A Decision Making Model of Influencing Behavior in Information Security

Iryna Yevseyeva, Charles Morisset, Thomas Groß, and Aad van Moorsel

Centre for Cybercrime and Computer Security
School of Computing Science, Newcastle University
Newcastle upon Tyne NE1 7RU, UK
firstname.lastname@newcastle.ac.uk

Abstract. Information security decisions typically involve a trade-off between security and productivity. In practical settings, it is often the human user who is best positioned to make this trade-off decision, or in fact has a right to make its own decision (such as in the case of 'bring your own device'), although it may be responsibility of a company security manager to influence employees choices. One of the practical ways to model human decision making is with multi-criteria decision analysis, which we use here for modeling security choices. The proposed decision making model facilitates quantitative analysis of influencing information security behavior by capturing the criteria affecting the choice and their importance to the decision maker. Within this model, we will characterize the optimal modification of the criteria values, taking into account that not all criteria can be changed. We show how subtle defaults influence the choice of the decision maker and calculate their impact. We apply our model to derive optimal policies for the case study of a public Wi-Fi network selection, in which the graphical user interface aims to influence the user to a particular security behavior.

1 Introduction

People continuously make information security decisions: should I use this wireless, should I put this person's USB stick in my laptop, how do I choose and memorize passwords? Almost always, the decision involves a trade-off between security and other concerns, such as being able to complete an important task or being able to easily do something that otherwise could be cumbersome. The decisions are often complex, with several objectives to be considered simultaneously, and the optimal decision may very much depend on the specific situation: while using a stranger's USB stick is not advisable, the importance of the job to be completed and/or knowledge about the owner of the USB stick may make it advisable to put the USB stick in one's laptop, despite the associated information security risks.

In situations such as above, a simple compliance policy (such as, not to allow USB sticks at all) would be suboptimal. Instead, one would want to allow some freedom for the owner of the laptop to decide the best course of action.

A. Horváth and K. Wolter (Eds.): EPEW 2014, LNCS 8721, pp. 194–208, 2014.

In general terms, unless one can specify a compliance policy that is optimal under all possible circumstances (which is a rare real-world case), there is room for improvement by allowing the user to make the final decision. There exist other situations, in which the user should play a role in the security decision making. For instance, in case of BYOD (bring your own device) [7], where the device owner uses their own device for work-related activities, the fact that the user owns the device puts certain restrictions on what the employer can decide without the owner's input. However, an employer might still want to influence the decisions of its employees, since the employer is impacted by these decisions.

In all these situations the end user is involved in the information security decision making, and is in fact responsible for the final choice. Then, and this is key for this paper, it may be advisable that service providers (telecoms, online banks), device vendors, employers, or other parties are able to *influence* the decision making, without restricting the end user. In the literature, this is often referred to as nudging [22] implemented widely in healthcare and social policies, see e.g. [21]. Nudging leaves the choice with the user, but aims to influence the decision so that the user is more likely to make a beneficial decision, e.g., by presenting choices in a particular manner that aims to impact the choice a person ends up making. There are many aspects to nudging that deserve discussion, but, in this paper, we do not debate the specific approach, but aim to derive results for influencing in general.

In [17] a first formalization of the concept of influencing was provided assuring it is as general as possible, but at the same time is intuitive and useful. Earlier, Heilmann [11] presented schematically the nudge success conditions from perspective of influencing autonomous system, also called System 1 (and not reflective, also called System 2) [13], and showed the difference of these conditions for different types of nudges with respect to taxonomy of Bovens [4].

In this paper, we provide a *model for influencing human decision making in security contexts*. A model aim to analyze users' decisions and behavior in order to be able to define better security policies and procedures from both an employer and its employees points of view. In particular, it gives an opportunity to an employer to influence decisions of its employees; however, leaving the final choice and responsibility for the decision to the employee who made it.

We believe such a model is necessary to enable a solid quantitative evaluation of influence. In particular, we want to be able to apply mathematical optimization to decision making as well as to the decision on how to influence, and for that we need a rigorous underpinning and understanding of the problem at hand.

Finally, we want to be able to evaluate the level of success of influencing behaviors, be it experimentally or theoretically–again, a formal model allows us to define the experimental or theoretical setting under which we carry out the evaluation. This paper will not reach all these goals, but provide the underlying quantitative model for human decision making evaluation for security decisions.

Our model is based on a well-known practical approach to modeling human decision making, *multicriteria decision analysis*, see e.g. [2], in particular, on *multiattribute utility theory* [15] presented in Section 2. We assume that such a model can be used both for the *decision maker* (e.g., the employee of a company), and for the *stakeholder* (e.g., the company). Given a set of alternatives evaluated on a set of criteria, we can define a policy that represents the choice of optimal decisions by the decision maker, and we can calculate the optimal modification of these criteria with respect to the stakeholder. A particular contribution of this work is to model the freedom of choice left by the stakeholder to the decision maker by considering that only a subset of all criteria are modifiable. We illustrate in Section 4 the case, where a stakeholder is effectively unable to influence the decision maker.

We illustrate each stage of our model and its merits using a public Wi-Fi selection scenario taken from [23] in each section. In the Wi-Fi example, a device user decides between networks and the device presents choices so as to influence the decision of the device user. In this case, the decision maker represents the device user and the stakeholder represents the company of the user. We show how changing presentation of some Wi-Fi's may alter the choice of decision maker. However, the approach is designed generally enough to be applied to other case studies, e.g. for choosing among access control policies [18]. Finally, Section 5 discusses possible extensions of the framework, in particular, considering influencing populations.

2 Decision Making

In order to model human decision making and to evaluate the different alternatives for a decision maker, we consider Multi-Criteria Decision Analysis (MCDA). MCDA is particularly useful in situations, where alternatives are evaluated on multiple, often conflicting, criteria, and in search of solutions that represent the best trade-off(s) between these criteria. In information security, this trade-off is usually between security and productivity/usability, for instance, a decision maker has to select between a more secure network and a faster one.

When compared to other approaches to model security decision making, e.g. through Markov Decision Process and reward models [3] or using the experience, e.g. by reinforcement learning [20], MCDA provides transparency to the process of making decisions and illustrates explicitly how trading-off between criteria is obtained. Transparency of the decision making process is desirable by both decision makers and stakeholders, who are interested in seeing how their preferences with respect to criteria are considered within a model. Moreover, MCDA allows for possible behavioral biases to be taken into account within a model, e.g. in a similar way as in [14].

For selecting a set of criteria that influence security decisions, it may be advisable to look at attributes related to technology, to management, to economy, to culture and to personal preferences. However, the general MCDA recommendation is to consider a set of criteria most relevant to a particular problem to be solved [10] from [15].

Here, by making a decision, we assume choice of an alternative among available ones. A decision maker is responsible for selecting an alternative a. We write $\mathcal{A}$ for the set of alternatives available to the decision maker.

In MCDA, alternatives are evaluated and compared using a set of criteria $\mathcal{G}$, such that each criterion should be either minimized or maximized (the direction of optimization). Each criterion comes with a scale, in which alternatives can be compared. Typical scales include real numbers, intervals, ratios, binary or verbal values (qualitative descriptions), which are ordered with respect to the optimization direction. Each criterion $g \in \mathcal{G}$ is, therefore, associated with a scale $\mathcal{K}_g$, and we write $g^{\min} \in \mathcal{K}_g$ and $g^{\max} \in \mathcal{K}_g$ for the minimal and maximal values of g, respectively. We write $\mathcal{K} = \bigcup_{g \in \mathcal{G}} \mathcal{K}_g$ for the set of all possible scales, and without any loss of generality, we assume that all criteria are maximized (minimized criteria can simply be multiplied by -1).

Each alternative is evaluated on each criterion $g \in \mathcal{G}$ by means of an evaluation function $\sigma_g : \mathcal{A} \to \mathcal{K}_g$. We write Σ_g for all possible σ_g functions, and $\Sigma_{\mathcal{G}} = \prod_{g \in \mathcal{G}} \Sigma_g$ for the cartesian product of all criteria evaluation functions. When no confusion can arise, we write $\sigma = (\sigma_{g_1}, \ldots, \sigma_{g_n})$ for a vector of evaluation functions, and $\sigma[g]$ for the evaluation function of σ corresponding to the criterion g.

We now present the basics of MCDA and Multi-Attribute Utility Theory, in particular. We then detail how to define the policy of a decision maker, and we illustrate this approach for selection of a public Wi-Fi case study.

2.1 Multi-Attribute Utility Theory

Multi-Attribute Utility Theory (MAUT) [15] is an MCDA approach, which assumes that decision makers aim to maximize their implicit utility function. MAUT is a compensatory technique, since it allows smaller values on a subset of criteria to be compensated by a large value on a single criterion, and is based on expected utility theory with some strong technical assumptions related to comparability, transitivity, continuity, and independence of outcomes (that assumes independence of criteria). MAUT is attractive because of its sound theoretical foundations (based on expected utility theory), its non-monetary nature and its applicability to be used as a basis for comparison of new, not yet considered alternatives with the same utility function constructed for the same decision maker. In addition, its natural approach to modeling risk behavior is particularly attractive for designing security decisions, where risk attitude of decision makers plays crucial role in their decision patterns.

The global utility of an alternative is obtained by aggregating individual criteria values amplified by criteria weights for all criteria. However, before aggregation, these criteria values must be *normalized*, in order to provide a fair basis for comparison. A normalization function, which in MAUT corresponds to *marginal utility function*, is a function $n_g : \mathcal{K}_g \to [0, 1]$. This function can change from one decision maker to another, thus, encoding some notion of preference/meaning interpretation.

In addition, preferences can be encoded using criteria weights, which in MAUT represent trade-offs between criteria. Here, a weight shows the relative importance of the criterion, when compared to other criteria. In particular, it defines how many units of one criterion can be traded-off for a unit of another criterion.

Here, we assume deterministic criteria weights defined by the decision maker with a criteria function $w : \mathcal{G} \rightarrow [0, 1]$ such that $\sum_{w(g) \in W} w(g) = 1$.

Determining weights explicitly may be difficult for decision makers. It may be cognitively hard to quantify weights also due to the meaning of weights may not be straightforward and even differ in different MCDA methods [2].

We can now define the notion of MAUT model.

Definition 1. *A MAUT model is a tuple $M = (\mathcal{A}, \mathcal{G}, \varSigma_{\mathcal{G}}, n, w)$, where $\mathcal{A}$ is a set of alternatives, $\mathcal{G}$ is a set of criteria, $\varSigma_{\mathcal{G}}$ is a set of criteria evaluation functions, n is a set of normalization functions with $n_g : \mathcal{K}_g \rightarrow [0, 1]$ for each g, and $w : \mathcal{G} \rightarrow [0, 1]$ is a weights function, such that $\sum_{w(g) \in W} w(g) = 1$.*

After mapping all criteria utilities to their scales, normalizing them and defining weights, the alternatives can be evaluated. For aggregating marginal criteria utilities for each alternative some form of aggregation function should be used, e.g. multiplicative, additive or some combination of both is usually applied. When compared to additive aggregation function, which allows some criteria for alternatives to be of zero value, multiplicative aggregation function requires presence of non-zero values for all criteria to make alternative useful.

For now, we introduce one of simplest forms of aggregating evaluations on all criteria values for each alternative, weighted sum, which we will also use for the Wi-Fi case study:

Definition 2 (Utility function). *Given a model $M = (\mathcal{A}, \mathcal{G}, \varSigma_{\mathcal{G}}, n, w)$, the utility of an alternative $a \in \mathcal{A}$, is defined as:*

$$v(a, w, \sigma) = \sum_{g \in \mathcal{G}} w(g) \cdot n_g(\sigma_g(a)).$$

Note that for the sake of simplicity, we consider that the normalization function is unique for all criteria, and therefore we do not pass it as an argument of v.

We now assume that decision makers base their decision making process using a MAUT model[1]. Given a vector of evaluation functions σ, and a weights function w, the policy of a decision maker is defined as:

$$\pi(w, \sigma) = \arg\max_{a \in \mathcal{A}} v(a, w, \sigma).$$

Note that in order for a decision maker to be deterministic, we assume the existence of an arbitrary ordering over alternatives, so that if there are several alternatives maximizing the utility function, the decision maker selects the highest one according to that ordering.

[1] We are aware of strong assumptions of MAUT and biases from rational behavior of the decision makers studied, e.g. by Kahneman and Tversky [13], [14], and here establish a basic model for influencing human decision making in security context also to initiate investigation of these biases.

2.2 Case Study: Selection of a Wi-Fi Network

As a case study, let us consider an example of influencing a choice of a publicly available wireless network (Wi-Fi). The dangers of choosing non-secure Wi-Fi are well documented: it exposes device and transmitted data to increased chances of spoofing and man-in-the-middle attacks [1], [5], and new attacks appear regularly.

For instance, recently, it was reported that the penetration testing tool BDF-Proxy (BackdoorFactory Proxy), which acts as a proxy for network communication, has the capability to infect any binary executable download the user makes with a Metasploit malware [16]. Thus, it can compromise the user's device with malicious software and gain control over the device. This attack is particularly problematic for untrusted Wi-Fi's as the Wi-Fi router can engage in ARP (Address Resolution Protocol) spoofing (i.e., by manipulating the lowest-level address resolution to make the user's client go through the proxy without the user's knowledge: The Wi-Fi can make the client fall for the trap without the user noticing anything.

Ability to influence choice of trusted Wi-Fi is of special interest in the context of recent consumerization of IT trend [24] and BYOD, in particular, since employees work on their own devices and define security protection of their devices by themselves, thus, potentially, exposing sensitive information [19]. In general, over one billion workers will work remotely by 2015, over a third of the total worldwide workforce [12]. A company that allows BYOD may want to influence the employee so that the trade-off decision between security and productivity is done in the company's interest. Alternatively, users may want to have influencing software on their phone to assist in making the information security decisions for work as well as home use.

Influencing was earlier introduced in the security context of Wi-Fi selection in [23]. There, the focus is on introducing the user interface design nudges, and on evaluating them with a user group. Here, we want to model human behavior further when modifying context of decision making by introducing an affect influencing factor, color, and by computing impact of such or another modification of the context on the decision to be made. In particular, a traffic light effect [8] is used with red-green colors (and associated emotions and meanings), which was also applied for framing choices to nudge individuals away from privacy-invasive applications in [6].

Let us consider a user in a coffee-shop having to choose between two different public wireless networks $\mathcal{A} = \{s, f\}$: s is a secure Wi-Fi with weak signal; f is a Wi-Fi of the coffee shop, with strong signal, but not necessarily safe. We want to illustrate with this example the trade-off between security and productivity/usability, and therefore consider the set of criteria $\mathcal{G} = \{t, r, l\}$, indicating the trust of the network, its strength and color in which its name is drawn, respectively. For the sake of simplicity, we assume that the scales for the trust and strength criteria are defined as $\mathcal{K}_t = \mathcal{K}_r = \{0, 1, 2\}$ (the higher the better). For the color criterion, a scale is defined as $\mathcal{K}_l = \{R, N, G\}$, corresponding to red, neutral and green colors of paint used for drawing names of networks,

Table 1. Decision matrix for $\sigma = [s \mapsto (1, 1, N), f \mapsto (0, 2, N)]$

	criteria			
	trust	strength	color	
	$\{0, 1, 2\}$	$\{0, 1, 2\}$	$\{R, N, G\}$	(scale)
	$t \rightarrow \max$	$r \rightarrow \max$	$l \rightarrow \max$	(direction)
	0.5	0.3	0.2	(weights)
alternative				
s	1	1	N	
f	0	2	N	
$\pi(w, \sigma)$		f		

respectively. This categorical scale can be mapped into a scale of quantitative values $\mathcal{K}_l = \{0, 0.5, 1\}$, taking into account traffic light similar effect, with red color associated with danger, green color – with no danger, and neutral color, e.g. white, with no special affect on users (when compared to a standard amber color with attention bringing effect).

Note that here, we consider a simple and abstract notion of trust, and, in practice, this notion can be defined using the presence of Wi-Fi network providers in a white list predefined by security officer or system administrator of the company or by the employee him-/herself. More sophisticated evaluation of 'trust' criterion may take into account other aspects, e.g. current location of an employee [9].

Finally, the decision maker has to define the criteria weights $w = (0.5; 0.3; 0.2)$, meaning that connecting to a trusted Wi-Fi is more important for the decision maker than choosing a Wi-Fi with strong signal. The color of the presented name of a Wi-Fi is less significant for the decision maker than the two other criteria.

In the following, for the sake of compactness, we write $\sigma = [s \mapsto (v_1, v_2, v_3), f \mapsto (v_4, v_5, v_6)]$, associating s Wi-Fi with a trust of v_1, a strength of v_2 and a color of v_3 values, respectively; and f Wi-Fi with a trust of v_4, a strength of v_5 and a color of v_6 values, respectively. Table 1 represents the traditional decision matrix [2] for a decision maker, evaluation of a set of alternatives on a set of criteria. Assuming that a decision maker uses a linear normalization function of the following form:

$$n_g(\sigma) = \frac{g - g^{min}}{g^{max} - g^{min}}, \tag{1}$$

we can calculate the utility for each alternative as follows:

$$v(s, w, \sigma) = 0.5 * 0.5 + 0.3 * 0.5 + 0.2 * 0.5 = 0.5$$
$$v(f, w, \sigma) = 0.5 * 0 + 0.3 * 1 + 0.2 * 0.5 = 0.4.$$

From these calculations it follows that the decision maker selects $\pi(w, \sigma) = s$.

3 Decision Evaluation

3.1 Impact

To be able to measure the efficiency of an alternative, we introduce an *impact function* such that, given an alternative a, a weight function w and evaluation functions σ, $\rho(a, w, \sigma)$ represents the impact of criteria weights and criteria evaluations of alternative on selection of that alternative as the final choice of the decision maker. In the rest of the paper, we consider that the impact function intuitively represents a benefit for the system, and as such, the aim of a stakeholder is to maximize the impact, i.e., a higher impact is 'better'. Note, the impact function should be seen as an ideal valuation of the possible alternatives, and as a way to evaluate the behavior of the decision makers, rather than as a way to define the behavior of the decision makers.

In general, this function can be defined in many different ways (for instance, through an access control policy stating which alternatives are secure [18]). We propose here to define it using a MAUT model, which is however slightly different from the one defined above. The impact function can be defined directly as the utility function v of M. However, in the context of information security, we want to clearly distinguish the alternatives, so that there are 'good' and 'bad' alternatives. Hence, given an alternative a, a weight function w and a set of criteria evaluation functions σ, we define the impact function as:

$$\rho(a, w, \sigma) = \begin{cases} 1 & \text{if } v(a, w, \sigma) \geq v(a', w_i, \sigma) \text{ for any } a', \\ 0 & \text{otherwise.} \end{cases}$$

In other words, an alternative has an impact if, and only if, it is maximal according to the utility function. Note that more complex impact functions can be considered, for instance, when different levels of security can be defined.

3.2 Utility Function Parameters

To evaluate the efficiency of a decision made by a decision maker, a stakeholder may compare it to his own choice or a choice of an 'ideal' decision maker from a company perspective in the same situation. Four cases are possible here: In ideal case, the stakeholder would wish the decision maker behaving in an optimal way from the stakeholder's point of view. This would mean the decision maker (user / employee) having the same with stakeholder (or company) preferences, or the same weight function $w_c = w_u$, where w_c is a stakeholder's weight function and w_u is a user's weight function, and at the same time having the same set of criteria evaluation functions: $\sigma_c = \sigma_u$, where σ_c is a stakeholder's set of criteria functions and σ_u is a set of user's criteria evaluation functions.

However, the reality might be different with the most general case with both sets of evaluation functions $\sigma_c \neq \sigma_u$ and weighting functions $w_c \neq w_u$ being different for a stakeholder and a decision maker. The two special cases with either different weights or different criteria evaluation functions will be considered below.

3.3 Evaluation in Case Study: Selection of a Wi-Fi Network

To illustrate decision evaluation scenarios for the case of public Wi-Fi selection, let us consider the stakeholder and the decision maker having the same weight functions $w_c = w_u = (0.5; 0.3; 0.2)$. However, they have different evaluation functions: $\sigma_c = [s \mapsto (1, 1, N), f \mapsto (0, 2, N)]$ for stakeholder and $\sigma_u = [s \mapsto (1, 1, N), f \mapsto (1, 2, N)]$ for the decision maker. Indeed, the decision maker considers the alternative f as being more trusted, with $\sigma_u[t](f) = 1$, when compared to the company, which assigns to it a smaller trust value with $\sigma_c[t](f) = 0$. This small difference results in the different utilities of the alternatives $v(s, w_u, \sigma_u) = 0.5$ and $v(f, w_u, \sigma_u) = 0.65$, and leads to the decision maker choosing $f = \pi(w_u, \sigma_u)$, while $\rho(f, w_c, \sigma_c) = 0$, meaning that the decision maker selects an alternative that is suboptimal for the stakeholder.

We may also consider another case of the company and the user having the same set of criteria evaluation functions $\sigma_c = \sigma_u$, but different preferences with respect to criteria weights. For instance, the stakeholder considers trust being more important $w_c(t) = 0.5$, when compared to the decision maker $w_u(t) = 0.3$. They may also have different opinions about importance of the strength of the Wi-Fi signal: the stakeholder assumes it is as less important $w_c(r) = 0.4$, when compared to the decision maker $w_u(r) = 0.6$. But they agree on color being not very important $w_c(l) = w_u(l) = 0.1$. Here, again $\pi(w_u, \sigma_u) = f$, while $\rho(f, w_c, \sigma_c) = 0$.

4 Influencing Decisions

As said in the introduction section, in the BYOD context, companies allow their employees to use personal devices for work (or company devices for personal purposes), and the border between personal and company data becomes blurred. In such situations, companies may try to take some control over personal devices for better protection of their data. Applying strong security policies for such personal devices may meet opposition reaction from their employees, since employees ownership perception of devices will be disturbed, which may push employees towards overriding such security policies. Therefore, companies must search for 'softer' ways of influencing their employees behavior.

In this work, we suggest a 'soft' strategy for stakeholders to assist in security decisions by their employees with limited changes to the information taken into account by the decision maker, based on the idea that even small changes can influence final choices of decision makers [13], [14]. Such an approach was considered widely for health and social solutions, see e.g. [21], [22], and recently studied in the context of security and privacy decision making [6].

Next, we examine an example of a company adopted BYOD strategy or stakeholder, which wants to protect its employees, users of devices, from non secure behavior and emphasize safer choices for them. Note, that we assume here a 'good' stakeholder, who wants to help and protect a user in a paternalistic way, and exclude a 'bad' influencer, for instance, aiming to attack users and manipulate their choices motivated by 'bad' incentives, leaving this special case as a future work. As our working case study for demonstrating influencing effect, we keep selection of a Wi-Fi to connect to in a public place among several available ones. We consider when influencing may be beneficial to both a stakeholder and a user, and how it may be performed, assuming MAUT model as a basis for human decision making.

4.1 Influence

Given a MAUT model $M = (\mathcal{A}, \mathcal{G}, \Sigma_{\mathcal{G}}, n, w)$, we can consider ways, in which a stakeholder may influence choices of a decision maker. By definition of the model, there are two ways to affect the result of the model evaluation: either by affecting a weighting function, and corresponding set of weights, or by affecting a set of criteria evaluation functions, and corresponding set of criteria values for alternatives.

Influencing weighting of criteria means influencing implicit trade-off preferences of decision makers with respect to different criteria. In principle, this approach may be efficient, but, in practice, it is time-consuming, since it requires training and education of users and their subsequent conscious reflection on the issues they were taught. For instance, for security decisions, it would require training sessions on the security policy of the company to increase employees awareness of risks; their education on the security issues related to the policy of their company and on possible consequences of such decisions for them and their company; and promoting a security culture, e.g., with rewards for secure behavior. These are efficient, but long-term approaches, which require time and involve user awareness and conscious decision making. Moreover, while users may be aware and intend to behave securely, these intentions do not always translate into actual behavior.

Therefore, an alternative and/or complementary approach would be to try to influence the behavior of decision makers directly at the moment of the decision making. This approach would involve changing values for some criteria. Having possibility to change all criteria would be ideal for a stakeholder. However, there are different reasons why it may not be possible in most cases; to name a few: it may not be legal or ethical to change values for some criteria or too costly for the company to do it. However, a stakeholder may still be able to change values for some 'modifiable' criteria via a set of evaluation functions, assuming the values for the rest of criteria are non-changeable.

Given a set of criteria $\mathcal{G}$, we consider a subset of 'modifiable' criteria $\mathcal{M} \subseteq \mathcal{G}$, for which stakeholder can change criteria values. The exact definition of this subset depends of course on the context, but intuitively, it corresponds to the aspects taken into account by the decision maker that are controlled by the stakeholder. Given a vector of evaluation functions σ, we define the set of possible modified functions as:

$$P_{\mathcal{M}}(\sigma) = \{\sigma' \mid \forall g \in (\mathcal{G} \setminus \mathcal{M})\ \sigma'[g] = \sigma[g]\}.$$

Table 2. Impact of all modifications to the color criterion for initial alternatives evaluations $\sigma_u = [s \mapsto (1, 1, N), f \mapsto (1, 2, N)]$ and criteria weights $w_u = (0.3; 0.5; 0.2)$ of the decision maker with $\rho(f, w_c, \sigma_c) = 1$ and $\rho(s, w_c, \sigma_c) = 0$

σ_{xy}	$v(s, w_u, \sigma_{xy})$	$v(f, w_u, \sigma_{xy})$	$a = \pi(w_u, \sigma_{xy})$	$\rho(a, w_c, \sigma_c)$
σ_{NN}	0.5	0.6	f	0
σ_{NR}	0.5	0.5	f	0
σ_{NG}	0.5	0.7	f	0
σ_{GN}	0.6	0.6	f	0
σ_{RN}	0.4	0.6	f	0
$\boldsymbol{\sigma_{\mathbf{GR}}}$	**0.6**	**0.5**	**s**	**1**
σ_{RG}	0.4	0.7	f	0
σ_{RR}	0.4	0.5	f	0
σ_{GG}	0.6	0.7	f	0

In general, more complex restrictions on $P_{\mathcal{M}}$ can be defined, for instance, reflecting an incremental change in the values of criteria (e.g., the value of a criterion can only be incremented or decremented by a given factor).

Hence, we assume that there is an *influence*, if and only if, the decision maker would behave differently without being influenced. The raw impact of an influence can be measured in a differential way: given a vector of evaluation functions σ and a weight function w, we say that the decision maker was influenced whenever $\pi(w, \sigma) \neq \pi(w, \sigma')$. Note that the set of alternatives $\mathcal{A}$ does not change with the application of criteria evaluation modifications. In other words, influencing a decision maker does not change the set of alternatives available to the decision maker.

We are now in position to define the optimal modification possible by a stakeholder over a decision maker.

Definition 3. *Given a stakeholder with a weights function w_c and a vector of evaluation functions σ_c, and a decision maker with a weights function w_u and a vector of evaluation functions σ_u, the optimal vector of modified evaluation functions for the decision maker is given by:*

$$\mathsf{opt}(w_u, w_c, \sigma_u, \sigma_c) = \arg \max_{\sigma'_u \in P_{\mathcal{M}}(\sigma_u)} \rho(\pi(w_u, \sigma'_u), w_c, \sigma_c).$$

4.2 Influence in Case Study: Selection of a Wi-Fi Network

Let us consider a subset of modifiable criteria $\mathcal{M} = \{l\}$, i.e., only the color, in which a network is displayed, can be modified. We now illustrate the optimal modification of the criteria evaluations opt for the influencing strategy applied by the stakeholder. Having a set of criteria weights $w_c = (0.5; 0.4; 0.1)$ and criteria evaluations of alternatives $\sigma_c = [s \mapsto (1, 1, N), f \mapsto (0, 2, N)]$, we have $\rho(s, w_c, \sigma_c) = 1$ and $\rho(f, w_c, \sigma_c) = 0$. In other words, the stakeholder wants to influence the decision maker towards selecting a more secure Wi-Fi s.

Let us consider now that the decision maker has different criteria weights $w_u = (0.3; 0.5; 0.2)$ and different criteria evaluations $\sigma_u = [s \mapsto (1, 1, N), f \mapsto (1, 2, N)]$, which leads to the decision maker choosing a faster network $\pi(w_u, \sigma_u) = f$. Since $\mathcal{M} = \{l\}$, only the color criterion value can be modified. Table 2 details all the possible cases, where, for the sake of compactness, we write σ_{xy} for the evaluation function $\sigma_{xy} = [s \mapsto (1, 1, x), f \mapsto (1, 2, y)]$. We also consider that when s and f have the same value, the decision maker selects f by default.

In other words, $\mathsf{opt}(w_u, \sigma_u, w_c, \sigma_c) = [s \mapsto (1, 1, G), f \mapsto (1, 2, R)]$, i.e., changing the color of s to green, and that of f to red, results in the influencing effect making the decision maker to swap his/her choice and to select an alternative preferred by the stakeholder.

However, note the impact of this modification depends on the set of non-modifiable criteria $\{t, r\}$. For instance, if the utility of f is null and of s is maximal, there is no effect that will make a decision maker with a weight on $l \neq 1$ and additive aggregation function to change its decision from s to f. Similarly, if the decision maker has a weight equal to 0 on the criterion t, then all effects have no impact. See details of the last case in Table 3 for the same set of criteria evaluations $\sigma_u = [s \mapsto (1, 1, N), f \mapsto (1, 2, N)]$ of the decision maker as in the previous example, but different set of weights $w_u = (0; 0.8; 0.2)$. This case demonstrates that if decision makers do not care about the trust of the network, there is no chance to make them to select a more secure alternative whatever modifications are applied to the modifiable criteria.

Table 3. Impact of all modifications to the color criterion for initial alternatives evaluations $\sigma_u = [s \mapsto (1, 1, N), f \mapsto (1, 2, N)]$ and criteria weights $w_u = (0; 0.8; 0.2)$ of the decision maker with $\rho(f, w_c, \sigma_c) = 1$ and $\rho(s, w_c, \sigma_c) = 0$

σ_{xy}	$v(s, w_u, \sigma_{xy})$	$v(f, w_u, \sigma_{xy})$	$a = \pi(w_u, \sigma_{xy})$	$\rho(a, w_c, \sigma_c)$
σ_{NN}	0.5	0.9	f	0
σ_{NR}	0.5	0.8	f	0
σ_{NG}	0.5	1	f	0
σ_{GN}	0.6	0.9	f	0
σ_{RN}	0.4	0.9	f	0
σ_{GR}	0.6	0.8	f	0
σ_{RG}	0.4	1	f	0
σ_{RR}	0.4	0.8	f	0
σ_{GG}	0.6	1	f	0

5 Influencing Population

In all previous sections, we have considered a deterministic decision maker by default. To allow modeling groups of users rather than single users, we may consider a *probabilistic* decision maker. We model this aspect by considering a probability distribution over weights, such that given a weight function w, $\psi(w)$

represents the probability of w. From a statistical point of view, $\psi(w)$ represents the percentage of the population with the weight distribution w.

The policy of the entire population can therefore be defined as given a MAUT model $M = (\mathcal{A}, \mathcal{G}, \Sigma_{\mathcal{G}}, n, \psi(w))$:

$$\pi(\psi, \sigma, a) = \sum_{w \in W} \{\psi(w) \mid \pi(w, \sigma,) = a\}. \tag{2}$$

For influencing a population of users, a stakeholder needs to look for an alternative (or subset of alternatives) with highest impact and a subset of modifiable criteria that makes this alternative (preferred by the stakeholder) to be selected by the majority of population.

$$\mathsf{opt}(w_c, \sigma_c, \psi_u, \sigma_u) = \arg \max_{\sigma' \in P_{\mathcal{M}}(\sigma_u)} \sum_{a \in \mathcal{A}} \rho(a, w_c, \sigma_c) \pi(\psi_u, \sigma'_u, a).$$

5.1 Population in Case Study: Selection of a Wi-Fi Network

As an example of population modeling, we can consider examples of three types of decision makers with the same criteria evaluation functions $\sigma_u = [s \mapsto (1, 1, N), f \mapsto (1, 2, N)]$, but different criteria weights $w_1 = (0.3; 0.5; 0.2)$, $w_2 = (0; 0.8; 0.2)$, and $w_3 = (0.8; 0; 0.2)$. Let us also consider a probability distribution ψ such that $\psi(w_1) = \psi(w_2) = \psi(w_3) = 1/3$. We can calculate that $\pi(w_1, \sigma_u) = f$, $\pi(w_2, \sigma_u) = f$, $\pi(w_3, \sigma_u) = s$, and therefore, following Equation 2, we have $\pi(f, \sigma, \psi) = 2/3$ and $\pi(s, \sigma, \psi) = 1/3$.

If a stakeholder wants to shift choices of a population of users, he/she may consider similar strategy as one proposed for influencing choice of individual decision makers, but taking into account weights of different groups of users.

6 Conclusion

In this work, we have proposed a model for influencing human decision making in security context. We have illustrated the approach with a case study of a public Wi-Fi selection, and have shown how optimal influence may be selected. Even though the resented multi-criteria model is simplified, when compared to possible real-life scenario, however, it establishes a basis for developing a more complex framework, which we consider as our future work.

The first step will be to consider more than two alternatives to select from. Moreover, it will be interesting to investigate more complex impact functions (e.g. non-monotonic), which may lead to a backfire of influencing, with decision maker selecting a worse alternative when compared to his/her initial intention. Another interesting aspect is related to studying different normalization functions and their interpretation by different decision makers. For instance, it was observed in [23] that a padlock sign, usually assigned to trusted Wi-Fi's, may be perceived as blocking of access by some users, who misinterpret, and, consequently, normalize differently evaluation of Wi-Fi's. Interesting aspects to study are dependence between different security decisions and applying a sequence or combinations of influencing effects. For instance, in [23], it was shown that the color effect has a higher impact when applied in combination with ordering effect of different networks presented to the decision maker by default.

Taking into account complexity of different criteria, such as ‘trust’ criterion, MAUT contribution may be further investigated by modeling more complex shapes of marginal utility functions, such as convex or concave utility functions corresponding to risk (risk-prone or risk-averse) attitude of decision maker, when compared to the linear marginal utility functions modeled here. Moreover, the quantities obtained through MAUT can be used to characterize the strength of the effect applied, following, for instance, the recent approach in the context of quantitative access control policies [18].

Finally, the Wi-Fi scenario provides an interesting basis for future work. The importance of name when choosing a Wi-Fi was studied in the context of trust in [9]. The ‘trust’ criterion is interesting as it may take into account various information, e.g. about decision maker’s location, to avoid situations, where the most trusted network for a researcher located in a coffee shop far away from universities appears to be the ‘eduroam’ Wi-Fi, an international network for all university staff of universities provided within campuses of universities only.

Acknowledgements. The authors acknowledge funding for Choice Architecture for Information Security (ChAISe) project EP/K006568/1 from Engineering and Physical Sciences Research Council (EPSRC), UK, and Government Communications Headquarters (GCHQ), UK, as a part of Cyber Research Institute. We gratefully acknowledge the support and contribution of our colleagues on the ChAISe project from Newcastle University: James Turland and Northumbria University: Debora Jeske, Lynne Coventry, Pam Briggs and Christopher Laing who worked with us to identify issues and solutions in this project. We are also thankful to unknown reviewers who helped to improve this paper.

References

1. Aime, M., Calandriello, G., Lioy, A.: Dependability in wireless networks: Can we rely on WiFi? IEEE Security Privacy 5(1), 23–29 (2007)
2. Belton, V., Stewart, T.: Multiple Criteria Decision Analysis: An Integrated Approach. Kluwer Academic Publishers, Dordrecht (2002)
3. Bishop, M.A.: The Art and Science of Computer Security. Addison-Wesley Longman Publishing Co., Inc., Boston (2003)
4. Bovens, L.: The ethics of nudge. In: Grüne-Yanoff, T., Hansson, S. (eds.) Preference Change: Approaches from Philosophy, Economics and Psychology. Philosophy and Methodology of Social Sciences, vol. 42, pp. 207–219. Springer, Theory and Decision Library (2009)
5. Chismon, D., Carter, T., Ruks, M., Hoggard, H.: Mobile devices: Guide for implementers. White paper, MWRInfoSecurity and Center for the Protection of National Infrastructure (CPNI), Basingstoke, UK (February 2013)

6. Choe, E.K., Jung, J., Lee, B., Fisher, K.: Nudging people away from privacy-invasive mobile apps through visual framing. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) INTERACT 2013, Part III. LNCS, vol. 8119, pp. 74–91. Springer, Heidelberg (2013)
7. Clarke, J., Hidalgo, M.G., Lioy, A., Petkovic, M., Vishik, C., Ward, J.: Consumerization of IT: Top risks and opportunities. ENISA deliverables, European Network and Information Security Agency (ENISA), European Network and Information Security Agency (ENISA) report (2012)
8. Farnham, G., Leune, K.: Tools and standards for cyber threat intelligence projects. Technical report, SANS Institute (2013)
9. Ferreira, A., Huynen, J.-L., Koenig, V., Lenzini, G., Rivas, S.: Socio-technical study on the effect of trust and context when choosing WiFi names. In: Accorsi, R., Ranise, S. (eds.) STM 2013. LNCS, vol. 8203, pp. 131–143. Springer, Heidelberg (2013)
10. Goodwin, P., Wright, G.: Decision Analysis for Management Judgment, 4th edn. J. Wiley (2009)
11. Heilmann, C.: Success conditions for nudges: A methodological critique of libertarian paternalism. European Journal for Philosophy of Science 4(1), 75–94 (2014)
12. AIDC worldwide mobile worker population 2010-2015 forecast. Technical report, IDC Australia (2012)
13. Kahneman, D.: Thinking, fast and slow. Farrar, Straus & Giroux, New York (2011)
14. Kahneman, D., Tversky, A.: Prospect theory: An analysis of decision under risk. Econometrica 47(2), 263–291 (1979)
15. Keeney, R., Raiffa, H.: Decisions with Multiple Objectives: Preferences and Value Tradeoffs. J. Wiley, New York (1976)
16. Kennedy, D., O'Gorman, J., Kearns, D., Aharoni, M.: Metasploit: The Penetration Tester's Guide, 1st edn. No Starch Press, San Francisco (2011)
17. Morisset, C., Groß, T., van Moorsel, A., Yevseyeva, I.: Formalization of influencing in information security. Technical Report CS-TR-1423, Newcastle University (May 2014)
18. Morisset, C., Groß, T., van Moorsel, A., Yevseyeva, I.: Nudging for quantitative access control systems. In: Tryfonas, T., Askoxylakis, I. (eds.) HAS 2014. LNCS, vol. 8533, pp. 340–351. Springer, Heidelberg (2014)
19. Seigneur, J.-M., Kölndorfer, P., Busch, M., Hochleitner, C.: A survey of trust and risk metrics for a BYOD mobile worker world. In: Proceedings of SOTICS 2013, pp. 82–91. IARIA (2013)
20. Servin, A., Kudenko, D.: Multi-agent reinforcement learning for intrusion detection: A case study and evaluation. In: Bergmann, R., Lindemann, G., Kirn, S., Pěchouček, M. (eds.) MATES 2008. LNCS (LNAI), vol. 5244, pp. 159–170. Springer, Heidelberg (2008)
21. Applying behavioural insights to reduce fraud, error and debt. Policy paper: Transforming government services to make them more efficient and effective for users, Cabinet Office, Behavioural Insights Team, UK (February 2012)
22. Thaler, R.H., Sunstein, C.R.: Nudge: Improving Decisions About Health, Wealth, and Happiness. Yale University Press, New Haven (2008)
23. Turland, J., Coventry, L., Jeske, D., Briggs, P., Laing, C., Yevseyeva, I., van Moorsel, A.: Nudging towards security: Developing an application for wireless network selection for android phones (in preparation, 2014)
24. Yevseyeva, I., Morisset, C., Turland, J., Coventry, L., Groß, T., Laing, C., van Moorsel, A.: Consumerization of IT: Mitigating risky user actions and improving productivity with nudging. In: Proceeding of CENTERIS 2014 - Conference on ENTERprise Information Systems. Springer (accepted, 2014)