Position Paper: Privacy Risk Analysis Is about Understanding Conflicting Incentives

Einar Snekkenes

Position paper: Privacy Risk Analysis is About Understanding Conflicting Incentives*

Einar Snekkenes

NISlab, Gjøvik University College

Abstract. We motivate and give a brief overview of the Conflicting Incentives Risk Analysis (CIRA) method, explaining key ideas and concepts, offering a small example and give an overview of remaining challenges in relation to the use of CIRA in large scale risk analysis projects.

1 Introduction

Privacy may be motivated by a multitude of reasons, e.g. its intrinsic value, legal or regulatory compliance or a general concern regarding improper use of personal identifiable information. For example, if an employer, insurance company or the press got hold of the medical record of a known individual, it may turn out to be very difficult to prevent this information from being used in a way that is unfavourable to the individual.

In this paper we restrict our attention to risks that emanate from deliberate actions. By risk we mean the concern relating to an undesirable surprise (i.e. a loss) caused by the implementation of some action. Risk is always subjective and relative to perception. E.g. I may have a concern in relation to the uncertainty regarding the publication of my medical records, i.e. I experience risk. We can extend this notion of risk to include opportunity risk. By opportunity risk, we mean the uncertainty that an individual will fail to seize the opportunity to implement actions that one could reasonably expect that he should implement.

Risk analysis answers the question: Are there any risks that require some kind of action by the individual exposed to the risk? Risk management is about implementing the necessary action to ensure that unacceptable risks are mitigated.

There is a need to improve the predictability and the coverage of the risk identification process in the context of intended human behaviour. This challenge is a consequence of limited availability of representative historic data relevant for new and emerging systems. Furthermore, to improve the efficiency of the discovery process, there is a need to identify issues that are key to risk discovery, and avoiding activities that shed little or no light on potential problem areas. In the following, we explain how the Conflicting Incentives Risk Analysis (CIRA) method[1] adresses these issues.

* This work is part of the PETweb II project sponsored by The Research Council of Norway under grant 193030/S10.

2 Summary of CIRA

CIRA identifies stakeholders, actions and perceived expected consequences that characterizes the risk situation. According to CIRA, we have risk if the stakeholder that is in the position to trigger the action and the risk taker would be in disagreement as to weather or not the action should be implemented.

In CIRA, a *stakeholder* is an individual (i.e. physical person) that has some interests relating to the outcome of actions that are taking place within the scope of interest. There are two classes of stakeholders: the *action owners* and the *risk owner*. The *action owner* is in the position to decide if and when the action in question is to be executed. Typically, each stakeholder has associated a collection of actions that he owns. The *risk owner* is the stakeholder whose perspective we take when performing the risk analysis - that is, he is the stakeholder at risk. By *utility* we mean the benefit as perceived by the corresponding stakeholder. Utility comprises of *utility factors*. Chule et. al..[2] identify utility factors relevant for our work. Each utility factor captures a specific aspect of utility e.g. prospect of wealth, reputation, legal compliance, social relationships.

The CIRA tasks identify stakeholders, actions and consequences of actions in terms of perceived value changes to the utility factors. Our first task is to identify who is to take the role of the risk owner (e.g. the data protection officer, a politician that has some ideological desire to create a society where privacy is a common good, the prime minister, a typical citizen). Next, we need to identify what utility factors go into the risk owners perception of utility. I.e. how these utility factors are defined and how the various utility factors are weighted, reflecting the trade-off judgements made by the risk owner. This is required to assess how actions change the values of the various utility factors and thus influence changes to the risk owners perception of the utility offered by the action. We estimate the cumulative utility by using techniques from Multi Criteria Decision Analysis[3].

We then go on to identify the other stakeholders of interest. That is, we identify the stakeholders who are in the position to implement actions that will modify the values of the utility factors (and weights) of the risk owner. For each of these stakeholders, we identify the actions that are at their disposal and to what extent they modify the risk owner's utility factors. See e.g. [4] for a taxonomy of actions relevant in a privacy context.

When modelling the extent to which actions modify utility factors, care must be taken to make sure that the complete picture is captured. In particular, each action can be viewed as a strategy in a potentially complex game[5], where the implementation of the action amounts to the participation in a game.

In short, CIRA identifies situations such as the following: for all the actions that can be taken, can those that are in the position to implement an action obtain a significant benefit and at the same time cause damage to the risk owner (in terms of loss of utility)? Are there actions that one can reasonably expect that the action owner should take, but for which the action owner would have to take a loss in utility and the risk owner have the prospect of a gain?

3 A small example

We consider a situation, involving a bank and a customer. The bank is offering financial advice to the customers. The advice is offered by the bank through a financial adviser. The question is: Is the customer at risk? I.e. is there a legitimate concern by the customer in relation to the uncertainty that the bank is set up such that the customer may receive bad advice? From the perspective of the risk analyst, is there an easy procedure that will help the risk analyst to determine if the customer is at risk?

We assume that the financial adviser is rational in the classical economic sense in that the only factor that he considers when providing advice to customers is the monetary value of the bonus he receives if the customer buys the product he suggests. The financial adviser may obtain personal information about customers such as mental capacity, education, financial maturity, degree of dyslexia, relationship status, mental problems etc. through the operators of an identity management system. The bank has very few ethical standards, and those that exist are not enforced. There is at least one financial service offered by the bank that provides the adviser with a significant bonus, and may result in a significant loss for the customer (A1). There is also a service offered by the bank where the advisor will have to do a significant volume of work without this effort resulting in what he perceives to be a fair return in terms of utility and where the customer is expected to gain significant utility (A2).

Applying the CIRA method, we easily see that the customer is facing the following risks: Being offered service A1 and accepting the offer. Not being offered service A2, and consequently not being able to take advantage of it. In both cases, we have situations of conflicting incentives.

In the context of CIRA, risk mitigation amounts to modifying the weights that the stakeholders assign to the relevant utility factors or to what extent actions modify the values of the utility factors. In the above example, changing the weights of the utility factors can be achieved by e.g. forcing financial advisers to complete ethical training and/or using empathy or ethics screening tests before hiring. Changing the impact that actions have on utility factors of the advisers can be realized by modifying the bonus scheme - e.g. converting any bonus awarded to a liability, and offering the customer a compensation that exceeds the customer's loss if there is a complaint from the customer that he received unfavourable advice. This new bonus rule must then be made available to all concerned.

4 Challenges and limitations

The CIRA method is still at an early stage of development. The method may benefit from further work on issues such as: Identification of a more broad action repertoire, e.g. along the lines of the taxonomy relevant in a privacy context[4], the capturing of uncertainties in relation to estimates using e.g. interval arithmetic [6] or bounded probabilities[7]. We also need a more comprehensive

taxonomy of utility factors to capture human goal directed behaviour to state but a few. The use of intervals or bounded probabilities instead of point values will provide a link between quantitative and qualitative interpretation of CIRA. In many cases, the risk analyst may not have access to the relevant individuals. Then the utility factors and their trade-off may have to be obtained from past behaviour, profiles or from distributions obtained from similar individuals. Some stakeholders may find the CIRA approach to risk analysis rather intrusive, trying to game the analyst, failing to provide the analyst with a correct set of utility factors and/or their weights. The combination of utility factors using linear weights may fail to correctly model the stakeholders real assessment of utility e.g. in the presence of threshold values and utility factor dependencies.

5 Conclusions

We have given an overview of CIRA and explained how this method can be used to identify privacy risks. Using CIRA, the analyst is provided with guidance with respect to the key issues that are the root sources of many risks. We argue that in spite of its lack of maturity, CIRA offers concepts and procedures that will improve the process of identifying and analysing risks relating to intended human behaviour.

References

1. Rajbhandari, L., Snekkenes, E.: Intended actions: Risk is conflicting incentives. In: Proceedings of the Information Security Conference (ISC 2012), Springer Verlag (2012) 370–386 LNCS.
2. Chulef, A., Read, S., Walsh, D.: A hierarchical taxonomy of human goals. Motivation and Emotion **25** (2001) 191–232
3. Greco, S., ed.: Multiple Criteria Decision Analysis: State of the Art Surveys. International Series in Operations Research & Management Science. Springer (2005)
4. Solove, D.J.: A Taxonomy of Privacy. University of Pennsylvania Law Review **154**(3) (January 2006) p477+
5. Rasmusen, E.: Games and Information: An Introduction to Game Theory. 4th edition edn. Wiley-Blackwell (October 2006)
6. Moore, R.E., Kearfott, R.B., Cloud, M.J.: Introduction to Interval Analysis. SIAM (2009)
7. Ferson, S., Hajagos, J.G.: Arithmetic with uncertain numbers: rigorous and (often) best possible answers. Rel. Eng. & Sys. Safety **85**(1-3) (2004) 135–152