A game-theoretic approach to assess adversarial risks

S. Meng, M. Wiens & F. Schultmann
Institute for Industrial Production, KIT, Karlsruhe, Germany

Abstract

In our complex world today almost all critical infrastructures are interdependent and thus vulnerable to many different external and internal risks. To protect them against the greatest risks, a well-functioning risk management process is required to develop appropriate safety and security strategies. There are many well-established risk analysis methods existing. They predominantly apply empirical models and statistical data to quantify the risks. Within the realms of natural or aleatory risks this approach is considered suitable and functional. However, it could be a fatal flaw to apply such conventional, history-orientated models in order to assess risks that arise from intelligent adversaries such as terrorists, criminals or competitors. Approaches of classic risk analysis generally describe adversaries' choices as random variables, thus excluding the adversaries' behaviour and ability to adapt to security strategies. One possibility for considering human behaviour when analysing risks is the recourse to game theory. Game theory is the paradigmatic framework for strategic decision-making when two or more rational decision-makers (intelligent adversaries) are involved in cooperative or conflictive decision situations. In our study we propose an approach for combining a classic risk analysis method with a game-theoretic approach. Using a defender-offender game as a basis, we simulate, exemplary for a terrorist attack against public transport, the behaviour and reactions (to applied security strategies of the defender) of a rational player acting as an adversary. Although risk analysis and game theory are very different methodologies, we show that linking them can significantly improve the quality of forecasts and risk assessments. If the behaviour and reactions of intelligent adversaries need to be considered, our approach contributes to enhance security through improving the allocation of scarce financial resources.
Keywords: risk, risk analysis, adversarial risks, game theory.


WIT Transactions on Information and Communication Technologies, Vol 47, © 2014 WIT Press
www.witpress.com, ISSN 1743-3517 (on-line)
doi:10.2495/RISK140131

1 Introduction

When analysing risks, the risk R of an event E is almost always defined as a function of its probability of occurrence p and the possibly resulting negative consequences c ($R_E = f(p_E, c_E)$) or as a function of the vulnerability of the element at risk v, the threat to the element at risk t and c ($R_E = f(v_E, t_E, c_E)$). However, this only works properly when dealing with random (or aleatory) risks. By this we understand risks which can be expressed by stochastic properties of random events, i.e. their probability distribution or frequency. Examples are, inter alia, throwing (ordinary) dice, weather forecasts, technical failure rates and simple investment decisions. Within the realm of random events, practical risk management is primarily based on an optimal adaptation to empirical distributions. However, this approach can lead to misjudgement and flaws when it comes to risks that arise from deliberate choices of two or more rational and interacting individuals [1]. In such cases it is not enough to determine the probability of an event on mere empirical grounds because this probability depends both on the beliefs and motives of the human counterpart as well as on the dynamics between this counterpart and the decision-maker himself. The common way to deal with such strategic decision situations with two or more interacting individuals is the recourse to game theory.

Game theory is originally a mathematical theory that can be applied to complex strategic decision situations. It models human interactions in order to analyse the reasoning process of (mostly) rational actors, to predict the outcome (equilibrium) of this interaction and to analyse the properties of this equilibrium in terms of efficiency, stability and plausibility. Fields of application are, inter alia, economic markets and networks, cooperative arrangements, competitive situations, crimes and recently also terrorism. The last two applications fall into the domain of adversarial risks, where the players' dominant interests are in harsh conflict.

In the following we want to show that adding game-theoretic elements to classic risk analysis approaches can be helpful when dealing with terrorist risks. We want to demonstrate the functioning of our proposed approach in the context of adversarial risk management and security of public rail transport. Public rail transport is relevant for the well-functioning and welfare of modern societies and therefore referred to as a critical infrastructure [2]. If such a system is partially or entirely interrupted, it has not only negative effects on public safety and security, public health, and public life, but also on politics and whole economies [3]. In the past decades, public rail transport has been several times the target of terrorist attacks. Terrorists seem to favour this critical infrastructure because attacks against them usually cause many casualties, large financial losses and extensive media coverage. Being a preferential target for terrorists, railway systems, and public railway transport systems in particular, are in need of particular protection against terrorist attacks and this protection demands additional scarce resources.

The remainder of our paper is structured as follows. In section 2, we describe the classic approach of risk management and risk analysis vis-à-vis terrorist risks. In section 3, a brief outline of the game-theoretic approach to analyse

adversarial risks is presented. Section 4 provides an example which allows a direct comparison of the classic risk analysis approach with game theory, also hinting at the advantages of a symbiotic approach. The paper closes with a discussion of the results and a conclusion in section 5.

2 Risk and classic risk analysis

In a first step it is necessary to define the term risk. Risk inherently incorporates both the possibility to make losses and gains (within the context of this paper we will solely consider the former). Knight [4] describes risk in his seminal work *Risk, Uncertainty and Profit* as a "measurable uncertainty" (and "true" uncertainty as immeasurable uncertainty). Bonß [5] and Ayyub [6] define risk as uncertainty about the outcome of an event. The International Organization for Standardization [7] similarly defines risk as an "effect of uncertainty on objectives", with effects being seen as "deviation from the expected - positive and/or negative", whereas Kaplan and Garrick [8] characterize risk as a complete set of triplets, with each triplet consisting of a scenario in combination with the corresponding likelihood and the consequences of that scenario. Similar, Haimes [9] defines risk as the product of impact, vulnerability and threat, and Aven and Renn [10] state that risk is the product of uncertainty about and severity of consequences of an event. What all definitions of risk have in common is, however, that they somehow associate risk with uncertainties and/or probabilities [11].

All activities, whether in private or business life, carry risks or, in other words, "we are all constantly exposed to risk" [12]. Individuals, companies or governments therefore permanently conduct some sort of risk management. Risk management can be defined as "the identification, assessment, and prioritization of risks followed by coordinated and economical application of resources to minimize, monitor, and control the probability and/or impact of unfortunate events" [13]. Key principles of risk management are, inter alia, that it creates values, is part of the decision-making process, explicitly addresses uncertainty, is based on available information, and that it takes human factors into account [7]. Risk management is usually an iterative cyclic multi-step process, with risk analysis being one of its key elements. Risk analysis is about finding initiating events (hazards, threats, opportunities) to understand and describe relevant risks, their consequences, and the associated likelihoods of those consequences [14, 15]. The results of a risk analysis can be represented qualitatively, semiquantitatively or quantitatively. There is a wide variety of risk analysis tools and methods, such as brainstorming and expert intuition, expert audits, Delphi method, structured what-if technique (SWIFT), probabilistic risk analysis (PRA), failure modes and effects analysis (FMEA), fault-tree analysis (FTA), event-tree analysis (ETA), Bayesian networks, Monte Carlo simulation and so forth [6, 13, 14]. If a risk analysis method is predominately based on intuition and/or probabilities, the authors refer to them as *classic* risk analysis approaches.

Risk has many features according to which it can be categorised (origin, cause, severity, controllability and so forth). One feature that is of special interest within

the scope of this paper is whether risk is the result of randomness or deliberate choice. This distinction highlights the important point that some categories of risk can arise from natural or aleatory uncertain hazards (e.g. natural hazards or engineered systems) and others from rational and intelligent adversaries [16]. Key differences between both risk types are summarized in Table 1.

Table 1: Uncertain hazards versus intelligent adversaries (after Parnell *et al.* [16]).

	Uncertain hazards	**Intelligent adversaries**
Available data	historical data, and data from experiments and simulations	limited reliable historical data
Risk of occurrence	risk reasonably well defined (e.g. models, expert opinions)	considerable ambiguity of risk (e.g. adversaries' adaptability)
Geographic risk	well-known specific areas at risk	all areas at risk (attacks are anywhere, anytime possible)
Information	information are shared (at least) within the scientific community	asymmetry of information; not all information are shared for national security reasons
Event type	natural or random (aleatory) event; occurrence not influenceable	intelligent adversaries event; occurrence influenceable (e.g. through governmental measures)
Preparedness and prevention	well-known safety and security measures available	no reliable safety and security measures available (lack of knowledge about capabilities and intentions of adversaries)

When applying classic risk analysis approaches to adversarial risks, risks may be underestimated in a systematic way [16, 17]. The problem is that there are no reliable probabilities available for risks arising from intelligent and rational acting adversaries because they are capable and willing to adapt their attack strategies in response to defensive measures [18]. As already mentioned, an alternative way to analyse adversarial risks is applying game theory.

3 Game theory

Game theory analyses human behaviour and social interactions within modelled strategic settings (games). It deals with interactive decision-situations where the decisions of two or more decision-makers (players) are interdependent. Players can be single actors or represent larger entities like groups, companies, societies, nations et cetera. Furthermore, all decision-makers are aware of the fact that they cannot predict the results of a decision-situation in isolation, i.e. without considering the behaviour and the decisions taken by all the other players. Game

theory provides an abstract, mathematical language to describe, analyse and solve strategic decision-situations. Games are abstract models which simplify complex real-life decision-situations in order to focus on the essentials of a (by a set of assumptions) neatly defined game [19]. Thus the advantages of this approach are at the same time the obstacles to a more prominent role in risk management: In real life decisions situations, individuals are usually confronted with a high degree of complexity (number of parameters which define the decision-situation), uncertainty (about motivations, the parameter values, the distribution of information et cetera) and with the often erratic, seemingly irrational, behaviour of humans. This is one reason why risk analysis reluctantly applies game theoretic models when analysing real-life strategic decision-situations. Game theory does not aim at instructing perfect decisions under real-life circumstances, rather it is a helpful tool for decision-makers to gain a profound understanding of the nature of strategic decision-situations and to analyse them in a coherent and structured manner [20].

In the following we present a small example for a sequential defender-attacker-game (see Figure 1).

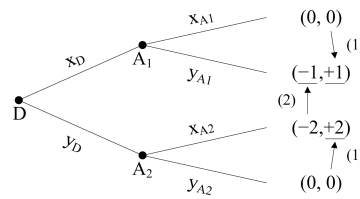

Figure 1: Example of an defender-attacker-game.

In this example the defender makes his move first (decision node D). Thus, the attacker has the chance to react on the move of the defender; he will consider the defender's move when making his own move (decision nodes $A1$ and $A2$). The pair of numbers behind each branch indicates the gains and losses for both the defender and the attacker, whereas the first number is assigned to the defender and the second number to the attacker. x and y indicate the available decision options at each decision node. Both players, who are assumed (to be intelligent and) to act rational, will take that decision option which maximises their individual pay-offs. Furthermore, the defender, who makes the first move, also needs to consider the reaction of the attacker (on his first move). This decision problem at hand of the defender can be solved using backward-induction: The defender first (1) identifies the attacker's preferred decision options for both decision nodes $A1$ and $A2$. Out of the remaining two decision options the defender then (2) chooses the one which maximises his pay-off. In this example the attacker would prefer y_{A1} over x_{A1} $(+1 > 0)$ if the defender chooses x_D, and x_{A2} over y_{A2} $(+2 > 0)$ if the defender chooses y_D. It follows that the defender now only has to compare the two

combinations (x_D, y_{A1}) and (y_D, x_{A2}) to find his optimal solution. As the former combination has a higher pay-off (for him) ($-1 > -2$), the defender will choose decision option x_D.

4 Example: optimal defence against adversarial risks

In the following we demonstrate with another simple yet more detailed example the distinctive advantage of game-theoretic reasoning compared to classic risk analysis when it comes to adversarial risks. The important and new aspect is the explicit consideration of the opponent's motivation which underlies all of his actions. This enables the defender of a critical infrastructure to direct scarce resources to very specific elements of the infrastructure where the opponent's adversarial motivation is strongest. In contrast, from the viewpoint of classic risk analysis the most vulnerable elements are those where the defender has experienced the highest losses (financial, material, immaterial) in the past. Of course former hotspots (infrastructure elements that already had been targets of an adversarial attack in the past) - can, and often do - coincide with future hotspots (infrastructure elements that will become attractive targets for attackers in the future). This is the reason why in many cases forecasts from classic risk analysis and game-theoretic approaches do not diverge. However, in the following we focus on a selected real-life decision-problem where game-theoretic reasoning actually will improve protection.

The basic setting of our example is a slightly modified version of the already introduced sequential defender-offender-game. We consider the case of a railway station which is highly frequented by passengers during rush hour, but sparsely frequented outside rush hour. The deterministic version of this defender-offender-game is quite easy to solve due to its unique solution. As it is not representing the ubiquitous uncertainty of the real world, we add some uncertainty to the decision-problem by assuming incomplete information on the side of the operator (defender) about the motivation of the terrorist (attacker). We distinguish two types of terrorists: The first type of terrorist ($T1$) strives for a new political leadership or system and thus regards as an enemy any entity which keeps the system running or symbolizes the systems strength. Consequently he seeks to achieve material or functional damages to critical infrastructures (and institutions) and symbolic landmarks. In his view any system-debilitating act is an appropriate measure. With regard to our example of the railway station, $T1$ could try to damage SCADA systems (functional inability) or destroy the structure of the station-building, which can be interpreted as a symbol power. $T1$ is also referred to as *anti-system-terrorist*. It is important to note that $T1$ seeks to avoid fatalities for two reasons. First, from his point of view the general public is not involved and, second, he depends on the support of the citizens to attain his goal (revolution or overthrow of the system). $T1$ is cautiously and extensively comparable with the German RAF or the Spanish ETA, which seek to change the economic or political system or simply to debilitate the state's authority. At the same time both depend on the public's support - at least to a significant extent.

The left table in Figure 2 shows the pay-offs of $T1$ for four different strategies: If he attacks selectively human targets during rush-hour he experiences a net-loss of -2 (utility-)units, because the extremely high number of fatalities is the last thing he wanted to achieve (Note: It is impossible to value human life with a number; all numbers for valuing human life in this example are fictitious, to demonstrate the principles of this approach). Concentrating on human targets outside rush hour reduces the number of fatalities, but after all, this is no effective strike against the system as such. That is why his pay-off in this case is still negative (-1). The best strategy for him is to attack the building or a SCADA system outside rush-hour in order to limit collateral damage (highest pay-off of $+2$ units). The second-best option is attacking a non-human-target during rush-hour which effectively hits the system. As this also causes significant collateral damages, his pay-off is reduced to $+1$ in this case.

anti-system terrorist (T1)	human target (H1)	non-human target (H2)
rush hour (R1)	-2	1
outside rush hour (R2)	-1	2

fanatic terrorist (T2)	human target (H1)	non-human target (H2)
rush hour (R1)	3	2
outside rush hour (R2)	2	1

Figure 2: Terrorist pay-offs depending on type and strategy.

In contrast, the second type of terrorist ($T2$) strives for a completely different society as a whole and is characterised by deep hatred concerning people's way of life and value systems. $T2$ is referred to as *fanatic terrorist*. This type corresponds to extremely radical and fanatic terror-groups such as extremist sects or religious warriors, with Al-Qaeda as its most prominent example. $T2$ seeks at both maximum damage and widespread fear. Consequently $T2$ does not shrink away from mass murder and has reached its goal if the general public is paralysed. The table on the right side of Figure 2 shows the pay-offs of $T2$. As this type has no qualms about killing or injuring people, he never receives a negative pay-off. He gains the highest pay-off ($+3$) when the number of fatalities is maximal, which corresponds to the strategy of attacking human targets during rush hour. Pay-offs then gradually decline with strategies, where the chance to wreak damage is limited. The least attractive option for $T2$ is to attack a non-human target outside rush hour. But as there is still enough damage (financial, material and immaterial), his pay-off remains positive ($+1$).

In the next step we define the pay-offs and strategic options of the defender who represents a system operator or security service. Whenever an attack occurs the defender experiences losses, represented by the negative pay-offs in the left table of figure 3. The worst constellation for the defender would be an attack against human targets during rush hour, which is evaluated with a pay-off of -6. It indicates that this is the worst case for $T2$. The negative utility, which is expressed by this number, reflects the impact of such a horrifying scene and the pressure

of the general public, which will certainly not tolerate these kind of aggressions. The minimum loss (-1) occurs when a non-human target is attacked outside rush hour, because the number of fatalities is low. The remaining two constellations are associated with a medium loss of -3 utility units.

defender (D)	human target (H1)	non-human target (H2)
rush hour (R1)	-6	-3
outside rush hour (R2)	-3	-1

defence-units	human target (H1)	non-human target (H2)
rush hour (R1)	d_2	d_1
outside rush hour (R2)	d_3	d_4

Figure 3: Defender pay-offs depending on type and strategy.

The strategy of the defender consists in maximising his pay-off through distributing defence units d_i, where the index i indicates the four attack scenarios as pictured in the right table of Figure 3 (an attack-strategy for the attacker is an attack-scenario from the defender's point of view). For example, attack-scenario 1 corresponds to the constellation *R1-H2*, attack-scenario 2 to *R1-H1* and so forth. For the sake of simplicity we set an upper limit D to the sum of all defence-units:

$$D \geq \sum_{i=1}^{4} d_i$$

The allocation of defence-units to the four scenarios is costly for both the attacker and the defender. We assume linear costs $c \cdot d$ with $c = 0.2$ for the defender and $c = 1.0$ for the attacker. On the part of the defender these cost represent the expenditure to install, maintain and operate the defence-units, whereas on the part of the attacker the defence-units have the effect of a barrier or a obstacle. This is because they reduce the attractiveness of a particular attack-strategy. Since we consider a sequential game with perfect, but incomplete, information, the terrorist can observe at any time the number of installed defence-units relevant for each of his four attack-options. Thus, the net pay-off for the terrorist is his raw pay-off (no defence measures existing) as shown in Figure 2 minus the number of defence-units deployed by the defender. Prerequisite for the terrorist to execute an attack is that his net pay-off is strictly positive. Let us take the case of $T1$ as an example. If the defender uses two defence-units to prevent attack-scenario 4 (*R2-H2*), the attractiveness of *R2-H2* to the terrorist drops from $+2$ to 0. This implicates that attack-strategy 4 is no longer interesting for the terrorist. However, if the defender uses only one instead of two defence-units, the attractiveness only drops from $+2$ to $+1$. This implicates that the terrorist is now indifferent between attack-strategy 1 (low primary attractiveness, no barriers) and attack-strategy 4 (high primary attractiveness, low barrier). By installing defence-units the defender can prevent an attack, but he cannot attenuate the impact of an attack. The defender's strategies can be represented by the vector $\mathbf{d} = \{d_1, d_2, d_3, d_4\}$.

In the next step, we show that classic risk analysis and game theory can lead to quite different results. We assume a time horizon of 25 periods. In each period the railway station can be attacked by $T1$ with probability p, or by $T2$ with probability $(1 - p)$. The defender initially knows the probability p and the pay-offs for both $T1$ and $T2$ as given in Figure 2. Thus, he has a priori a realistic picture about the threats arising from $T1$ and $T2$ and the underlying motives of the terrorists. This information comes from different sources (own experiences, intelligent services, public authorities). The defender has access to $D = 5$ defence-units. We consider the classic risk analysis approach first. If we look at the setting from a classic risk analysis perspective, the defender distributes the maximum number of defence-units D with respect to the distribution of experienced damage and intelligence information. Let x_i be the cumulated number of attacks which occurred in scenario i and u_i^D the corresponding pay-off to the defender. The defender then devotes a proportion d_i to scenario i with:

$$d_i = \frac{x_i u_i^D}{\sum_{j=1}^{4} x_j u_j^D} \cdot D$$

In a second step "nature" randomly determines which type of terrorist carries out an attack. For the corresponding type we determine the net pay-off for each of the scenarios according to the function $u_i^T[net] = u^t - d_i$. The terrorist executes the attack where the net pay-off is positive and highest (if two or more net pay-offs are equal, the terrorist is indifferent and "flips a coin"). The defender experiences a loss from the attack and, thus, adjusts the x-value of the particular scenario where the attack occurred. This procedure is repeated 25 times. Technically spoken, this risk-analysis-procedure is a specific variant of *fictitious play*. This is an adaptive procedure where the actors try to forecast the counterpart's most probable strategy on the basis of historic frequencies. Table 2 shows the aggregated results of this simulation for $p = 1$, $p = 0.5$ and $p = 0$.

Table 2: Simulation results for the classic risk analysis approach.

	p = 1 (anti-system-terrorist)				p = 0.5				p = 0 (fanatic terrorist)			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Scenario	attacks absolute	attacks relative	defence-units	losses + cost	attacks absolute	attacks relative	defence-units	losses + cost	attacks absolute	attacks relative	Defence-units	losses + cost
1 (R1-H2)	1	0.25	59.8	15.0	4	0.16	26.1	16.2	7	0.29	32.3	27.5
2 (R1-H1)	0		3.7	0.7	4	0.16	23.8	28.8	6	0.25	55.7	53.1
3 (R2-H1)	0		1.9	0.3	4	0.16	24.0	16.6	7	0.29	31.6	27.3
4 (R2-H2)	3	0.75	59.6	15.0	12	0.50	51.1	22.2	4	0.17	5.4	5.1
	4		125	31	24		125	83.8	24		125	113

Column *(1)* reports the absolute number of attacks which occurred under each of the four attack-scenarios, column *(2)* reports the relative frequency of each attack-scenario, column *(3)* reports the number of used defence-units, and column *(4)* reports the *disutility* of the defender which is made up of the experienced losses and the costs for the defence-units.

Classic risk analysis is quite successful vis-à-vis $T1$ ($p = 1$). As he never attacks in scenario 2 and 3, the remaining 'scenarios under risk' are relatively easy to handle. There are just four attacks, one under the constellation of scenario 1 and three attacks under scenario 4. These attacks occur because the defender needs some initial periods to learn and adjust the defence-units. Once he has learned (by the simple adaptive heuristic described above), he focusses exclusively on the critical scenarios 1 and 4 which is sufficiently deterrent in this case. Overall cost sum up to 31 units: 6 result from the losses of the four attacks and 25 are the cost for installing 5 defence-units in every period. The defender's losses are highest if he faces $T2$ ($p = 0$). In this case there are three critical scenarios, with the most dangerous one absorbing more than half of the defence-units. Whereas in the case of $T1$ damage costs were low (or even negligible), damage cost are the bulk of the overall cost to the defender in the case of a $T2$. The main problem is that the defender cannot fully deter the terrorist, but non-stop reacts to the observed first mover's past decisions without trying to anticipate the next move by strategic thinking. This leads to a kind of "hare-and-tortoise-tragedy" for the defender. He tends to invest defence-resources in a "grazed willow". In the mixed constellation ($p = 0.5$), the defender is randomly confronted with one of both types. Since the defender always has higher pay-offs when confronted with $T1$ than with $T2$, the overall costs are reduced compared to when he is solely confronted with $T2$. However, cost are not averaged, because the defender cannot deter both types and, thus, faces again the "hare-and-tortoise-tragedy" - although just to a limited extent.

Let us now turn to pure game-theoretic reasoning. Following this approach the defender distributes an effective number of defence-units D with respect to the pay-off structure of the terrorist. So the defender considers Figure 2 and tries to minimize the losses resulting from the most critical scenarios by 'neutralising' the terrorist's pay-off. This approach can already slightly improve the results in the case of $T1$. If the defender installs one defence-unit on R1-H2 and two units on R2-H2 from the first period on, his optimal strategy is thus $\mathbf{d} = \{1, 0, 0, 2\}$. By doing so he fully deters the terrorist and saves two defence-units each round. The resulting damage from attacks is zero (instead of six in the classic, adaptive approach) and the cost for defence is reduced to 15 (instead of 25 in the former approach). However, the results are significantly better in the case of $T2$. The first and obvious point is, that the defender can prevent the harshest strike by installing three defence-units on *R1-H1* (scenario 2). Additionally, $T2$ has two second-best options which are both associated with a damage of -3 for the defender. The defender cannot prevent an attack because it requires two defence-units each. But he can lower the attractiveness of the two second-best strategies by using one defence-unit for each of them. Now the terrorist is indifferent between the two second-best options and the low-risk third-best option, which he otherwise would not have cared about. The optimal strategy for the defender is thus $\mathbf{d} = \{1, 3, 1, 0\}$. The expected damage per period is -2.3 which sums up to $58,3$ over all 25 periods. Together with the defence-costs (25 units) the total cost add up to 83.3 (compared to 113 in the classic approach). The coherent consideration of the terrorist's motives reduces total cost by more than 25% in this example. The

mixed case ($p = 0.5$) is treated similar. The first and important step is again to neutralise the high-risk scenario vis-à-vis $T2$ (cost of -3 defence-units). In the mixed-scenario it is not optimal, however, to distribute the remaining two defence-units on the second-best options of $T2$. By this, $T1$ can act without hindrance which means an expected loss of $0.5 \cdot 3 = 1.5$ units per period. The best strategy is actually to cover the high-risk-options for both types and to save the remaining defence-units, thus $\mathbf{d} = \{1, 3, 0, 0\}$. The defender then has minimal damage-costs of 50 and defence-cost of 20. His total (expected) loss adds up to 70. Compared to the adaptive approach (83.8) cost is reduced by more than 16%.

5 Summary and conclusions

In this paper we argue for a stronger game theoretic foundation of classic risk analysis in the context of adversarial risks. We have proved with an illustrative example that game-theoretic reasoning can substantively improve defence-decisions, because it takes the motives of the adversary directly into account. By a direct focus on the motivation it is possible to allocate defence-resources in a more efficient manner compared to an adaptive heuristic approach where the defender as second-mover constantly drags behind. Game theory performs well in our example because we assume that the defender disposes very precise information about the distribution of terrorists' types and their motives. Of course, this is a very strong assumption. But our core arguments can be maintained if we allow a limited degree of uncertainty regarding terrorists' pay-offs. Clearly, game theoretic parameters have to be filled with content which cannot be provided by game theory only. In future work, we generalize this model and try to determine the critical boundaries for robust results. This should also give more insight into the question, in what respect assumptions about information are most critical. Furthermore, we have demonstrated that game theory is a practical tool for decision-makers to structure and simplify adversarial strategic decision-situations. Also, game theory is an appropriate measure to mathematically describe human behaviour and to gain profound knowledge about the nature of such decision-situations. We think combining game theory with classic risk analysis can contribute to sustainably improve risk analysis in companies (governments) when threatened by human beings in form of competition, crime or terrorism. Research has to be done in identifying ways to combine game theory and classic risk analysis.

Acknowledgement

The research documented in this paper was supported in part by the security research project *RIKOV*, financed by the German Federal Ministry of Education and Research (BMBF).

References

[1] Hall, Jr, John R., The elephant in the room is called game theory. *Risk Analysis*, **29(8)**, p. 1061, 2009.

[2] Abou El Kalam, A., Deswarte, Y., Baïna, A. and Kaâniche, M., Polyorbac: A security framework for critical infrastructures. *International Journal of Critical Infrastructure Protection*, **2(4)**, pp. 154–169, 2009.

[3] Freudenberg, D., *Theorie Des Irregularen: Partisanen, Guerillas und Terroristen im modernen Kleinkrieg*. Vs Verlag Fur Sozialwissenschaften, 2007.

[4] Knight, F.H., *Risk, Uncertainty, and Profit*. Hart, Schaffner and Marx and Houghton Mifflin: Boston MA, new edition, 1921.

[5] Bonß, W., Risk. dealing with uncertainty in modern times. *Social Change Review*, **11(1)**, 2013.

[6] Ayyub, B.M., *Risk Analysis in Engineering and Economics*. Chapman & Hall/CRC: Boca Raton, Fl, 2003.

[7] International Organization for Standardization, *Risk management — Principles and guidelines (ISO 31000:2009)*. ISO: Geneva, 1st edition, 2009.

[8] Kaplan, R.S. and Garrick, B.J., On the quantitative definition of risk. *Risk Analysis*, **1(1)**, pp. 11–27, 1981.

[9] Haimes, Y.Y., On the definition of vulnerabilities in measuring risks to infrastructures. *Risk Analysis*, **26(2)**, pp. 293–296, 2006.

[10] Aven, T. and Renn, O., On risk defined as an event where the outcome is uncertain. *Journal of Risk Research*, **12(1)**, pp. 1–11, 2009.

[11] Aven, T. and Renn, O., *Risk Management and Governance*. Springer: Berlin and Heidelberg, 2010.

[12] French, S., Morton, A. & Renn, O., Special issue on risk management. *EURO Journal on Decision Processes*, **1(3-4)**, pp. 165–168, 2013.

[13] Hubbard, D.W., *The failure of risk management: Why it's broken and how to fix it*. Wiley: Hoboken and NJ, 2009.

[14] Aven, T., *Risk analysis: Assessing uncertainties beyond expected values and probabilities*. Wiley: Hoboken and NJ and USA, 2008.

[15] Purdy, G., Iso 31000:2009—setting a new standard for risk management. *Risk Analysis*, **30(6)**, pp. 881–886, 2010.

[16] Parnell, G. S., Smith, C. M. and Moxley, F. I., Intelligent adversary risk analysis: A bioterrorism risk management model. *Risk Analysis*, **30(1)**, pp. 32–48, 2010.

[17] Dillon, R. L., Liebe, R. M. and Bestafka, T., Risk-based decision making for terrorism applications. *Risk Analysis*, **29(3)**, pp. 321–335, 2009.

[18] Ezell, B.C., Bennett, S.P., Winterfeldt, D.v., Sokolowski, J. & Collins, A.J., Probabilistic risk analysis and terrorism risk. *Risk Analysis*, **30(4)**, pp. 575–589, 2010.

[19] Myerson, R., *Game theory: Analysis of conflict*. Harvard University Press: Cambridge and Mass, 1991.

[20] Dixit, A.K. & Skeath, S., *Games of strategy*. W.W. Norton: New York, 2nd edition, 2004.