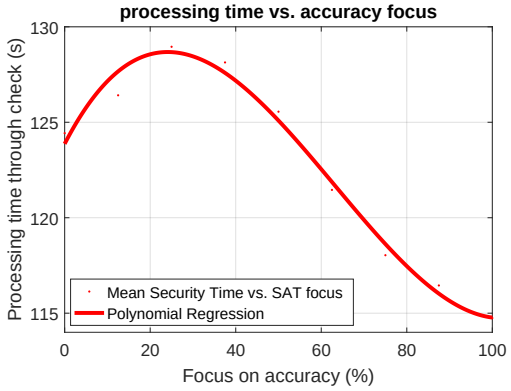
Adjusted R^2: 0.8714
$RMSE$: 10.6930

Boxplot of simulation results, case study 2 (SAT): processing times vs. accuracy focus



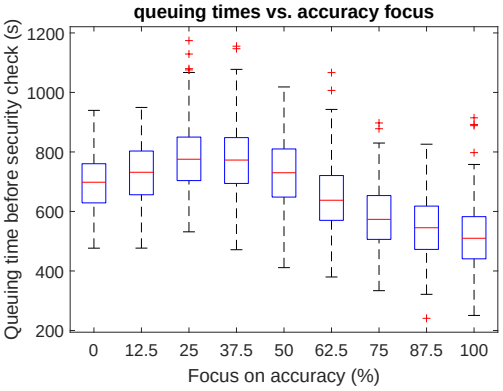
n = 250, 9 set-ups

Polynomial regression on simulation results, case study 2 (SAT): processing time vs. accuracy focus



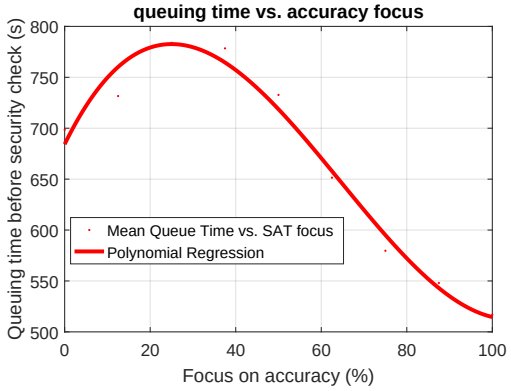
Adjusted R^2: 0.9855
$RMSE$: 0.7963

Boxplot of simulation results, case study 2 (SAT): queuing times vs. accuracy focus



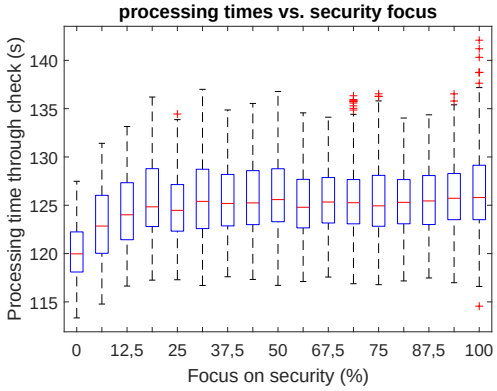
n = 250, 9 set-ups

Polynomial regression on simulation results, case study 2 (SAT): queuing time vs. accuracy focus



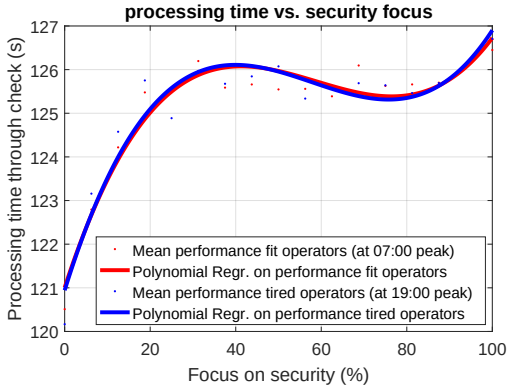
Adjusted R^2: 0.9657
$RMSE$: 18.5347

Boxplot of simulation results, case study 3 (fatigue): processing times vs. security focus

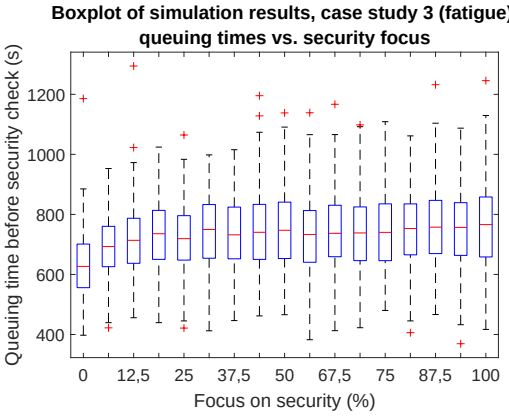


n = 250, 17 set-ups

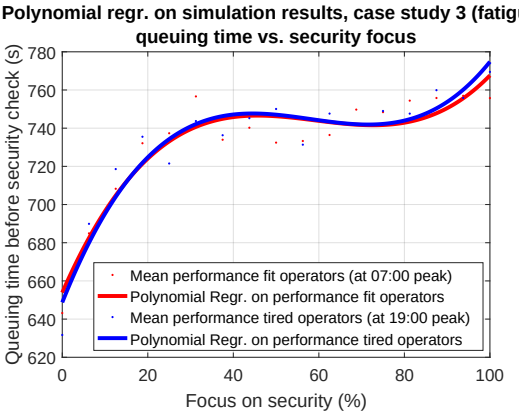
Polynomial regr. on simulation results, case study 2 (fatigue): processing time vs. security focus



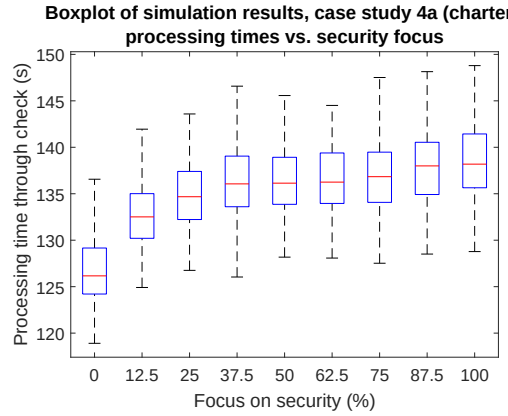
Adjusted R^2: 0.8891
$RMSE$: 0.4782



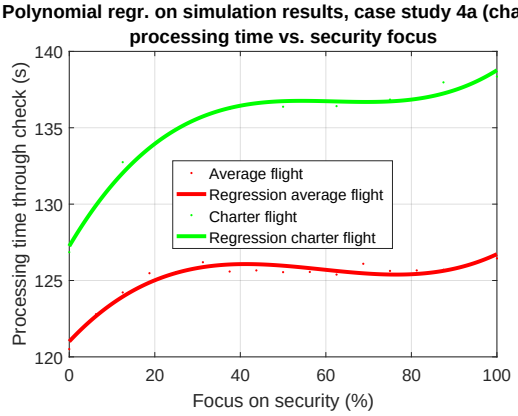
n = 250, 17 set-ups



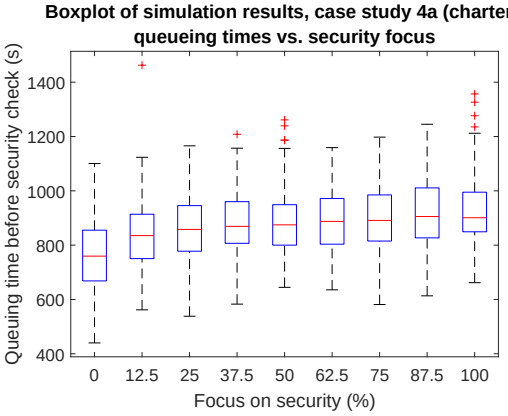
Adjusted R^2: 0.8891
$RMSE$: 10.6506



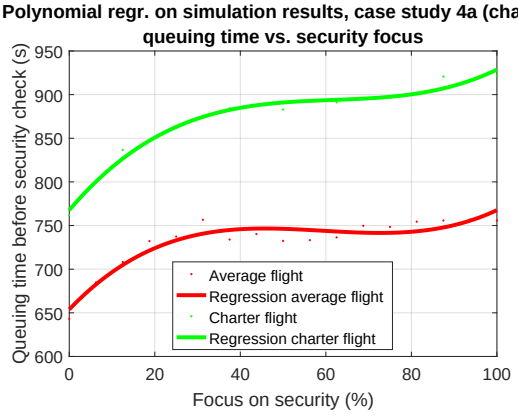
n = 250, 9 set-ups



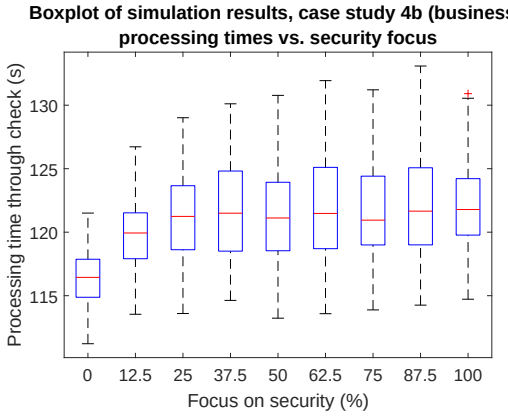
Adjusted R^2: 0.9743
$RMSE$: 0.5665



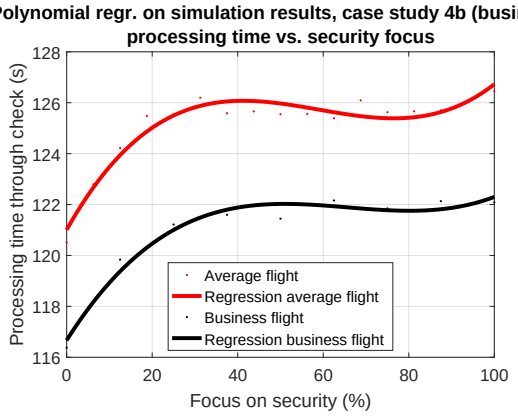
n = 250, 9 set-ups



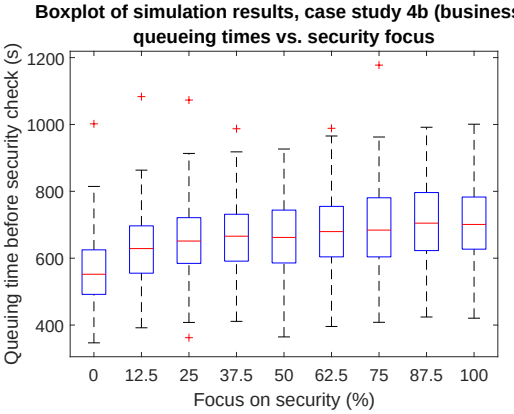
Adjusted R^2: 0.9635
$RMSE$: 9.4386



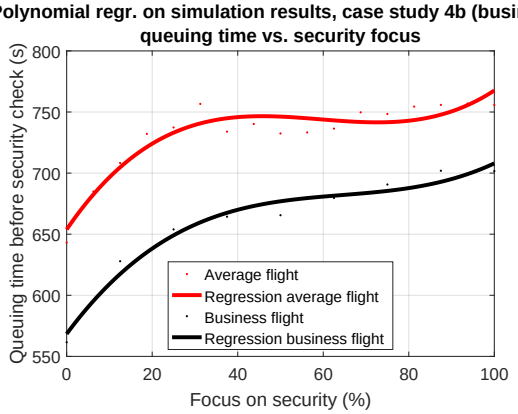
n = 250, 9 set-ups



Adjusted R^2: 0.9469
$RMSE$: 0.4299

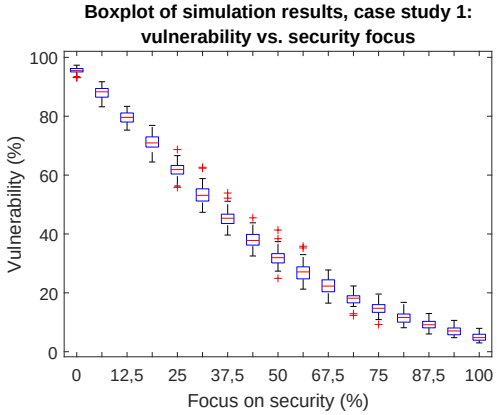


n = 250, 9 set-ups

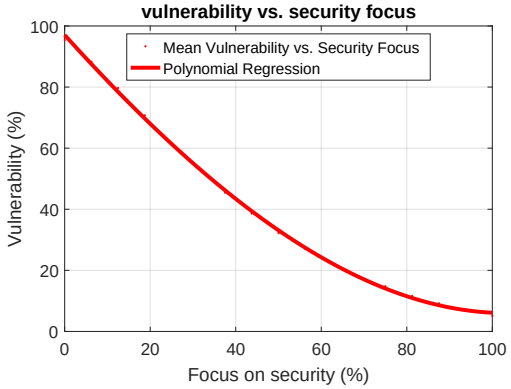


Adjusted R^2: 0.9696
$RMSE$: 9.7379

C.2. Security performance simulation results and fits

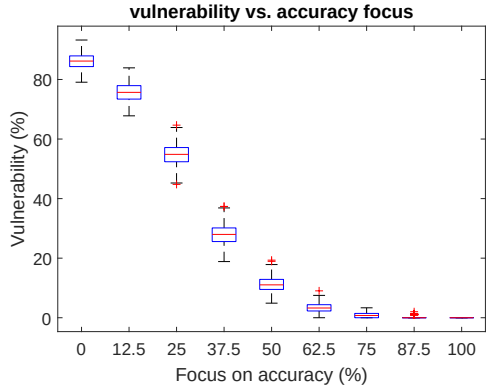


n = 250, 17 set-ups



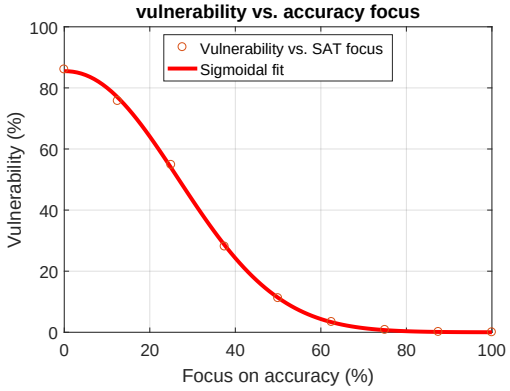
Adjusted R^2: 0.9991
$RMSE$: 0.9121

Boxplot of simulation results, case study 2 (SAT): vulnerability vs. accuracy focus



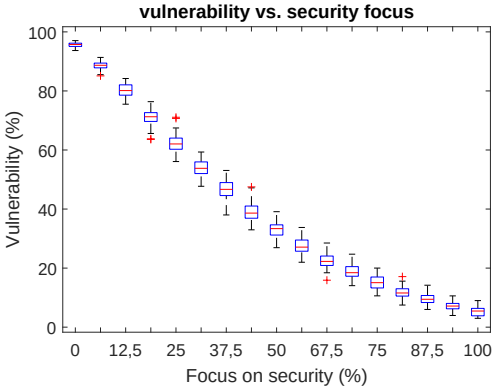
n = 250, 9 set-ups

Sigmoidal regression on simulation results, case study 2 (SAT): vulnerability vs. accuracy focus



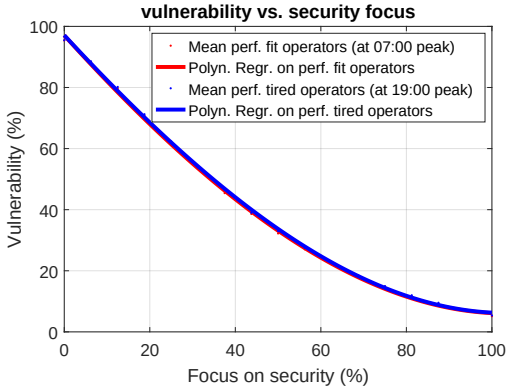
Adjusted R^2: 0.9993
$RMSE$: 0.8621

Boxplot of simulation results, case study 3 (fatigue): vulnerability vs. security focus



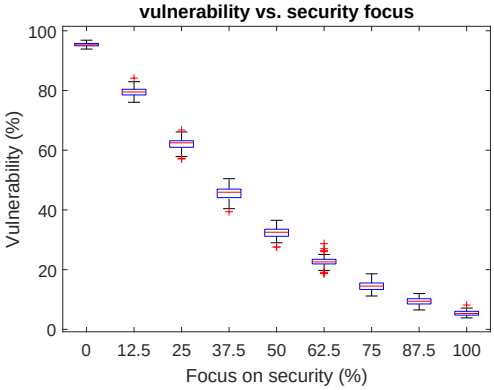
n = 250, 17 set-ups

Polynomial regr. on simulation results, case study 3 (fatigue): vulnerability vs. security focus



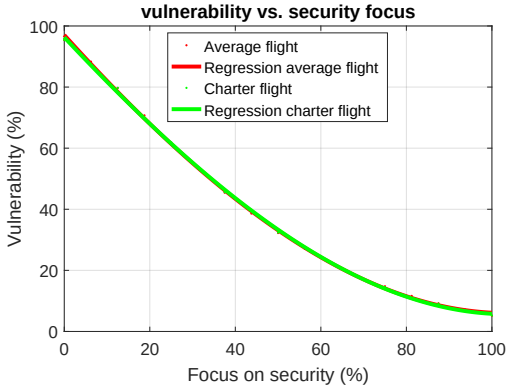
Adjusted R^2: 0.9991
$RMSE$: 0.9018

Boxplot of simulation results, case study 4a (charter): vulnerability vs. security focus

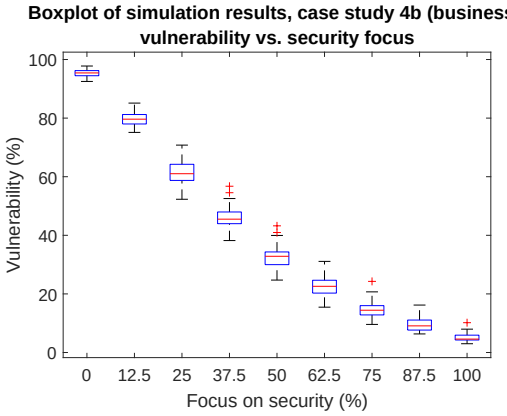


n = 250, 9 set-ups

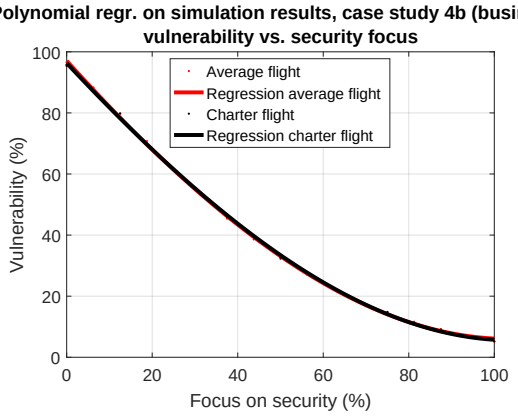
Polynomial regr. on simulation results, case study 4a (charter): vulnerability vs. security focus



Adjusted R^2: 0.9635
$RMSE$: 9.4386

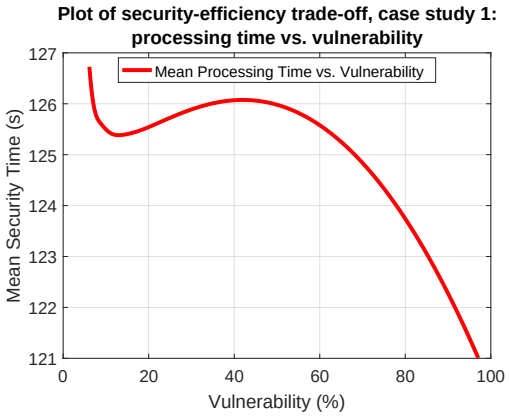


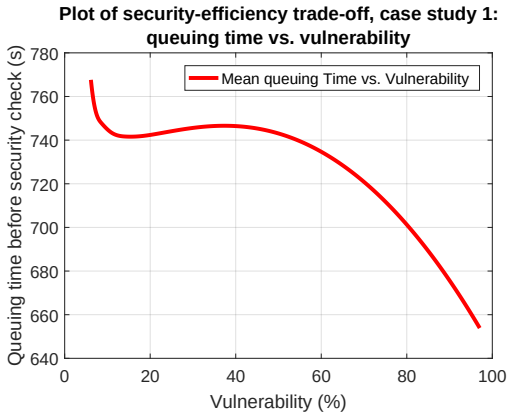
n = 250, 9 set-ups

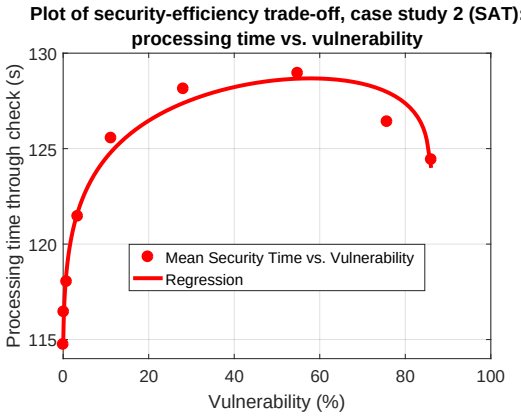


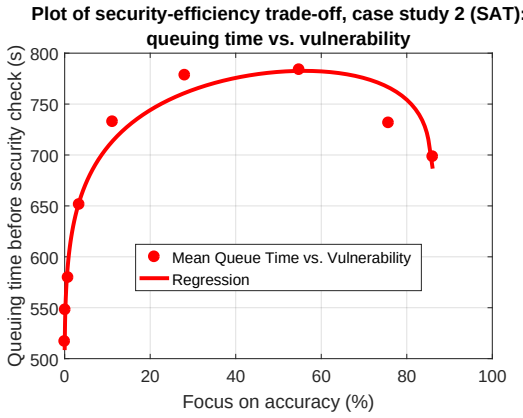
Adjusted R^2: 0.9989
$RMSE$: 1.0534

C.3. Security-efficiency trade-off plots

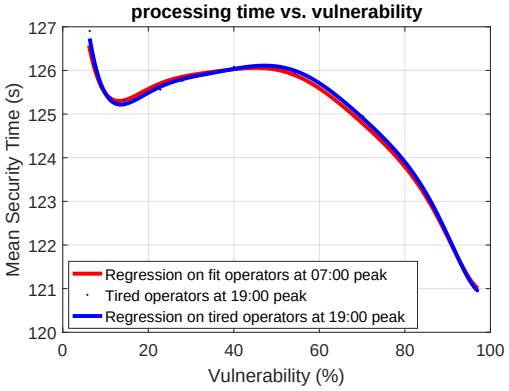




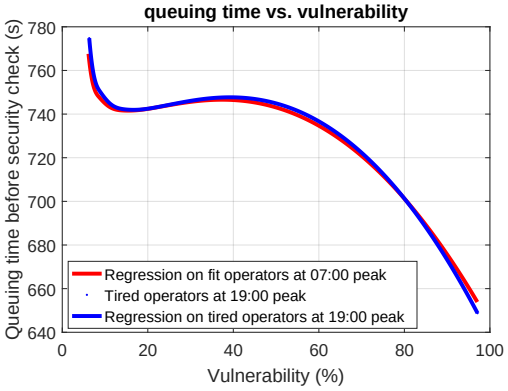




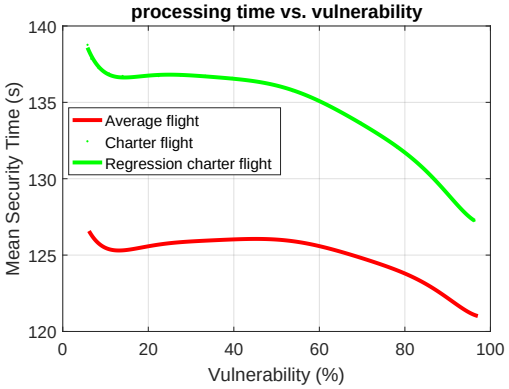
Plot of security-efficiency trade-off, case study 3 (fatigue): processing time vs. vulnerability



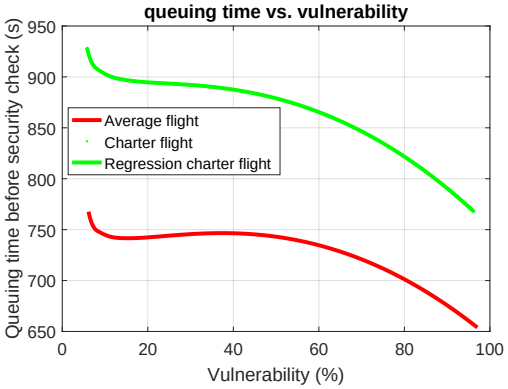
Plot of security-efficiency trade-off, case study 3 (fatigue): queuing time vs. vulnerability



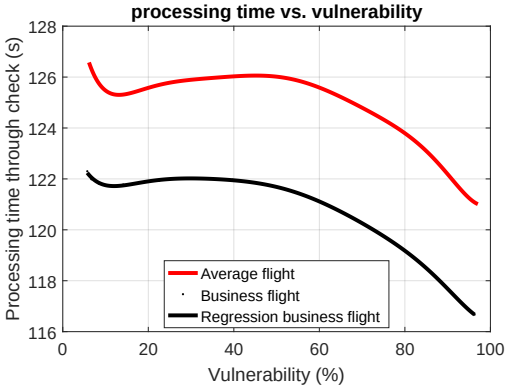
Plot of security-efficiency trade-off, case study 4a (charter): processing time vs. vulnerability



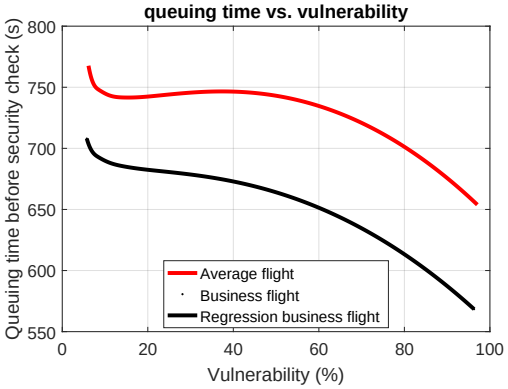
Plot of security-efficiency trade-off, case study 4a (charter): queuing time vs. vulnerability



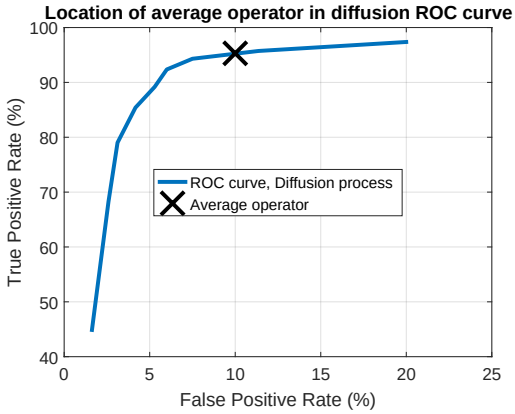
Plot of security-efficiency trade-off, case study 4b (business): processing time vs. vulnerability

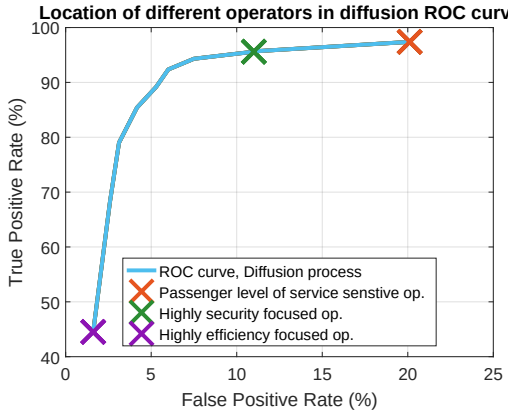


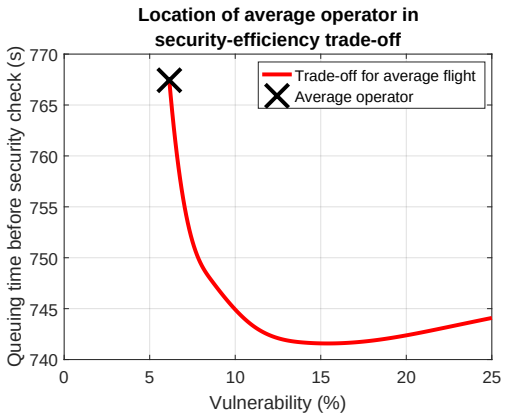
Plot of security-efficiency trade-off, case study 4b (business): queuing time vs. vulnerability

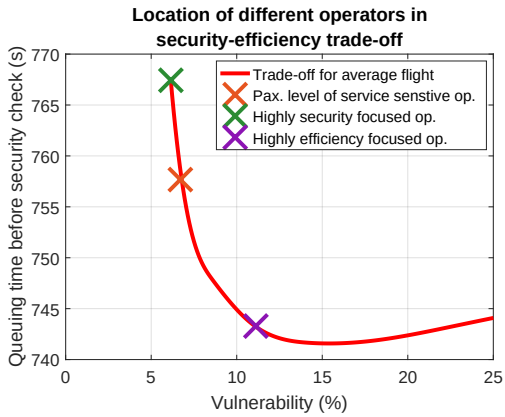


C.4. Placement of real operators in security-efficiency trade-off plots









D

Overview of calibrated parameters

The calibration of the AATOM model is an important part of this research. In order to provide a clear overview of the calibrated parameters, and to allow for the experiments to be replicated, all calibrated parameters are given in Table D.1.

Table D.1: Overview of calibrated parameters implemented in AATOM model

Parameter	Standard value	Variable range
Departure flight time schedule	RTHA, October 5th 2017, 05:00 - 10:00	-
Load factor of aircraft	0.9	-
Arriving passengers at $t_{non-peak_1}$	10% ~U(1:40:00, 1:20:00)	-
Arriving passengers at t_{peak}	80% ~U(1:20:00, 0:40:00)	-
Arriving passengers at $t_{non-peak_2}$	10% ~U(0:40:00, 0:20:00)	-
# opened security lanes	3	-
$t_{luggagedrop}$	N(54.6, 36.09)	-
$t_{luggagecollect}$	N(71.5, 54.95)	-
μ_a, σ_a	N(0, 1)	-
μ_f, σ_f	N(2.949, 1.12)	-
$thres_{\text{threat}}$	1.966	[0.842, 3.090]
d'	2.778	[0.5, 3]
$p_{\text{forbidden}}$	0.236	-
$p_{\text{randomcheck}}$	0	[0, 1]
$t_{\text{luggagecheck}}$	N(104.67, 80.86)	-
$t_{\text{physicalcheck}}$	N(43.00, 20.96)	-
$t^1_{\text{x-ray}}$	N(10.28, 5.06)	-
$t^2_{\text{x-ray}}$	N(16.44, 9.08)	-
$t^3_{\text{x-ray}}$	N(20.82, 11.04)	-
$t^4_{\text{x-ray}}$	N(21.00, 11.98)	-
p^1_{box}	0.317	-
p^2_{box}	0.485	-
p^3_{box}	0.188	-
p^4_{box}	0.010	-
z	0.4522	(0.0910, 0.8134)
a	0.9585	(0.5325, 4.7925)
b	0	(-3.834, 0.426)
v_a	-0.0859	-
v_f	0.0392	-
T_{er}, Physical check	32.713	-
T_{er}, Luggage check	61.819	-
T_{er}, X-ray, 1 box	3.670	-
T_{er}, X-ray, 2 boxes	7.861	-
T_{er}, X-ray, 3 boxes	12.242	-
T_{er}, X-ray, 4 boxes	15.359	-
s_z	0.158	-
s_t, Physical check	25.000	-
s_t, Luggage check	47.243	-
s_t, X-ray, 1 box	1.299	-
s_t, X-ray, 2 boxes	12.007	-
s_t, X-ray, 3 boxes	15.356	-
s_t, X-ray, 4 boxes	16.738	-
η_a	0.04988	-
η_f	0.02364	-
a_{v_a}	1.591E-03	(1.424E-03, 1.766E-03)
a_{v_f}	-7.259E-03	(-8.057E-04, -6.496E-04)
b_{v_a}	-9.244E-02	-
b_{v_f}	4.218E-02	-
$p^{dverage}_{\text{forbidden}}$	0.2067	-
$p^{business}_{\text{forbidden}}$	0.1270	-
$p^{charter}_{\text{forbidden}}$	0.4167	-
p^1_{box}	0.317	-
p^2_{box}	0.485	-
p^3_{box}	0.188	-
p^4_{box}	0.010	-

E

Contribution to AATOM

The major contribution to AATOM has been described elaborately in this thesis: the expansion of the model with diverse, autonomous and dynamic agents, and the implementation of this expansion. On this new AATOM model, new researches can be further built, as is described in the implications for related research in Chapter 7. But next to this, during the research process, a larger contribution to AATOM has been delivered. This contribution is described in this appendix.

At the beginning of this thesis research, a first version of the AATOM model was present. This model was until then however only created as a basis for the research of PhD candidate Stef Janssen. Because the AATOM model was a very promising concept also for other research, the idea was brought up to generate a *Baseline AATOM model*, which would serve as a basis platform for airport terminal operations research. This baseline model would also be the basis for this thesis research and would incorporate everything that was required to use AATOM when performing any study to airport terminal operations.

Together with Janssen and another MSc thesis student, this thesis research started with achieving consensus on what would be necessary for such a baseline model. The next step was to write a document that describes the model in a way that someone who is not familiar to AATOM could understand the components that constitute the model. The completed document is "AATOM - An Agent-based Airport Terminal Operations Model"[30]. This document describes into detail the environment, the agents and the interactions that are present within AATOM. As a part of this thesis research, a considerable contribution has been delivered in the writing and editing of this document.

The AATOM simulator is built in JAVA. Next to understanding the design of the constituent entities, it is therefore also important for researchers who are new to AATOM to understand the implementation approach of the model in JAVA. In order to help with this, Janssen wrote a document that introduces the implementation language of the AATOM simulator to people who are new to the simulator.[31] This research was the first one to use this AATOM implementation guide. This meant that a part of this thesis research was to identify the teething problems of a person that was new to AATOM. Together with Janssen, the implementation guide was improved and as the work continues, new additions are contributed in order to make sure that new people will not have the same questions every time. An initiative was started to add a Frequently Asked Questions (FAQ) chapter to the document, in which typical questions were documented and answered.

At the moment of writing, three other MSc students are working in the AATOM model. As this is the first research that started in AATOM, it is a part of this thesis research to be the first contact point for other students with questions, as they may encounter the same problems that have been encountered in this thesis research.

Bibliography

[1] IATA Consulting & ACI. Improved Level of Service Concept, 2017.

[2] Christopher J. Anderson. The psychology of doing nothing: Forms of decision avoidance result from reason and emotion. *Psychological Bulletin*, 1(129):139–167, Jan 2003.

[3] ASME Innovative Technologies Institute LLC. Executive Summary RAMCAP™ – A 7 Step Approach. pages 1–8, 2005.

[4] Drake Baer. The scientific reason why Barack Obama and Mark Zuckerberg wear the same outfit every day, April 28, 2015.

[5] Elias Bartholomew. *Airport and Aviation Security*. CRC Press, Taylor & Francis Group, 2010. ISBN 978-1-4200-7029-3.

[6] Rand Beers. *Risk Management Fundamentals* . US Department of Homeland Security, April 2011.

[7] Valerio Biscione. Drift Diffusion Process algorithm in MATLAB, December 2014.

[8] Bruce J. Boudreau. Improving the Airport Customer Experience. 2016.

[9] Gerald G. Brown and Louis Anthony Cox Jr. How Probabilistic Risk Assessment Can Mislead Terrorism-Risk Analysts. *Risk Analysis*, 13, No. 2:196–204, 2011.

[10] Andrew Caplin and Daniel Martin. The Dual-process Drift Diffusion Modle: Evidence from Response Times. *Economic Inquiry*, 54(2):1274–12832, April 2016. doi: ISSN0095-2583,10.1111/ecin.12294.

[11] Louis Anthony Cox Jr. Some Limitations of "Risk = Threat × Vulnerability ×Consequence" for Risk Analysis of Terrorist Attacks. *Risk Analysis*, 28, No. 6:1749–1761, 2008.

[12] Louis Anthony Cox Jr. Game Theory and Risk Analysis. *Risk Analysis*, 29, No. 8:1062–1068, 2009.

[13] Prof. dr. K.G. Curran, Prof dr. R andLangendoen, dr. O.A. Sharpanskykh, and S.A.M. MSc Janssen. Efficient and Secure Airports, 2017.

[14] Oxford Dictionary English. Definition of "Threat" in English.

[15] Jonathan Levav et al. Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences of the United States of America*, 108(17), April 2011. doi: 10.1073/pnas.1018033108.

[16] J.S. McCarley et al. Visual Skills in Airport-Security Screening. *Psychological Science*, 15(5):302–306, May 2004. doi: 10.1111/j.0956-7976.2004.00673.x.

[17] McCauley et al. Dynamic Circadian Modulation in a Biomathematical Model for the Effects of Sleep and Sleep Loss on Waking Neurobehavioral Performance. *Sleep*, 36(12):1987–1997, Dec 2013. doi: 10.5665/sleep.3246.

[18] Nicole Hattenschwiler et al. A First Exploratory Study on the Relevance of Everyday Object Knowledge and Training for Increasing Efficiency in Airport Security X-ray Screening . *Proceedings of the 49th Annyual IEEE International Carnahan Conference on Security Technology*, September 2015.

[19] Tibor Bosse et al. A Language and Environment for Analysis of Dynamics by Simulation. *International Journal on Artificial Intelligence Tools*, 2006.

[20] et al. Andrea Saltelli. *Sensitivity Analysis in Practice*. John Wiley & Sons, Ltd, 2004. ISBN 0470870931.

[21] BEMOSA (Behavioral Modeling for Security in Airports). A Behavioral Model of Security in Airports: Preliminary Results - Aerodays Madrid, 2011.

[22] Reuters Gilbert Kreijger. Schiphol buys 60 body scanners .

[23] Konstantina Gkritza, Debbie Niemeier, and Fred Mannering. Airport security screening and changing passenger satisfaction: An exploratory assessment. *Journal of Air Transport Management*, 12(5):213–219, 2006. ISSN 09696997.

[24] Hayward J. Godwin. The Influence of Real-World Factors on Threat Detection Performance in Airport X-Ray Screening. *PhD Thesis*, December 2008.

[25] Donald Gross. *Fundamentals of queueing theory*. Willey series in prbability and statistics, 2008. ISBN 978-0-471-79127-0.

[26] ANP Het Parool. Hoog overleg over problemen Schiphol, 8th of May 2017.

[27] Franziska Hofer and Adrian Schwaniger. Using threat image projection data for assessing individual screener performance. *WIT Transactions on the Built Environmen*, 82:417–426, 2005. doi: 10.5167/uzh-97993.

[28] Robert R. Hoffman and David D. Woods. Beyond Simon's slice: Five fundamental trade-offs that bound the performance of macrocognitive work systems. *IEEE Intelligent Systems*, 26(6):67–71, 2011. ISSN 15411672.

[29] IATA. Level of Service Concept.

[30] Knol Janssen, Blok. AATOM, An Agent-based Airport Terminal Operations Model. 20th of October, 2017.

[31] Stef Janssen. AATOM - An Agent-based Airport Terminal Operations Model Simulator. 20th of October, 2017.

[32] Stef Janssen and dr. Alexei Sharpanskykh. Agent-based Modelling for Security Risk Assessment.

[33] PhD Jeffrey T. Fairbrother. *Fundamentals of Motor Behavior, Human Kinetics' fundamentals of sport and exercise science series*. Human Kinetics, 2010. ISBN 9780736077149.

[34] Marson S. Jesus. Trade-off analysis between security and efficiency of airport operations. January 2018.

[35] Alan (Avi) Kirschenbaum. The cost of airport security: The passenger dilemma. *Journal of Air Transport Management*, 30:39–45, 2013. ISSN 09696997.

[36] Arthur Knol. Analysing Security and Efficiency of Airport Terminal Operations, Literature Review. 2017.

[37] De Kwis. More restrictions in security check on airports, March, 2017.

[38] Adrian J Lee and Sheldon H Jacobson. The impact of aviation checkpoint queues on optimizing security screening effectiveness. *Reliability Engineering and System Safety*, 96(8):900–911, 2011. ISSN 09518320.

[39] Hans van Dongen Matthew M. Walsh, Glenn Gunzelmann. Computational cognitive modeling of the temporal dynamics of fatigue from sleep loss. *Psychonomic Society*, 24:1785–1807, Feburary 2017. doi: 10.3758/s13423-017-1243-6.

[40] TUV NORD Matthias Springer. What's the difference between safety and security?

[41] Merriam-Webster. Definition of Security Risk.

[42] Eve Mitleton-Kelly, Scale Emergency, and Transport Domains. *Co-evolution of Intelligent Socio-technical Systems*. 2013. ISBN 978-3-642-36613-0.

[43] National Research Council of the National Academies. Review of the Department of Homeland Security's Approach to Risk Analysis - Chapter: 4 Evaluation of DHS Risk Analysis. page 52, 2010.

[44] RTL Nieuws. Extra veiligheidscontroles bij Schiphol om 'signaal dreiging'. July 2016.

[45] Martin Hagger Nikos Chatzisarantis. *The Social Psychology of Exercise and Sport*. Open University Press, McGraw-Hill Education, 2005. ISBN 0335216188.

[46] Roger Ratcliff and Gail McKoon. Modeling response times for two-choice decisions. *Neural Comput.*, 9: 347–356, September 1998. doi: 10.1111/1467-9280.00067.

[47] Roger Ratcliff and Jeffrey N. Rouder. The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. 20(4):873–922, April 2008.

[48] Security management Rotterdam The Hague Airport. Passenger predictions for security check. 2017.

[49] Adrian Schwaninger. Reliable Measurements of Threat Detection. *AIRPORT*, 1:22–23, 2003.

[50] D. Scotti. Measuring Airports' Technical Efficiency: Evidence from Italy. *PhD thesis, University of Bergamo, Italy*, 2011.

[51] Alexei Sharpanskykh. TU Delft AE4422-16 Agent-based Modelling and Simulation in Air Transport. unpublished lecture slides, 2017.

[52] Arthur O'Sullivan; Steven M Sheffrin. *Economics : principles in action.* Needham, Mass.: Prentice Hall, 2003.

[53] H.A. Simon. The Sciences of the Artificial. *MIT Press*, 1969.

[54] Philip L. Smith and Roger Ratcliff. Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, 27(3):161–168, March 2004. doi: https://doi.org/10.1016/j.tins.2004.01.006.

[55] Mark G Stewart and John Mueller. Cost-benefit analysis of airport security: Are airports too safe? *Journal of Air Transport Management*, 35:19–28, 2014. ISSN 09696997.

[56] Thales. Airport Infrastructure Security Towards Global Security. *Airport Infrastructure Security Towards Global Security*, pages 1–16.

[57] Imri Sofer Thomas V. Wiecki and Michael J. Frank. HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Neuroinform*, 02 August 2013. doi: https://doi.org/10.3389/fninf.2013.00014.

[58] John Tierney. Do you suffer from decision fatigue?, August 17, 2011.

[59] Tolga Ülkü. Efficiency of German Airports and Influencing Factors 1 1 Introduction. pages 1–28, 2010.

[60] A.J. van den Berg. Agent-Based Threat Identification Literature Review. pages 4–7, 2017.

[61] Koen van Impe. Simplifying Risk Management, March 28, 2017.

[62] Joachim Vandekerckhove and Francis Tuerlinck. Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods*, 40(1):61–72, 2008. doi: 10.3758/BRM.40.1.61.

[63] Meritxell Vinas Tio. Study of Airport Capacity cs. Efficiency SESAS challanges. (January):64, 2010.

[64] William E Weiss. Dynamic security: an agent-based model for airport defense. *Proceedings of the 2008 Winter Simulation Conference*, (2001):1320–1325, 2008.

[65] Wayne A. Wickelgren. Speed-Accuracy Tradeoff and Information Processing Dynamics. *Acta Pshychologica*, (41):67 – 85, 1977.

[66] David Woods. Chapter 2: Essential Characteristics of Resilience. *Resilience Engineering*, pages 21–34, 2002.

[67] Kwang Eui Yoo and Youn Chul Choi. Analytic hierarchy process approach for identifying relative importance of factors to improve passenger security checks at airports. *Journal of Air Transport Management*, 12(3):135–142, 2006. ISSN 09696997.