A Game-Theoretic Framework for Safety and Security

Tim Zander

Vision and Fusion Laboratory
Institute for Anthropomatics
Karlsruhe Institute of Technology (KIT), Germany
tim.zander@kit.edu

Technical Report IES-2018-06

Abstract

The purpose of this paper is to describe the Framework described in [BG16] in a game theoretic way. The idea behind this is that for modelling security (i. e. the assumption of an intelligent attacker) the language of game theory seems to be a very good choice. Game theory can deal with the problem where the actions of each subject are interdependent, e.g. an attacker will change his strategy whenever a new security feature will prevent his old strategy from succeeding or a new attack seems to be more promising. Moreover, game theory has been thoroughly studied and hence changing the description language of the model, gives access to many results. Additionally, we view the Beyerer and Geisler Framework as part of security economics.

1 Introduction

When faced with the task to build or improve a system in terms of both safety and security at the same time, one has to rely as to the author's best knowledge on heuristics and intuition as there exists very little rigorous theory which can be used in practice. Moreover, we can see that there are serious logical limitations in achieving safety and security. Take for example the problem of a virus scanner on a computer. Then perfectly detecting whether a program is either malicious

or safe to execute is impossible as this would solve the halting problem (see for example [Coh87]). For the same reason it is impossible to decide in general whether a program will crash or not. This of course does not imply that we cannot improve the safety and security of a computer such that it will be good enough in practice. Another theoretical problem is the uncertainty of the NP=P-Problem, as for example elliptic curve cryptography relies on the problem that factoring in the associated group is a computational hard problem, which would brake down if NP=P would be true (with a reasonable value). Again this does not mean that we should consider cryptography as unsafe. But it implies that we have to constantly question our belief about the effectiveness of the cryptography used and its implementations.

Some security or safety incidents easily lead to monetary loss. The easiest case is just some amount of stolen money. In many cases likely-hood of a security or safety incident is quite rare, but in (not exclusively) financial terms the incident could be catastrophic (e. g. fire, flood-damage, full-loss of data). Further, often a similar risk for such incidents is shared by many players. In such cases, they could form a group and pay for each other's damages. In case the group is large enough and the risk of each player is independent, then by the law of large numbers they should only pay roughly the same amount every year. Of course in reality this is done via buying an insurance (if there is one available for the specific problem). Of course in some cases such as the loss of data or in case of stolen personal information of customers the true damage is much harder to quantify and it can be much more case specific. Also, note that there can be a significant difference in the damage done to the subject and what an attacker can gain. Think for example about a blown up automatic teller machine, where on top of the monetary loss, there are also the costs of repair. Even if there is no damage beyond the stolen item, there can still be a big gap. Take for example a famous painting, which is if stolen almost impossible to sell and hence is likely to have a far smaller monetary value for the thief. A more rigorous type cost analysis for cybercrimes can be found in [ABB$^+$13]. The paper [Her12] is investigating this type of question by asking *Why Do Nigerian Scammers Say They are From Nigeria?*.

Furthermore, this type of economic analysis has been already heavily studied in the area of internet security. As a starting point for this, see for example the survey [MA11] or the website[Uni] of Ross Anderson. In the paper [And01] the hardness of information security is evaluated; It is concluded that information security is more than the technical problem alone, many problems can be better explained

with the ideas of microeconomics such as *network externalities, asymmetric information, moral hazard, adverse selection, liability dumping and the tragedy of the commons.*

2 Summary of Beyerer and Geisler's Framework

We give a short summary of the relevant aspects of the framework for modelling safety and security introduced in [BG16]. The general idea is that there are several agents, each of them belongs to a certain role, which is either a *sources of danger*, subjects with flanks of vulnerabilities or protectors. The subjects are denoted by a set S and the set of vulnerabilities of a subject $s \in S$ are denoted as F_s. Part of *sources of dangers*, which is denoted as D, is purely stochastical, i. e. it resembles random events. The other agents have subjective views about the world and update their beliefs according to Bayes' theorem. More precisely, D splits up into wilful danger D_{w}, i. e. the attacker acts intentionally and intelligently to maximise their utility, and into unintended danger D_{U}, i. e. the result of random events. Further, D_{W} splits up into D_{WP} where the attacker wants to achieve a purpose[1] and into D_{WM} where the attacks follow only the purpose of the attack itself[2]. The source of unintended danger D_{U} splits up into D_{UC} which is the danger coming from carelessness or negligence[3] and D_{UR} purely random events[4].

3 Strategic game

In this section, we will now translate the Framework of Beyerer and Geisler to the language of game theory.

Definition 1. *A **game** is a tuple $G = (A_p, u_p)_{1 \leq p \leq N}$ where $\{1, \ldots, N\}$ is the set of players[5], A_p is the set of action of player p and $u_p : A \to \mathbb{R}$ is the utility function of player p (i. e. the payoff) where $A := \prod_{p=1}^{N} A_p$.*

[1] e. g. copy data, steal money or goods etc.
[2] e. g. vandalism
[3] e. g. inattention, breach of duties
[4] e. g. natural disasters, technical failure etc.
[5] In Beyerer2016 this are the agents

We have 5 different types of players $D_{\text{WP}}, D_{\text{WM}}, D_{\text{UC}}, D_{\text{UR}}, P$. We let the action space of each player of the game is a subset of the Cartesian product of the union of all flanks of vulnerability of each subject, i. e. $\bigcup_{s\in S} F_s$ and a subset of the union of the following; A be the space of all attacks, I be the space of all incidents and M be the space of all measures (of defence).

3.1 Action space

Each player $d \in D_{\text{W}}$ has an action space $A_d \subset \bigcup_{s\in S} F_s \times A$ (which can change over time, and he can take according to his budget $b_d(t)$). Each player $d \in D_{\text{UC}} \cup D_{\text{UR}}$ has an action space $I_d \subset \bigcup_{s\in S} F_s \times I$. The players of D_{U} play their actions at random, but for a player $d \in D_{\text{UC}}$ we assume that the probability of causing an incident (i, f) is negatively correlated to $\int_0^1 k(i, f, \beta)d\beta$ where $k(i, f, \beta)$ are the cost of d for causing an incident i on flank f with success β. A protector player $p \in P$ has an action space M_p. His goal is to minimise the threats to some subjects $S_p \subset S$ and hence $M_p \subset \bigcup_{s\in S_p} F_s \times M$. We let M_p^* be the action taken by p.

3.2 Utility functions

The utility of a player $d \in D_{\text{W}}$ for an action (a, f) with success $0 \leq \beta \leq 1$ is

$$c_{\text{effort}}(a, f) + (c_{\text{penalty}}(a, f) \Pr(\text{penalty}|a, f\beta) + g(f, \beta)).$$

We denote this utility as $u_d(\bigcup_{p\in P} M_p^*, a, f, \beta)$. Further, let $p(\beta|\bigcup_{p\in P} M_p^*, a, f)$ be the probability density of success β when executing action (a, f) while the protector players choose $\bigcup_{p\in P} M_p^*$ as their action. The idea here is that the application of a measure will decrease the probability of success of some attack or incident on some flanks. Now we let

$$\int u_d(\bigcup_{p\in P} M_p^*, a, f, \beta)p(\beta|\bigcup_{p\in P} M_p^*, a, f)d\beta$$

be the definition of the utility of the player for the action (a, f).

An action $m(f) \in M_p$ costs the player $c(m(f))$.[6] But applying similar measures to different subjects potentially reduces cost per measure (economies of scale).[7] The overall actions he can take are according to the budget given, i.e. $b(p)$. Applying the measures M_p^* will cost the protector $\sum_{m \in M_p^*} c(m(f))$. The protector's utility density function $u_p(a, f, \beta)$ for any attack or incident a with success β on flank f is

$$c(f, \beta)p(\beta| \bigcup_{p \in P} M_p^* a, f)$$

And hence the utility for the protector for an attack or incident a on flank f

$$\int_0^1 u_d(a, f, \beta)p(\beta| \bigcup_{p \in P} M_p^*, a, f)d\beta.$$

Now we are finished with our definition of the game. Each player now has an action space and a utility function which is interdependent on the actions the other players choose. We can now start reasoning about this model by applying game theoretic results. So we can conclude, if we assume that the utility function is continuous (or each player has only finitely many actions) and the action spaces are compact metric spaces, that then the game has a Nash equilibrium. If we are in doubt whether our agents will behave fully-rational, we could use other strategies as suggested in [WLB17].

What we have not dealt with is the issue that the success of an attack depends on the success of other attacks or incidents of players. Take for example the hostile takeover of a computer for the purpose of bitcoin mining, now if another attacker has also access to the very same computer and also uses it for bitcoin mining then the expected gain should only be less than half of what would be otherwise expected. Even worse if some thunderstorm destroys the computer, before any bitcoins can flow, then the gains should be zero.

[6] Note that a measure costs can change over time, such as some measures have a large initial cost but then cost almost nothing (e. g. a fence).

[7] For example, developing some piece of security software costs some fixed amount, but the price of a copy is negligible.

3.2.1 Example

A producer of security measures wants to decide whether he should develop some security measure m. He estimates the fix costs at c_m and the cost per measure applied as c_a. Now he wants to know whether he can sell enough measures (let S be the set of all tuples (m, x) measures sold for x) for a good enough price such that

$$c_d + \sum_{(m,x)\in S} c_a(m) + \sum_{(m,x)\in S} x \geq 0. \tag{3.1}$$

For that we have to determine, if there exists new games where measure m is available for to all the protectors P for a certain price x, enough protectors are going to apply the measure m for their price such that Equation (3.1) is full-filled. Note that the price of the producer is not necessarily the price of the protector. Take for example a big fence with some barbed wire, it may not be allowed to install (so we may assume that the costs for the protector would be infinite). Or the protector has to stop the production line of his company in order to install the measure, which will then of course result in additional costs.

4 Bayesian game

We extend the above game to follow the rules of Bayesian game. As in the above game, it is assumed that all the players will have full knowledge about their own and the others players' action spaces and utility functions. In a Bayesian game on the other hand, the player have only incomplete information available, but have beliefs about the action spaces and utility functions of the other players. So lets first formally define what a Bayesian game is.

Definition 2. *A **Bayesian game** is a tuple* $\Gamma = ((T_i, \mathcal{T}_i), A_i, u_i, p)_{1\leq i\leq N}$ *with* $A := \prod_{i=1}^{N} A_i$ *and* $T := \prod_{i=1}^{N} T_i$ *where*

- $\{1, \ldots, N\}$ *is the set of players;*
- $(T_i, \mathcal{T}_i)$ *is a measurable space*[8]*, where* T_i *is the* i*'s non-empty type space. Further we let* $\mathcal{T} = \otimes_{i=1}^{N} \mathcal{T}_i$*;*[9]

[8] Note that, if T_i is discrete, then we may ignore $\mathcal{T}_i$, as in this case it is the power set.
[9] The product σ-algebra.

- *A_i is the space of actions of player i, a non-empty metric space*[10]
- *$u_i : T \times A \to \mathbb{R}$ is i's bounded utility function, which is measurable by the σ-algebra $\big(\mathcal{T} \otimes [\otimes_{i=1}^{N} \mathcal{B}(A_i)], \mathcal{B}(\mathbb{R})\big)$;*
- *p is a probability measure on $(T, \mathcal{T})$ which denotes the common prior over the type profiles.*

Now if we want to model the Beyerer-Geisler framework as a Bayesian game we need to define the type space T. The type space T_p for player p in P consist of tuple of functions which map the objective costs and objective probabilities to the player's subjective view. So it consists of functions $\nu_{p,c(m)}$ which maps the objective cost of some measure m (on flank f of subject s) to the subjective cost of p. Also, some function ν_p which maps $c(f, \beta)$ to the subjective cost of p. Then some function π_p which maps the probabilities $p(\beta|\bigcup_{j\in P} M_j^*, i, f)$,$\Pr(i|\bigcup_{j\in P} M_j^*, f)$ $p(\beta|\bigcup_{j\in P} M_j^*, a, s, f)$ and $\Pr(a|\bigcup_{j\in P} M_j^*, s, f)$ to those subjectively assumed by p. We may also assume that the player p does only know his type up to some probability measure X_p on T_p Together this leads to some new subjective utility function $R_{p,\bigcup_{p\in P} M_p^*}(\nu_{p,c(m)}, \nu_p, \pi_p)$.

The type space for player d in D_W consists of a map $\nu_{d,c(a)}$ which maps the objective costs $c_{\text{effort}}(a, s, f)$, $(c_{\text{penalty}}(a, s, f)$ and $g(s, f, \beta))$ to the subjective costs of d. Also, there is a function π_d which maps $\Pr(\text{penalty}|a, s, f, \beta)$ and $p(\beta|a, s, f)$. to the subjective probabilities assumed by d. Again we may assume that the player d only knows his type up to some probability measure X_i.

Having set up the framework of Beyerer and Geisler like this we can apply the results of [CNM14] and know that the game (under some minor continuity assumptions or in the discrete case) has a Bayes-Nash-equilibrium.

5 Introducing temporal dynamics

We introduce some temporal dynamics now. So assume that the finite time horizon is given by $\mathfrak{T} = \{0, 1, \ldots, k\}$. Now the state of the system has three components

[10] If A_i is finite, we may assume that it is a set without any additional structure.

at time i: The type space $(T_i, \mathcal{T}_i)$, the common knowledge (prior) $\mathfrak{p}_i$ and the private assumptions $X_i^p = (\nu_{p,c(m)}^i, \nu_p^i, \pi_p^i)$ of player p. We denote the action space at time i of each player p A_i^p. Note that the action space of an protector player depends on $b_p(t)$. We may also assume that rather than having a new budget every round a protector player has fixed amount of money for some fixed number of rounds, say $t_{k'}$. So the action a player can take depends on the actions already taken.

For a player in $d \in D_W$, the actions the player can take are again limited by his budget $b_d(t)$ plus sometimes he can reinvest the eventual gains. Again he may assume that d has some fixed amount of money for some fixed number of rounds, say $t_{k'}$. So the action a player can take depends on the actions already taken.

So in this case if the subjective assumption of the players match the private assumptions, hence if our model is a non-Bayesian game, we can think of this temporal dynamic as an extensive-form game. When the game is finite, we can think of such a game in terms of a game tree. This game has again Nash-equilibria, but in this context they can be unrealistic. A solution to this problem are subgame perfect equilibria, which compared to ordinary Nash-equilibria have the additional property, that they are also equilibria for every subgame. Their existence can be shown via backward induction.

In terms of the Bayesian game model, we can go over to sequential Bayesian game. Again the concept of Bayes-Nash-equilibrium leads to unrealistic equilibria, but in case our game is finite, we can show the existence of perfect Bayesian equilibria which overcome this issue.

Another idea would be to use the framework of [OTT17]. This framework assumes that the common knowledge evolves as

$$P_{i+1} = f_i(P_i, A_i, W_i^C),$$

where f_i is function of common knowledge and W_i^C is a random variable which represents the randomness of the evolution. Then we assume that a player observes

$$Y_i^p = l_i^p(X_i^p, A_i, W_i^p),$$

where W_i^C is again a random variable which represents some random noise. We further assume that f_i^p is common knowledge among all players.[11] There are no

[11]This just means that if the players would swap their positions they would observe the same.

hard general results for this type of framework, but it is at least suspected that there exist equilibria solutions for this framework in general.

6 Conclusion and future work

We redefined the framework of Beyerer and Geisler in terms of game theory. This opens access to a deep theory for example the existence of the Nash equilibria. This immediately raises the question on the suboptimalities of these stability points of the game in terms of the social optimum (cf. [GCC08]). Moreover, many other results of game theory seem to be relevant as well, such as the study of how humans behave in these kind of strategical interactions. What also can be seen is that the task of the protector player can be a very hard problem. Not only do they have to think about their own flanks of vulnerabilities, but they also have to get a good idea of the adversary's capabilities and their motivation, e. g. their utility function.

We raise the question whether we can improve our predicting abilities of different security polices. Take for example the choice of password polices. Could we have predicted that the policies of forcing to change the passwords regularly will lead to questionable security (cf. [ZMR10]). Another question we can ask is the model's ability to predict what happens for the problem, when the protector does not have to bear the cost of failure. This is for example the case for proprietary software, the protector is the copy-right holder but the one who will suffer first is the user (cf. [MA11]).

This leads to the question of how the model we defined can be used apart for a purely theoretical quality analysis. One idea we want to investigate in the future is the question whether multi-agent simulations with our model in mind will lead to good predictions in terms of security engineering but also in terms of what political decisions such as the European General Data Protection Regulation will achieve for the personal data protection of its citizens.

Bibliography

[ABB+13] Ross Anderson, Chris Barton, Rainer Böhme, Richard Clayton, Michel JG Van Eeten, Michael Levi, Tyler Moore, and Stefan Savage. Measuring the cost

of cybercrime. In *The economics of information security and privacy*, pages 265–300. Springer, 2013.

[And01] Ross Anderson. Why information security is hard-an economic perspective. In *Computer security applications conference, 2001. acsac 2001. proceedings 17th annual*, pages 358–365. IEEE, 2001.

[BG16] Jürgen Beyerer and Jürgen Geisler. A framework for a uniform quantitative description of risk with respect to safety and security. *European Journal for Security Research*, 1(2):135–150, 10 2016.

[CNM14] Oriol Carbonell-Nicolau and Richard McLean. On the existence of nash equilibrium in bayesian games. Departmental working papers, Rutgers University, Department of Economics, 2014.

[Coh87] Fred Cohen. Computer viruses. *Comput. Secur.*, 6(1):22–35, February 1987.

[GCC08] Jens Grossklags, Nicolas Christin, and John Chuang. Secure or insure?: a game-theoretic analysis of information security games. In *Proceedings of the 17th international conference on World Wide Web*, pages 209–218. ACM, 2008.

[Her12] Cormac Herley. Why do nigerian scammers say they are from nigeria? *WEIS*, June 2012.

[MA11] Tyler Moore and Ross Anderson. Economics and internet security: A survey of recent analytical, empirical, and behavioral research. 2011.

[OTT17] Yi Ouyang, Hamidreza Tavafoghi, and Demosthenis Teneketzis. Dynamic games with asymmetric information: Common information based perfect bayesian equilibria and sequential decomposition. *IEEE Transactions on Automatic Control*, 62(1):222–237, 2017.

[Uni] Ross Anderson Cambridge University.

[WLB17] James R Wright and Kevin Leyton-Brown. Predicting human behavior in unrepeated, simultaneous-move games. *Games and Economic Behavior*, 106:16–37, 2017.

[ZMR10] Yinqian Zhang, Fabian Monrose, and Michael K Reiter. The security of modern password expiration: An algorithmic framework and empirical analysis. In *Proceedings of the 17th ACM conference on Computer and communications security*, pages 176–186. ACM, 2010.