Modelling Security Risk in Critical Utilities:

The System at Risk as a Three Player Game and Agent Society

J.S. Busby[1], A. Gouglidis[2], S. Rass[3], S. König[4]
[1] Lancaster University, Department of Management Science, Lancaster, UK
[2] Lancaster University, School of Computing and Communications, Lancaster, UK
[3] Universität Klagenfurt, Institute of Applied Informatics, System Security Group, Klagenfurt, Austria
[4] Austrian Institute of Technology GmbH, Klagenfurt, Austria
{j.s.busby, a.gouglidis}@lancaster.ac.uk; stefan.rass@aau.at; sandra.koenig@ait.ac.at

***Abstract*— It becomes essential when reasoning about the security risks to critical utilities such electrical power and water distribution to recognize that the interests of producers and consumers do not fully coincide. They may have incentives to behave strategically towards each other, as well as toward some third party adversary. We therefore argue for the need to extend the prior literature, which has concentrated on the strategic, adaptive game between adversary and defender, towards 3-player games. But it becomes hard to justify modelling a population of consumers as a single, decision making actor. So we also show how we can model consumers as a group of mutually-influencing, yet not centrally co-ordinated, heterogeneous agents. And we suggest how this representation can be integrated into a game-theoretic framework. This requires a framework in which payoffs are known by the players only stochastically. We present some basic models and demonstrate the nature of the modelling commitments that need to be made in order to reason about utilities' security risk.**

***Index Terms* – Security, utilities, game theory, agent-based model.**

I. INTRODUCTION

Industrial control systems have become increasingly vulnerable to cyber-attacks due to their interconnection with insecure corporate networks. The risk is particular serious for critical utilities such as water and electricity distribution systems. The recent attack in the Ukraine [1] illustrated how attackers could ultimately gain physical control over such a system and inflict widespread harm on a society that is heavily dependent on basic commodities and services.

How to assess this risk has emerged as a problem of significant interest in recent years, drawing from the more general literature on assessing risk from deliberate attacks – called 'adversarial risk analysis' [2, 3] to distinguish it from the traditional 'probabilistic risk analysis'. This shift in thinking about the risk agent as being an adaptive, strategic actor rather than chance failure has, in particular, led to the use of adversary-defender games to analyse risk in utility systems [4, 5].

Our claim is that producers and consumers of a utility have interests that do not completely coincide. They may act strategically towards each other, and this possibility can be of considerable importance. For example, a producer might not disclose a threat to its consumers in order to avoid a panic response that might lead to over-consumption and exacerbation of an attack. Consumers might exaggerate the harm they experience in order to make compensation claims. Either strategy might influence the payoff to the adversary. Thus there are various possibilities for producers and consumers to act strategically and adaptively towards each other as well as towards a third party adversary. We explore these in this paper, showing in a basic way how they can be modelled, and indicating the nature of the games that need to be represented.

However, when we separate the interests of producer and consumer as distinct players, we encounter the problem that the consumer, in reality, is not a unitary decision maker, but a set of heterogeneous agents that influence each other, without central coordination, through a social network. We therefore also show how an agent-based model of consumers can be integrated into the overarching game. Such a model is inherently stochastic in nature because the specific characteristics of heterogeneous agents that shape their risk responses are randomly endowed. This means that the populations' responses are probability distributions, and the game between the utility and adversary has stochastic payoffs that need to be indexed appropriately to analyse the game.

In the remainder of this paper, we suggest key scenarios, show how these can be represented as a game, show how we can replace one player with an agent-based model of responding, and show how the outcomes this produces can be integrated back into a game. We briefly discuss the main insights that this work yields.

II. THE ADVERSARY-PRODUCER-CONSUMER INTERACTION

A. The first scenario: the threat of an attack

The clearest scenario involves a system such as a water utility whose control systems are under threat from a specific, directed cyber-attack – an example of which can be found in the Maroochy case [6]. The focus is on the strategies of adversary and utility and the payoffs combine the costs of attack and defence with the harms arising from an attack. The complication is that consumers also make choices relevant to their payoffs and the payoffs of others. For example, they may choose to boil contaminated drinking water when warned to do so, or they may substitute bottled water. The costs and harms they experience produce gains for the adversary and losses

The authors wish to acknowledge funding from the European Union Seventh Framework Programme under grant agreement no. 608090, Hybrid Risk Management for Utility Networks (HyRiM).

for the utility. And the adversary and utility choose strategies in the knowledge that consumers make such choices and influence their own payoffs.

The extensive form of this scenario in its simplest form is shown in Fig. 1. The attacker has a choice of whether to attack or not. The utility simultaneously has two choices: first, whether to enhance or just maintain its defences; second, whether to amplify the risk in its communications with the consumer, admit the objective level of risk, or attenuate it. The consumer receives a risk communication containing a risk level, but it does not know if this is an amplification or attenuation of the objective value. So it must simultaneously choose how to interpret it (by inflating it on the assumption that it was attenuated, accepting it as it is, or discounting it on the assumption that it was amplified). In this representation, the consumer's strategies concern its choices about how to interpret communications, not choices about whether to protect itself. Protection, or not, follows the interpretation choice.

We assume that if the attack takes place, there is an objective probability B of harm H to the consumer if the utility does not enhance its defences. This probability is reduced by a factor $D < 1$ if the utility does enhance the defences at cost K. The consumer incurs H with probability B or $D \cdot B$ unless it takes its own precautions (such as boiling drinking water, or substituting bottled water) at cost C. We also assume that if the utility chooses to amplify the objective risk it will communicate a risk message $M = 2B$, and to attenuate $0.5B$. If the consumer chooses to inflate the communicated risk M the consumer will believe the risk is $2M$ and if it chooses to discount it will believe the risk is $0.5M$.

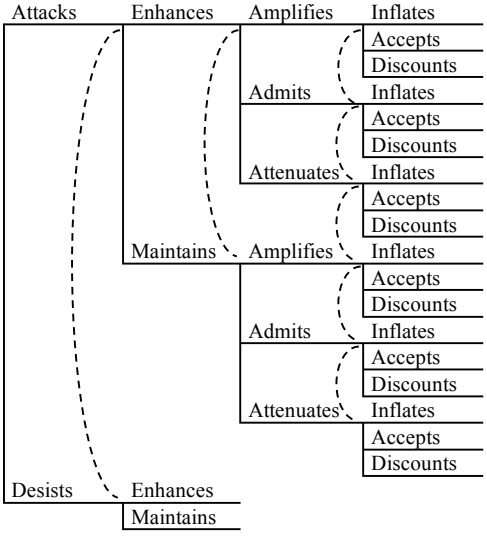
To define the consumer's payoffs, we give the consumer a threshold level, T. The consumer takes precautionary action only when its belief L about probability of incurring the objective harm H exceeds this. Thus, for example, in a state in which the utility's strategy is to enhance defences and amplify the risk, with the consumer choosing to inflate the communicated risk, the consumer will believe the risk is $4B$. It will *not* protect itself if $4B < T$, *anticipating* it will incur a payoff $-4BDH$. If $T < 4B$, on the other hand, the consumer will protect itself at a cost of C and incur no harm.

The payoff to the utility is the possible cost of defence K, plus some proportion $E_1 < 1$ of the cost incurred by a consumer that decides to protect itself at cost C, plus some proportion $E_2 < 1$ of the harm incurred by the consumer. This sharing of the consumer's costs reflects loss of goodwill and reputational damage, so the consumer's costs are *not* reduced by this sharing. Any monetary compensation, not represented here, from utility to consumers, would reduce the consumer's losses by this amount. The factors E_i also reflect the way in which the consumer's protective action could damage the utility's interests – for example if consumers boycotted the utility's product. But the utility has a different view of the probability of an unprotected consumer incurring H, knowing it to be B, not L. So, for example, when the consumer is anticipating a payoff of $-4BDH$ the utility knows the payoff to be $-BDH$ (even though it knows the consumer 'knows' it to be higher), and so its payoff component reflecting this is $-E_2BDH$. The game remains one of complete information as the players are assumed to know that other players will make different evaluations of the same risk, having chosen particular communication or interpretation strategies.

To save space we do not give the adversary's payoff, but assume this is the sum of the losses inflicted on the legitimate actors, utility and consumer, less its cost of attack, A. This assumes that the utility and consumer are both right in the evaluations of their own payoffs, despite potentially different beliefs in the level of risk. In Fig. 1, $[a]\ b : c$ denotes 'if a then b else c'.

B. The second scenario: the spoofed or hoaxed attack

In the second scenario, an attacker spoofs an attack by spuriously announcing an attack against a water utility on social media. The immediate object of the attack is the consuming public, and the public can choose whether to respond (for example incurring the costs of substitution by bottled water). It could choose to make no response, to avoid the cost and perhaps also anticipating that a response would encourage future attacks. The attacker can



Message M, Belief L	*Utility payoff P_u*	*Consumer payoff P_c*
$2B, 4B$	$-K + [T < 4B] - E_1C : -E_2BDH$	$[T < 4B] - C : -4BDH$
$2B, 2B$	$-K + [T < 2B] - E_1C : -E_2BDH$	$[T < 2B] - C : -2BDH$
$2B, B$	$-K + [T < B] - E_1C : -E_2BDH$	$[T < B] - C : -BDH$
$B, 2B$	$-K + [T < 2B] - E_1C : -E_2BDH$	$[T < 2B] - C : -2BDH$
B, B	$-K + [T < B] - E_1C : -E_2BDH$	$[T < B] - C : -BDH$
$B, 0.5B$	$-K + [T < 0.5B] - E_1C : -E_2BDH$	$[T < 0.5B] - C : -0.5BDH$
$0.5B, B$	$-K + [T < B] - E_1C : -E_2BDH$	$[T < B] - C : -BDH$
$0.5B, 0.5B$	$-K + [T < 0.5B] - E_1C : -E_2BDH$	$[T < 0.5B] - C : -0.5BDH$
$0.5B, 0.25B$	$-K + [T < 0.25B] - E_1C : -E_2BDH$	$[T < 0.25B] - C : -0.25BDH$
$2B, 4B$	$[T < 4B] - E_1C : -E_2BH$	$[T < 4B] - C : -4BH$
$2B, 2B$	$[T < 2B] - E_1C : -E_2BH$	$[T < 2B] - C : -2BH$
$2B, B$	$[T < B] - E_1C : -E_2BH$	$[T < B] - C : -BH$
$B, 2B$	$[T < 2B] - E_1C : -E_2BH$	$[T < 2B] - C : -2BH$
B, B	$[T < B] - E_1C : -E_2BH$	$[T < B] - C : -BH$
$B, 0.5B$	$[T < 0.5B] - E_1C : -E_2BH$	$[T < 0.5B] - C : -0.5BH$
$0.5B, B$	$[T < B] - E_1C : -E_2BH$	$[T < B] - C : -BH$
$0.5B, 0.5B$	$[T < 0.5B] - E_1C : -E_2BH$	$[T < 0.5B] - C : -0.5BH$
$0.5B, 0.25B$	$[T < 0.25B] - E_1C : -E_2BH$	$[T < 0.25B] - C : -0.25BH$
	$-K$	0
	0	0

Figure 1. Representation of the simple attack scenario

choose to invest in amplifying its credibility – for example by hacking the utility's customer communications systems but without staging an attack in reality. But, as a 2-player game, this ignores the way in which the attack on consumers' psychological well-being is therefore also an attack on the utility, and the way in which a utility therefore has an incentive also to anticipate and act against such spoofing – for example by publicizing the strength of its controls, and even improving such controls.

In reality, this is not a separate game from the previous one, since spoofing is a third option for the adversary's choice, to put beside a real attack and no action at all. But the game is somewhat different because the consumer finds out about the supposed attack from the attacker or its communication proxy (such as an anonymous news report), not from the utility. The utility might know that a spoof is a spoof and can choose to communicate this with the consumer. The considerations behind such communications could be different from those in the previous game: denying a spoof may lend it more credibility than ignoring it, and may damage the utility's reputation by suggesting it is complacent to those consumers who believe strongly in the spoof.

A very simple version of a spoofing game is shown in Fig. 2. It is assumed here that the utility has the same choice to make about its defences, and the choice in its communications is now simply whether to deny the spoof or to ignore it. The consumer similarly has a binary choice of whether to believe the spoof is real or not. If it believes it to be real, it takes precautions at cost C and knows then there is no harm; if it believes it to be a spoof, it believes the payoff is zero. The utility's payoffs include the cost of enhancing defences, K, but also the reputational costs Q and R of denying or ignoring a spoof that the consumer believes is real. The parameter A' is the modified cost to the attacker. It is assumed that the utility's choice to defend its system is unknown to the consumer, or at least uninfluential, so in the context of a spoof makes no difference to the consumer payoffs. Again the adversary's payoff is simply the negative sum of the other players' payoffs plus its cost of attack, $-A' - P_u - P_c$. The structure for the initial analysis in the previous scenario is brought forward as Θ_i so the result is a single, more complex game.

For illustration, we put the probabilities $T = 0.2$, $B = 0.3$, $D = 0.2$, $E_1 = 0.4$, $E_2 = 0.5$. The payoffs, costs and harms are expressed as a subjective consequence, allowing the consequences experienced by the three players to be comparable. A payoff or certainty equivalent of 0 means a negligible subjective impact. A payoff of -10 means a completely negative subjective effect. And one of +10 means a completely positive one. For illustration we put $K = 3$, $C = 3$, $H = 9$, $A = 2$, $A' = 0.5$, $Q = 4$, $R = 2.0$. When the combined game is analysed with these payoffs, there turn out to be 8 pure-strategy Nash equilibria, in which the utility *always* enhances defences. But we suggest that a more descriptive analysis is obtained by using the 'Level-k' approach, recently used in the risk literature for an attacker-defender game [2]. This is defined recursively, such that a player playing at level 0 makes a random choice, and a player playing at level k uses a maximin decision rule expecting other players to play at level k - 1. In our game, with the given parameters, level 2 plays are for the adversary to attack, for the utility *not* to enhance defences and amplify, and for the consumer to discount. The contrasting strategies in the equilibria and the level-k choices show how sensitive any prediction of a player's behaviour is to their game-theoretic rationality. Only the equilibria include an adversary's choice of spoofing or desisting from an attack.

Adversary Attacks	*Utility* Θ_i	*Utility*	*Consumer*	*Message M, Belief L*	*Utility payoff* P_u	*Consumer payoff* P_c
Spoofs	Enhances	Denies	Believes	0, B	$-K - Q$	$-C$
			Rejects	0, 0	$-K$	0
		Ignores	Believes	-, B	$-K - R$	$-C$
			Rejects	-, 0	$-K$	0
	Maintains	Denies	Believes	0, B	$-Q$	$-C$
			Rejects	0, 0	0	0
		Ignores	Believes	-, B	$-R$	$-C$
			Rejects	-, 0	0	0
Desists	Enhances				$-K$	0
	Maintains				0	0

Figure 2. Representation of the spoofing scenario

C. The third scenario: the alleged attack as distraction

In a third scenario, a producer either orchestrates an attack or claims that there is an attack as a distraction – typically from an act of neglect or negligence, or perhaps from weak financial performance, an unreliable water supply, or billing errors. Evidence of such scenarios is very hard to come by, for obvious reasons, but it is plausible, and in general such decoys provide an important social strategy for dealing with blame and responsibility. A report on the Ukrainian power grid attack argued that failures for other reasons are sometimes attributed to cyber-attacks [1]. A producer thus has a strategic choice of whether or not to initiate such a scenario. The other main player is the consumer, who can choose whether to behave as though there were no crisis, whether there were a crisis caused by an attack from a third party, or whether there were a crisis, but not as a result of such an attack. The payoffs concern reputation. This might have an effect on consumption, but is most likely to involve complaint, protest, lobbying, and political or regulatory action. Again on face value this is a 2-player game, between utility and public. But again, if less clearly, some real adversary has a role as a third player. It may receive a payoff: a positive payoff if it experiences benefit from being implicated in some crisis (whether or not it played a positive part in it); or a negative payoff if it is wrongly blamed for a crisis and the authorities act against it. If the crisis provokes an investment in risk controls, spuriously, it might gain in the sense that the utility thereby incurs an additional cost, or lose in the sense that a real attack will then be more costly. And it might have a choice whether to admit blame, untruthfully, or deny it. This choice will have some influence over the payoffs to the other two players.

Unlike the previous two games, the state of the world at the start of the game is quite different: not the potential for a move from an attacker, but the occurrence of some event from which the utility might seek a distraction. There is no simple intersection with any of the choices in the previous games. The consumer has to gauge the credibility of the utility's communication, as in the first game, but it has the adversary's communication to judge in combination. The utility is the first mover, and now its adversary has perfect

information and makes a sequential response, either disowning the attack or claiming (spuriously) it was theirs, with various costs and benefits. The consumer chooses an interpretation – no crisis, a crisis arising from an attack, or a crisis wrongly attributed to an attack – as last mover, but not knowing whether the utility announcement *is* a distraction, and not knowing whether the attacker's statement is true or not.

In Fig. 3, in order to simplify, we ignore the question of whether there is a crisis or not, assuming that the consumer has objective knowledge on which it acts to protect itself. Instead, the figure depicts a game where the central issue is whether the consumer chooses to believe the utility's claims of an attack. If the consumer chooses to believe the alleged attack, whether or not the adversary disowns a role in the attack, it suffers some subjective insecurity about attacks, I_1. If it chooses to disbelieve it suffers a subjective insecurity about accidents or negligence, I_2. The adversary's reputation is enhanced to some degree by a supposed attack that it claims for itself, by W_1, but only if the consumer believes it really was an attack. It is enhanced to a lesser degree W_2 if it disowns the attack and yet the consumer still believes it was an attack. In both cases, however, it is then at risk from retaliation or prosecution, Z. If it claims an attack that the public disbelieves, the adversary loses some reputation, V. The utility also loses reputation N when the consumer disbelieves it, and makes some small gain in reputation M when the consumer believes it, on the basis that each time someone believes it that becomes more of a habit. The utility, if it does not attempt the distraction but admits to some other failure, also faces a reputational loss D.

We do not take this specific scenario further, but simply use it to show that there are other games that can be identified around the security of a utility. Another possibility that this game has in common with the previous one is that of coalition. Clearly the utility and the consumer have interests in common. We suggested in the first scenario that some of the consumer's losses are reflected in the utility's payoff function. Perhaps more importantly, in any reasonably advanced society, the arrangements for compensation enforce a kind of coalition. But these arrangements are exogenous to the game and mean that the division of payoffs is not determined by the nature of specific games such as the ones represented here. It is also possible for ad hoc coalitions to form in principle, however, and this could be modelled in contexts where it is plausible. Equally, the basic representation could be extended into a conditional game [7], to deal with the social relationship between utility and consumer. It is potentially misleading to neglect this and to treat the players – especially utility and consumers – as isolated decision makers connected only through the game.

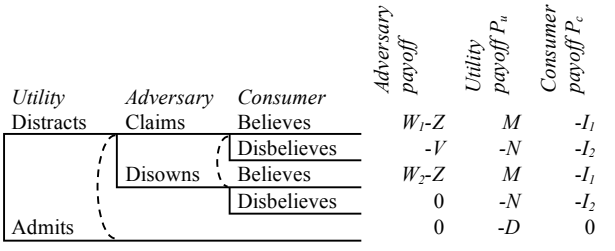


Figure 3. Representation of the distraction scenario

But we return to the reality that the 'consumer' is a set of heterogeneous actors who, although they influence each other within a social network, have no central coordination. This is the second main theme of our paper, and so - rather than develop further scenarios - we now move on to modelling this aspect.

III. The Multiple Heterogeneous Consumer Model

A. The agent based model

The central issue for the consumer in the game theoretic model is how its payoff (and therefore the payoffs of utility and adversary) depends on its risk belief and action threshold. Its strategy does *not* involve choosing whether it protects itself or not, but how it forms its risk belief in the light of the utility's risk communication. This then decides whether precautionary measures are taken or not. This premise comes from recent work on 'social risk amplification' [8]. Both differential equation-based diffusion-like models [9, 10] and agent-based models [11] of social risk amplification have been developed to explore how the inhabitants of a social network respond to risk. The phenomenon is somewhat similar to the spread of physical viruses through an agent network [12]. In Figs. 1 and 2, the single consumer (or homogenous consumer population) strategically decides how to interpret the utility's risk communication. We now replace this unitary decision maker with an agent population, and we give the agents simple decision rules that both 1) respond to a utility's risk communications, and 2) shape and are shaped by the risk beliefs of social neighbours. This agent-based model predicts the proportion of the agent population that takes precautionary measures, and it is this proportion that then shapes the payoffs in a reduced game. We describe the agent model briefly, and then show how the prediction is used in the main game.

An individual within a consumer population, $c \in C$, exists in a social network G, an undirected graph in which its neighbours $N(c) = \{n \in C \mid (c, n) \in G\}$. The period of a crisis is divided into a large number T of discrete 'ticks', at each of which one consumer is chosen at random with equal probability for activation. It interacts with one neighbour n, chosen at random with equal probability from $N(c)$. This interaction is itself a simple game of complete information in which the two consumers each decide whether to take precautionary action, $a_c \in \{0, 1\}$. The payoffs from this local game involve both potential physical harm and social rewards. Consumer c believes the physical harm will be $h_c \in [0, 1]$ with probability q_c if no precaution is taken, the cost of which is believed to be $k_c \in [0, 1]$. In terms of social rewards, c gains a subjective payoff of $f_c \in [0, 1]$ for conforming or being part of a consensus, and a subjective payoff $u_c \in [0, 1]$ for being regarded as being prudent (when it takes precautions). If c has a preference for the social approval of recklessness rather than prudence it gains a payoff of $1 - u_c$ (when it takes *no* precaution). This sets up a local game for every consumer-consumer interaction as shown in normal form Fig. 4. We use the 'Level-k' approach, described briefly above, with $k = 2$ to model the decisions of both parties in all interactions.

		Consumer n	
		Takes precaution	Takes no precaution
Consumer c	Takes precaution	$u_c + f_c - k_c$, $u_n + f_n - k_n$	$u_c - k_c$, $(1 - u_n) - q_n h_n$
	Takes no precaution	$(1 - u_c) - q_c h_c$, $u_n - k_n$	$(1 - u_c) + f_c - q_c h_c$, $(1 - u_n) + f_n - q_n h_n$

Figure 4. Local game for consumer interactions

The traits, f_c and u_c, are randomly endowed from a beta distribution with specified modes, minima and maxima, and are fixed for the duration of the model. The beliefs h_c and k_c are also randomly endowed and fixed. The belief in the probability of harm, q_c, is randomly endowed (with very small mode) at the start of the model period, before a crisis begins. When c is active, it takes a weighted average of its prior belief and the risk level b (a probability of incurring harm) being broadcast by the utility, which is zero until a crisis starts. Thus $q_c(t) = r_c b(t) + (1 - r_c)\, q_c(t\text{-}1)$, where the weight r_c is the consumer c's credulity or acceptance of the utility's risk message, another randomly endowed and fixed trait. The value of $q_c(t)$ weights the harm in the payoffs of the interaction game in Fig. 4.

The utility's broadcast risk level b is the product of an objective risk level B and a chosen amplification factor (2, 1 or 0.5) – both inherited from the overall game specified earlier. Thus the agent model is parameterized by the amplification factor: we get a different model, and different outcome, for the different amplification choices made by the utility, as in the original game. In this way, over the model period, the consumers' beliefs in the probability of some harm evolve over time, in response to the utility's broadcast and their own earlier beliefs. And their choice of whether or not to take precautionary action develops both in response to this changing probability belief and to the socially strategic interaction with their neighbours defined by this game.

The agent model was simulated with a simple system of 1000 consumers connected in a scale-free social network with a power law distribution of link numbers κ in which the number of nodes with κ links is proportional to $\kappa^{-\gamma}$. This applies to many networks of social contacts, in which γ generally lies in the range of 2 to 3 [13]. We set $\gamma = 2.5$. Modal values of the fixed endowments for k_c, u_c, f_c, r_c, h_c were 0.7, 0.5, 0.7, 0.9 and 0.8. The objective value of risk, $B = 0.3$, as in the game described earlier. To use the outcome in the game, we use the peak values of the fraction of the consumer population taking precautionary measures. It is this peak response that is probably most relevant to the other two players (the utility and the adversary). As the model is stochastic, we generate a frequency distribution over this peak value for 1000 runs of the agent model simulation.

B. Using the results of the agent model

The first part of the game (based on the first scenario) now appears as in Fig. 5. There is no unitary consumer to make strategic plays. The consumer payoff is determined by the peak proportion X of consumers in the consuming population who take precautions when given the risk message that defines the utility's communication strategy. Using the parameters from earlier, the net consumer payoff would be $-XC - (1 - X)BDH$ when the utility has chosen to defend, or $-XC + (1 - X)BH$ when it has not. It is important to say that these payoffs, although incurred by the consumer population, are not the payoffs of a consumer player. The consumer is no longer a strategic actor in the game, and its payoff is needed only because it contributes to the adversary's payoff. The net utility payoff, using the same approach as earlier, is $-K - X E_1 C - (1 - X) E_2 BDH$ (if the utility defends) or $-X E_1 C - (1 - X) E_2 BH$ (if it does not defend). The adversary's payoff as before is the sum of the negative payoffs to the utility and consumers, less its own cost. The game is now simpler in structure, with two players, but with payoffs defined by probability distributions over the consumer response. In Fig. 5, X_M is subscripted to denote that its distribution $F_X(M)$ depends on M through the operation of the agent model.

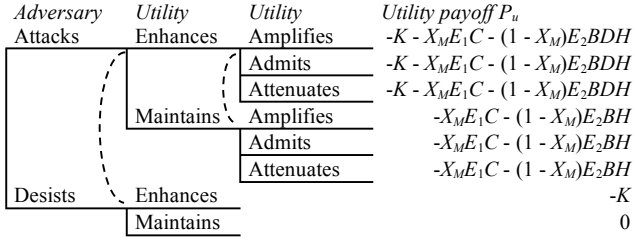


Figure 5. Revised game with consumer responses from agent model

Limitations of space preclude description of the extension into the second scenario, involving spoof attacks. This requires a similar agent model, with the utility's denial or ignoring of the spoof as an exogenous variable. Instead we move on to the final element of our method: the analysis of the game when we end up, as here, with payoffs defined using random outcomes rather than crisp values.

IV. The Analysis of a Game with Stochastic Payoffs

As the consumer model revolves around perceptions of risk, and as we essentially take a risk perspective on this analysis, it is natural to deal with the probability distributions defining the payoffs by using a risk index. Recently, MacKenzie [14] has proposed using Value At Risk (VaR), which originated in the context of financial risk, more widely. VaR is based on choosing a specific quantile in a probability distribution over a payoff in order to summarize a distribution by stating how great a harm, or more, should be expected with a certain probability. Thus $\mathrm{VaR}_{0.01}[F_Z]$ indicates that there is 0.01 probability of a harm of at least that value of Z, given a probability distribution F_Z over payoff Z. The conditional value at risk, $\mathrm{CVaR}_{0.01}$, is the expected value of losses given that they exceed the value at risk [14].

If a 1% VaR adequately represents the utility's risk attitude, the critical criterion for the utility is to choose a strategy that minimizes the worst $\mathrm{VaR}_{0.01}$. For the adversary, the support for its payoff distribution will normally be mostly positive. And its decision rule is likely to be a lot less conservative. But it will still reasonably maximin over its gains, and its payoffs can still reasonably be summarised by a quantile, for example $\mathrm{VaR}_{0.10}$. This means that the adversary payoff is no longer the simple sum of the legitimate actor harms with the adversary's costs – since its view of the legitimate actor harms is summarized with a different VaR index. The agent model

outcome, the proportion of consumers taking precautions X, is the only random variable. So, for example, for the strategy combination <adversary attacks, utility defends, utility amplifies> the utility's payoff is $\text{VaR}_{0.01}[-K - X_M E_1 C - (1 - X_M) E_2 BDH] = -K - \text{VaR}_{0.01}[F_X(2B)] E_1 C - (1 - \text{VaR}_{0.01}[F_X(2B)]) E_2 BDH$. The adversary's payoff is $-A - K - \text{VaR}_{0.10}[F_X(2B)](1 + E_1)C - (1 - \text{VaR}_{0.10}[F_X(2B)])(1 + E_2)BDH$.

The results of applying this to the game in Fig. 5, using parameter values carried over from the 3-player game where appropriate, produces a single equilibrium where the adversary attacks and the utility both enhances defences and amplifies its risk message. The level-2 strategy for the adversary is to attack, but for the utility to maintain (not enhance) its defences and to amplify.

Much more elaborate approaches to dealing with payoffs as probability distributions have been developed, replacing distributions with the sequence of their moments and applying them to a mixed strategy solution for a game [15]. This provides an avenue for further work. But, in the same way that level-k game strategies capture how actors might actually reason, in reality, the use of a VaR index expresses an intuitive and straightforward way in which actors can actually reason about uncertain payoffs. Actors do not have to exhibit unrealistic levels of technical rationality, and do not have to assume this of other actors against whom they play.

V. Discussion and Conclusions

The intended contribution is to show how we can model utility cyber-security risks – and more generally how we can model situations in which there is strategic action not just between an adversary and a defender, but among adversary, defender and a third party that bears much of the risk of any attack. In the simplest approach, this leads to a 3-player game with a set of parameterized payoffs. This combines the traditional view of an adversary-defender game with a game involving risk communications among the 'good guys' – the utility and its consumers – a game that is important in defining whether consumers then take precautionary action. Strategies indicated by different analyses – for example Nash equilibria and level-k decision rules – are quite different, showing how sensitive the strategies and ultimate outcomes are to what we assume about the rationality of the players. In a more refined approach, we show how to replace the idea of the third party (in this case the 'consumer') as a unitary decision maker with a heterogeneous population in which agents are mutually influencing, but not centrally coordinated. Again, how the consumer population interprets the utility's risk communications is important, but so is the way in which consumers observe and interpret each other's risk responses in defining the response of the population as a whole. The outcomes of this agent-based model lead to stochastic payoffs and the need to analyse games with this property, for which we suggest the Value at Risk index.

There are several limitations in the basic conceptualization of a 3-player game. The reality of the risk of cyber-attacks on utilities is that there are more than three actors. Regulators, journalists and politicians may be involved, and themselves have strategic interests. Recent work, for example, deals with governments as actors who make disclosures about critical utilities [16]. Our particular 3-player game is also perhaps the simplest that could be envisaged in this context, so in reality the available strategies and payoff structures may both be more complex. And there are basic questions about the completeness of this game: it seems unlikely that the players can have the comprehensive knowledge of each other's payoffs that the game assumes. The agent-based model with which we model the heterogeneous consumer population similarly is based on a basic set of assumptions about the social network and individual agent decision rules that do not fully represent the relevant mechanisms within such a population.

References

[1] L. M. Robert, Assante; Tim, Conway, "Analysis of the Cyber Attack on the Ukrainian Power Grid," SANS and E-ISAC2016.

[2] C. Rothschild, L. McLay, and S. Guikema, "Adversarial risk analysis with incomplete information: A level-k approach," *Risk Analysis,* vol. 32, pp. 1219-1231, 2012.

[3] J. Rios and D. R. Insua, "Adversarial risk analysis for counterterrorism modeling," *Risk analysis,* vol. 32, pp. 894-915, 2012.

[4] A. Rose, G. Oladosu, and S. Y. Liao, "Business interruption impacts of a terrorist attack on the electric power system of Los Angeles: customer resilience to a total blackout," *Risk Analysis,* vol. 27, pp. 513-531, 2007.

[5] J. S. Simonoff, C. E. Restrepo, and R. Zimmerman, "Risk-Management and Risk-Analysis-Based Decision Tools for Attacks on Electric Power," *Risk Analysis,* vol. 27, pp. 547-570, 2007.

[6] M. Abrams and J. Weiss, "Malicious control system cyber security attack case study–Maroochy Water Services, Australia," *McLean, VA: The MITRE Corporation,* 2008.

[7] W. C. Stirling, "A Game-Theoretic Social Model for Multiagent Systems," in *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, 2013, pp. 2718-2723.

[8] R. E. Kasperson, O. Renn, P. Slovic, H. S. Brown, J. Emel, R. Goble*, et al.*, "The social amplification of risk: A conceptual framework," *Risk analysis,* vol. 8, pp. 177-187, 1988.

[9] W. J. Burns and P. Slovic, "The diffusion of fear: Modeling community response to a terrorist strike," *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology,* vol. 4, pp. 298-317, 2007.

[10] J. S. Busby and S. Onggo, "Managing the social amplification of risk: a simulation of interacting actors," *Journal of the operational research society,* vol. 64, pp. 638-653, 2012.

[11] J. S. Busby, B. S. S. Onggo, and Y. Liu, "Agent-based computational modelling of social risk responses," *European Journal of Operational Research,* 2015.

[12] M. Tang and X. Mao, "An Agent-Based Artificial Society Approach to Analyzing Social Propagation," in *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*, 2015, pp. 777-782.

[13] A.-L. Barabási, "Scale-free networks: a decade and beyond," *science,* vol. 325, p. 412, 2009.

[14] C. A. MacKenzie, "Summarizing risk using risk measures and risk indices," *Risk analysis,* vol. 34, pp. 2143-2162, 2014.

[15] S. Rass, S. König, and S. Schauer, "Uncertainty in Games: Using Probability-Distributions as Payoffs," in *Decision and Game Theory for Security*, ed: Springer, 2015, pp. 346-357.

[16] M. Yoshida and K. Kobayashi, "Disclosure strategy for critical infrastructure under common knowledge of 'naive government'," in *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, 2011, pp. 3463-3470.