

A Game-Theoretic Analysis of Cooperation in Anonymity Networks

Mu Yang¹, Vladimiro Sassone¹, and Sardaouna Hamadou²

¹ ECS, University of Southampton, UK

² Università Ca' Foscari, Venice, IT

Abstract. Anonymity systems are of paramount and growing importance in communication networks. They rely on users to cooperate to the realisation of an effective anonymity service. Yet, existing systems are marred by the action of ‘selfish’ free-loaders, so that several cooperation incentives are being proposed. We propose a game-theoretic model of incentives in anonymity networks based on parametric utility functions, which make it flexible, adaptable and realistic. We then use the framework to analyse the cost of cooperation and the performance of the gold-star incentive scheme in the Crowds protocol.

1 Introduction

Anonymity of electronic communication is rapidly becoming an essential requirement of today’s society, in particular as far as tracking web browsing and handheld, mobile devices is concerned. Its importance is increasingly recognised as crucial in many fields of computer-supported human activities, such as e-commerce, web surfing, consumer profiling, personalised advertising. Anonymity is needed both by individuals and organisations who want to keep their identities, interests and activities confidential. Cryptographic techniques, firewalls, VPNs, and similar, can only provide partial protection; indeed, they can only protect the contents of a communication, not its origin, destination and occurrence. This constitutes a problem because in general a lot of potentially sensitive information can be inferred by the mere presence of a communication between two parties. To address this issue, many anonymity systems and protocols have been proposed in the literature. Their purpose is to support anonymous communications, at least to some extent [6, 22, 11, 21]. Since public visibility is the default condition on today’s main networks, most notably the Internet, anonymity cannot be enforced by either senders or receivers, but must be created by using messages to hide messages. In fact, the consumers of the anonymity service are at the same time its providers, as they cooperate to generate the network activity that grants anonymity to the system as a whole. Typically, cooperation entails relaying other users’ messages in order to create sufficient ‘doubt’ as to whom the real message originator actually is.

Anonymity systems have a broad range of users, ranging from ordinary citizens who want to avoid being profiled for targeted advertisements, to companies trying to hide information from their competitors, to entities requiring untraceable communication over the Internet. With these many potential users, it would seem that anonymity services based on consumer/provider users will naturally be well-resourced and able

to operate efficiently. However, cooperation cannot be taken for granted. Just because functioning as a relay may cost a significant amount of processing power and bandwidth, not all the users are going to be *cooperative*. Some will indeed act *selfishly*, and only use the system to send their messages whilst ignoring the requests to forward others' messages. Obviously, with not enough cooperative users, the systems will hardly operate at all, and will certainly not be able to afford adequate anonymity guarantees. Observe that this is not a trivial problem as it may appear superficially. In fact, as part of the anonymity requirements which lay at their very core, these systems do not monitor their users' behaviours nor their identities, making it virtually impossible to detect selfish users. In other words, even without considering the several documented attacks against anonymity networks (cf. e.g., [15, 19, 20]), inducing users to cooperate to the anonymity mechanisms is among the most critical aspects of maintaining the security and viability of the network. Due to the demand for strong anonymity from large numbers of cooperative users, it is therefore vital that these systems are able to deploy '*incentives*' to encourage users' cooperation and so make the anonymity provision effective. Some interesting approaches to achieve that have been proposed, such as make running relays easier and provide better forwarding performance [23].

To evaluate whether these approaches are effective, we need a framework which empowers us to analyse them, as well as provide guidelines and some mechanism design principles for incentive schemes. This much we provide in the present paper, exploiting notions and techniques from Game Theory.

Game theory [12] concerns strategic decision-making by rational entities – referred to as *players* – who behave according to a given set of rules – the *game* – with the explicit purpose of maximising their own benefit. Benefits are expressed by *utility functions*, whose value is determined by the players' actions and, ultimately, by their decisions. The '*rationality*' hypothesis is very significant, meaning that players are always capable to choose their actions exactly as required to maximise their utilities. Game theory is an excellent tool to model anonymity networks. In such systems users compete for anonymity services in the way prescribed by specific protocols and, in doing so, they invest their own resources. At the core of their participation in the system is therefore a need to balance their gain in terms of security (viz., the level of anonymity guaranteed to them) and/or performance (viz., the speed at which their transactions are processed) against their costs (e.g., in terms of bandwidth, software, etc). Rationality here is reflected in the micro-economic mechanisms which underpin such cost/benefit decisions.

In this paper we build a game-theoretic framework to study incentive mechanisms in anonymity systems. We model user behaviours (viz., cooperative or selfish) as a non-cooperative game, and compute the equilibrium strategies for the users involved in the game. As games model systems, we rely on game-theoretic principles to predict whether the users will or will not be cooperative, i.e., under exactly what conditions they will participate in the system or just exploit it. We also use the game theoretic notion of *Nash Equilibrium* and *Dominant Equilibrium* to analyse the strategic choices made by different users. Our objective is achieved by considering a rich and flexible set of parameters for users' payoffs, including aspects such as anonymity, cost and performance. In general, a user's utility function $U_i(\cdot)$ in our framework is a sum of factor

functions $\Phi_k(\cdot)$, each representing on a given payoff relevant to the analysis at hand. Furthermore, such functions can be weighed differently for each individual user, so as to afford great flexibility to the model.

In the paper we apply this framework to a relative simple anonymity scenario: the Crowds protocol. We show that, if we consider anonymity as the only parameter of importance, then there is exactly one Nash (dominant) equilibrium, whose equilibrium strategy is to behave cooperatively. Whilst this may explain why in standard Crowds there are no incentives to cooperation, it fails to match the real-world experience that users can indeed behave uncooperatively. In fact, the picture changes radically as soon as we consider the cost of communication as a parameter: as cost is a potent dissuader, users will soon start to contemplate the opportunities of selfishness. We observe that users who constantly behave selfishly enjoy *no anonymity at all*: as they only forward their own messages, there can be no ambiguity whatsoever as to the origin of a message intercepted from one of them. Strategic users will therefore engineer complex strategies whereby cooperative and selfish phases alternate. This leads us straight to our framework of mixed-strategy games, in which we study – both through analysis and simulations – the collective equilibrium behaviour of strategic users. We furthermore focus on the impact on equilibria of environmental parameters such as the number of attackers, the volume of network traffic, etc, and investigate the mechanics they induce on the equilibrium points.

We anticipate that the full power of our model only comes to the forefront when incentive mechanisms are involved: the ability to analyse the dynamics of the users’ chosen behaviours – viz., the equilibrium strategies – as contextual parameters vary, make the model suitable to design and analyse incentive mechanisms. To exemplify this, we focus on the *gold-star incentive mechanism* [23], whereby cooperative users are rewarded for their behaviour by enhanced performance, in the form of quicker delivery time for their anonymous messages. Precisely, messages carrying a gold-star are routed with priority over other messages. Users gain ‘gold-star’ status, i.e., the ability to send gold-star messages, according to whether they “achieved a satisfactory performance for at least R times out of the last V measurements.” We conduct for gold-star-incentivised Crowds the same set of analyses we carried out for Crowds. In particular, we study the equilibrium strategies as typical parameters vary, and illustrate how at equilibrium a strategic user will be selfish at most with frequency $1 - R/V$. In other terms, our results confirm the effectiveness of the gold-star mechanism as an incentive to cooperation.

To the best of our knowledge, ours is the first application of game-theory to yield an applicable framework to model incentives in anonymity systems. We compare our approach with the existing literature in §6 and assess it in the concluding section §7.

Structure of the paper: §2 and §3 introduce the framework from its game-theoretic foundations. In §4 we present our analysis of Crowds, and in §5 that of Crowds extended with the gold-star mechanism. The appendices contain most of the proofs and some of the figures that could not find space in this exposition.

2 Game-Theoretic Incentive Framework

2.1 Strategies and equilibriums

In anonymity networks, honest users compete for anonymity services with limited resources, such as bandwidth from servers. We model honest users' behaviours in such networks as a non-cooperative game. Each player (user) is a rational agent trying to maximise her own utility and choosing her actions (e.g., cooperative, selfish, etc) strategically. The actions she chooses are so-called *strategies*; the players' chosen *strategies* are drawn from a (finite) set of actions, and determine their utilities. A *dominant strategy* for a player is a strategy which guarantees her an optimal utility irrespective of the strategies chosen by the other players. It is thus natural for a player to adopt a dominant strategy, if any such strategy exists. A game reaches a *dominant-strategy equilibrium* if each player has a dominant strategy. However this may not be possible, since in general a user's utility depends not only on her own strategy, but may be affected by other players' strategies. In such cases, one typically considers a weaker property called *Nash Equilibrium* (NE), which represents a strategy profile in which each player's utility is optimal, given that the other players also play their optimal strategies. We remark that if a dominant strategy equilibrium exists, then at least one Nash Equilibrium does.

2.2 The general model in anonymity systems

We consider an anonymity system of n members $\{1, \dots, n\}$ where n_h users are honest and the other $n_m (= n - n_h)$ are malicious. Each honest member i has a finite set of strategic actions \mathcal{Act}_i and we write S_i and U_i for respectively user i 's strategies and utilities. Here the U_i depends on several factors which we discuss below. Typically, the set of strategic actions for user i include actions \mathcal{C} and \mathcal{S} , respectively for *cooperative* and *selfish*. In this paper, we are only interested in these two actions, thus the set \mathcal{Act}_i is independent of i . For every user, we then denote \mathcal{Act}_i simply as $\mathcal{Act} = \{\mathcal{C}, \mathcal{S}\}$. Here we define \mathcal{C} as the behaviour of forwarding messages for any requests and \mathcal{S} as the behaviour of always refusing others' requests but only forwarding one's own messages. User i will be called *cooperative* or *selfish* user according to whether $S_i = \mathcal{C}$ or $S_i = \mathcal{S}$. Note that cooperative and selfish actions refer to the behaviour of honest users: in the paper, we make the standard assumption that malicious users always act cooperatively in order to be chosen on the honest users' paths and, so, to de-anonymise the system. We leave to future work the investigation of the case where attackers may act selfishly.

Definition 1. A game Γ with n_h players over anonymity systems of n members consists of a set of utility functions U_1, \dots, U_{n_h} , where $U_i; \mathcal{Act}^{n_h} \rightarrow \mathbb{R}$, the set of real numbers.

For each $i \in \{1, \dots, n_h\}$, the utility function $U_i(S_1, \dots, S_{n_h})$ describes the payoff to user i under each combination of strategies. We assume that the utility functions take the form of a linear combination of factor functions $\Phi_k(-)$, each accounting for a parameter k relevant to the specific application. That is, using $\rho_{ik} \geq 0$ to indicate the (relative) weight that user i attributes to parameter k , then $\sum_k \rho_{ik} = 1$ and

$$U_i(-) = \sum_k \rho_{ik} \cdot \Phi_k(-) \quad (1)$$

In the paper we will only use the factors of *anonymity*, *performance* and *cost*. The former quantifies the value a user attaches to their anonymity, whilst the second to the speed of their network activity. These parameters often need to be traded off against each other, as a higher anonymity level often requires more complex protocols which, as a side effect, reflect in longer delivery times. The ability to give them different weights in the *same* utility function allows i to select their individual strategy to finely balance their payoffs. Similarly, the ‘cost’ factor measures the importance that i attaches to any payments she incurs for using the anonymity network. We believe these three factors are the most important ones, and as such are sufficient to cover several significant applications, as in this paper; yet, additional factors can easily be included as required.

As honest users will in general vary their behaviour, and not always act according to a fixed strategy S_i , we shall use probabilities to describe the likelihood of i choosing each possible strategy. More precisely, in our context we assume that with probability x_i (resp. $1 - x_i$), user i will act cooperatively (resp. selfishly). Such randomness yields a so called *mixed strategy*. A mixed strategy is said *pure* if $x_i = 0$ or $x_i = 1$, i.e., when i in fact never varies her strategy.

Let $\mathbf{X} = [0, 1]^{n_h}$ be the set of all possible combinations of n_h honest users’ mixed strategies. Given a combination of mixed strategies $x = (x_1, \dots, x_{n_h}) \in \mathbf{X}$, we denote by x_{-i} the combination $(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_{n_h})$ of $n_h - 1$ mixed strategies obtained from it by removing i ’s, and, for a mixed strategy y , we let define

$$(x_{-i}; y) \triangleq (x_1, x_2, \dots, x_{i-1}, y, x_{i+1}, \dots, x_{n_h}),$$

which differs from x as user i switches from strategy x_i to y .

Since a user action is determined by its mixed strategy x_i , we rewrite his utility U_i as a function from \mathbf{X} to \mathbb{R} and define the notion of equilibrium as follows.

Definition 2. For Γ a game, a mixed strategy z is a ‘best response’ for user i to x_{-i} if

$$U_i(x_{-i}; z) \geq U_i(x_{-i}; y) \quad \text{for all mixed strategies } y.$$

A combination of strategies $x = (x_1, \dots, x_{n_h}) \in \mathbf{X}$ is a mixed Nash equilibrium if x_i is a best response to x_{-i} , for $i = 1, \dots, n_h$; the equilibrium is called a pure Nash Equilibrium if every x_i in x is a pure strategy.

Observe that definition above the just formalises the idea that no user can improve their own utility by unilaterally deviating from the mixed strategy combination x .

Following [12], we compute the Nash equilibrium by studying the players’ best-response correspondences. From Definition 2, i ’s utility maximisation problem is

$$\max_{x_i \in [0, 1]} U_i(x_{-i}; x_i).$$

3 CROWDS

For the reader’s convenience, we report a detailed description of the Crowds protocol [24]. In this section, we succinctly recall the fundamental mechanism of the protocol, and the related notion of probable innocence. We opt for the algorithmic description below, where $\text{aSend}(\mathbf{M}, \mathbf{D})$ represents the anonymous send of a message \mathbf{M} to a

destination D provided by CROWDS, $M \rightarrow D$ a standard communication, and $\{M\}_\kappa$ a link-encryption via a shared symmetric key κ , of which there exists one for each pair of participants in the protocol. The *forwarding probability* p_f is (together with n) the key parameter, as it determines the average length of the forwarding paths.

<pre> function aSend(M,D) begin $j := \text{Random_Pick}(\{1, \dots, n\})$ $\{\text{Relay}(M,D)\}_{\kappa_j} \rightarrow j$ end </pre>	<pre> function Relay(M,D) begin if(Flip_biased_coin(p_f)) $M \rightarrow D$ else $j := \text{Random_Pick}(\{1, \dots, n\})$ $\{\text{Relay}(M,D)\}_{\kappa_j} \rightarrow j$ endif end </pre>
--	--

Replies, if any, travel the path in reverse to reach the initiator. This is realised in the obvious way, whereby j sends any reply back to the user she received the corresponding Relay message from.

Reiter and Rubin have proposed in [24] a hierarchy of anonymity notions in the context of CROWDS. These range from ‘*absolute privacy*,’ where the attacker cannot perceive the presence of an actual communication, to ‘*provably exposed*,’ where the attacker can prove a sender-and-receiver relationship. Clearly, as most protocols used in practice, CROWDS cannot ensure absolute privacy in presence of attackers or corrupted users, but can only provide weaker notions of anonymity. In particular, in [24] the authors propose an anonymity notion called *probable innocence* and prove that, under some conditions on the protocol parameters, CROWDS ensures the probable innocence property to the originator. Informally, they define it as follows:

A sender is probably innocent if, from the attacker’s point of view, she appears no more likely to be the originator than to not be the originator. (2)

Since anonymity only makes sense for honest users, we define the set of anonymous events as $\mathcal{A} = \{a_1, a_2, \dots, a_{n_i}\}$, where a_i indicates that user i is the initiator of the message.

We assume that attackers will always deliver a request to forward immediately to the end server, since forwarding it any further cannot help them learn anything more about the identity of the originator. Thus in any given path, there is at most one detected user: the first honest member to forward the message to a corrupt user. Therefore we define the set of observable events as $\mathcal{O} = \{o_1, o_2, \dots, o_{n_i}\}$, where o_j indicates that user j forwarded a message to a corrupt user. In this case we also say that j is *detected* by the attacker. Halpern and O’Neill in [13] formalised condition (2) mathematically as:

$$P(a_i | o_j) \leq \frac{1}{2} \quad \text{for all } i, j. \quad (3)$$

Also, it was proved in [24] as one of the fundamental properties of the framework that, under the assumption that each honest user is equally likely to initiate a transaction

(which we adopt in this paper too), probable innocence (2) holds if and only if

$$n \geq \frac{P_f}{p_f - 1/2} (n_m + 1) \quad \text{and} \quad p_f \geq \frac{1}{2} \quad (4)$$

We remark that the concept of probable innocence was recently generalised in [14] to encompass the frequent situations where attackers have extra knowledge on users. The idea formalised in [14] is that the gain obtained by the attacker by observing an event must be relative to the knowledge that the attacker has of the users independently of that acquired through the protocol (whence the attribute ‘extra’). The authors express the extra information in terms of a random variable S with observable values $s_1 \dots s_\ell$, and the conditional probabilities $p(s_k | a_i)$. Probable innocence in presence of extra information can then be expressed by the condition:

$$P(a_i | o_j, s_k) \leq \frac{1}{2} \quad \text{for all } i, j, k. \quad (5)$$

4 Cooperation Analysis in Crowds

We now specialise our general game model to the setting of Crowds. Our assumptions identify a tractable yet realistic case for us to analyse how different utility factors affect the users’ cooperation behaviour.

Honest users. In Crowds, paths are static, and each user creates only one path per time period. At the end of the period, all existing paths are destroyed, and a new session starts where each user creates a new path for her anonymous communications. The reason for that is that dynamic paths tend to decrease the overall system anonymity. Therefore, we assume that honest users play the following mixed strategies game: (1) at the beginning of each session, each player i chooses her strategy by flipping a coin governed by her mixed strategy x_i , and then acts accordingly for the entire session;³ (2) selfish users will only cooperate to route their own messages. Observe that this is a reasonable assumption in Crowds, since messages are received in cleartext, and can therefore be recognised by their originators.

Attackers. As stated earlier, attackers will always cooperate. Moreover, we assume they do not originate messages, as they focus on de-anonymising the system.⁴ Finally, we assume that in their attempt to guess the identity of the initiator, the attackers always bet on the previous user on the path, the so-called detected user, because the latter is the most likely initiator (cf. Proposition 4). That is why we measure the anonymity degree of i against such attackers via the popular metric $P(a_i | o_i)$, as expressed in (3) above.

4.1 The Anonymity Analysis

Since the purpose of joining an anonymity system is to enjoy anonymous communications, the anonymity payoff is typically a very important factor for honest users. Each

³ We plan to investigate in future work the case when users may flip their behaviour at each interaction, by resorting to more advanced notions from game theory.

⁴ We leave to future work the case where attackers may flood part of the network to break some users’ anonymity or to perform DOS attacks; for recent work related see, e.g., [10, 4, 25].

cooperative user contributes to provide anonymity to all users, including herself. We first focus on how cooperation affects the overall anonymity of the system, and then investigate the anonymity payoffs of individual users.

During the creation of a path initiated by user i , both i and the malicious users will forward i 's message with probability 1, while a generic honest user j will do so with probability x_j . Thus, i has on average η_i users to pick from for a path, where

$$\eta_i = 1 + \sum_{j \neq i} x_j + n_m,$$

and $\zeta_i = \eta_i - n_m$ of these are honest. We can then prove that Reiter and Rubin's condition (3) to ensure probable innocence to the initiator i becomes as follows.

Proposition 1. *Let \bar{x}_i be the average cooperation probability of users other than i , viz., $(\sum_{j \neq i} x_j)/(n_h - 1)$, and assume $p_f > 1/2$ and that i is cooperative, i.e., $x_i = 1$. Then, i has probable innocence against n_m malicious users if and only if*

$$n \geq \frac{p_f + (1 - \bar{x}_i)/2\bar{x}_i}{p_f - 1/2} (n_m + 1).$$

Proposition 1 can be proved similarly to condition (4), with respect to which it expresses a more stringent constraint on n : indeed, as the honest users may behave less cooperatively, there more users are required in the system altogether to guarantee probable innocence against the same number of malicious users.

4.2 Measuring anonymity payoffs

Let $A_i(\cdot)$ denote i 's anonymity payoff function, whose value can be computed using the anonymity degree metric $P(a_i | o_j)$, as a function of the honest users' mixed strategies. Since the lower the anonymity degree, the better the anonymity guaranteed, we define $A_i(\cdot)$ to be $(1 - P(a_i | o_j))a$, where the parameter a can be used to normalise the value of a 'unit' of anonymity and $a \geq 0$. Let us start by evaluating the probability $P(o_j | a_i)$, for which we obtain the following result.⁵

$$\text{Proposition 2. } P(o_j | a_i) = \begin{cases} \frac{n_m}{\eta_i} + \frac{n_m p_f}{\eta_i(\eta_i - \zeta_i p_f)} & i = j \\ \frac{n_m p_f}{\eta_i(\eta_i - \zeta_i p_f)} x_j & i \neq j \end{cases}$$

Note that $P(o_j | a_i)$ does not depend on x_i , that when i is the initiator, the probability of detecting user j is not influenced by i 's strategy. This is because that no matter what strategy i chooses, she will forward her own messages with probability 1.

Now, let us compute the probability of detecting a user $P(o_j)$. Assuming a uniform distribution for anonymous events a_i , the following results hold.

$$\text{Proposition 3. } P(o_j) = \frac{1}{n_h} \left(\frac{n_m}{\eta_j} + \frac{n_m p_f}{\eta_j(\eta_j - \zeta_j p_f)} + \sum_{k \neq j} \frac{n_m x_k p_f}{\eta_k(\eta_k - \zeta_k p_f)} \right).$$

⁵ Due to space limitations, proofs are omitted from §4 and §5, and reported for the reader's convenience respectively in Appendix B and D.

Proposition 4. $P(a_i | o_j)$ can be expressed as $P(o_j | a_i)P(a_i)/P(o_j)$ from Proposition 2 and 3.

It is easy to show that $P(a_i | o_i)$ is a decreasing function of x_i and that, in particular, $P(a_i | o_i) = 1$ when $x_i = 0$. Therefore, a fully selfish user has zero anonymity degree.

Corollary 1. $\frac{\partial P(a_i | o_i)}{\partial x_i} \leq 0$.

Since $A_i(\cdot)$ is a decreasing function of $P(a_i | o_i)$, we have $\frac{\partial A_i(x)}{\partial x_i} \geq 0$. Therefore, when anonymity is the sole value taken into account, cooperation is the dominant strategy in CROWDS. Why is then the case that in the real world users often opt for the selfish behaviour? In the following sections, we shall explain this apparent mismatch by investigating the impact on users' behaviour of cost factor.

4.3 The Cost Analysis

To fulfill the forwarding demands of CROWDS, user i incurs a cost $C_i(\cdot)$ and which can be evaluated as

$$C_i(\cdot) = C_{i0} + \sum_{j \leq n_h} C_{ij},$$

where C_{i0} is a fixed cost. i.e., incurred whether or not i is involved in any communication, and C_{ij} is the cost incurred for forwarding messages from j .

In the CROWDS protocol, the expected length of a path is

$$E(L) = \frac{p_f}{1 - p_f} + 2.$$

Each path starts with the initiator while the last node is occupied either by a honest user or by an attacker. The path's internal nodes can only be honest users, because once a malicious user is encountered, the previous user is detected and the path terminated. The expected number of internal nodes is $E(L) - 2$. We can then evaluate the average number of times i appears on her own paths as

$$1 + \frac{E(L) - 2}{\zeta_i} + \frac{1}{\eta_i} = 1 + \frac{p_f/(1 - p_f)}{\zeta_i} + \frac{1}{\eta_i}. \quad (6)$$

Similarly, the average number of times i appears on other users' paths is:

$$\frac{p_f \cdot x_i/(1 - p_f)}{\zeta_j} + \frac{1 \cdot x_i}{\eta_j}.$$

Let us now define τ_i as i 's network traffic, i.e., the number of the messages sent by i , and c as the cost of forwarding each single one of them. Assuming that all users will incur the same cost c , we can compute the cost of forwarding by summing up the two cost components above and C_{i0} .

Proposition 5. $C_i(x) = C_{i0} + \left(1 + \frac{p_f/(1-p_f)}{\zeta_i} + \frac{1}{\eta_i}\right)\tau_i c + \sum_{j \neq i} \left(\frac{p_f \cdot x_i/(1-p_f)}{\zeta_j} + \frac{1 \cdot x_i}{\eta_j}\right)\tau_j c$

An immediate consequence is that the x_i derivative of cost is greater than zero.

Corollary 2.
$$\frac{\partial C_i(x)}{\partial x_i} = \sum_{j \neq i} \left(\frac{p_f}{1-p_f} \frac{(1 + \sum_{k \neq i, j} x_k)}{\zeta_j^2} + \frac{1 + n_m + \sum_{k \neq i, j} x_k}{\eta_j^2} \right) \tau_j c \geq 0.$$

Clearly, an increase in cooperation level will result in an increase in cost. Thus, if only the cost factor is considered, the dominant strategy in CROWDS is to behave selfishly.

4.4 Balancing between Cost and Anonymity in CROWDS

In this section, we apply our game-theoretic model to the CROWDS protocol when users considers both cost and anonymity factors at the same time. We substitute the cost and anonymity from Propositions 4 and 5 in our utility function, and assume that the normalisation factor a and c are such to put both utility factors on a same scale.

$$U_i(\cdot) = -\rho_{iC} \left[C_{i0} + \left(1 + \frac{p_f/(1-p_f)}{\zeta_i} + \frac{1}{\eta_i} \right) \tau_i c + \sum_{j \neq i} \left(\frac{p_f \cdot x_i / (1-p_f)}{\zeta_j} + \frac{x_i}{\eta_j} \right) \tau_j c \right] + \rho_{iA} \left[1 - \frac{\frac{n_m}{\eta_i} + \frac{n_m p_f}{\eta_i (\eta_i - \zeta_i p_f)}}{\frac{n_m}{\eta_i} + \frac{n_m p_f}{\eta_i (\eta_i - \zeta_i p_f)} + \sum_{k \neq i} \frac{n_m x_i p_f}{\eta_k (\eta_k - \zeta_k p_f)}} \right] a. \quad (7)$$

Differently from the cases of the anonymity and cost utilities, to find the equilibrium points for $U_i(\cdot)$ appears to be hard, although we know that there always exists in mixed strategies games. Therefore, in order to illustrate the effect on a user's utility of the combination of the two factors, we resort to simulation techniques, focussing on relevant parameters such as, the user strategy x_i , her choice of factor weights ρ_{iC} , and the number of the attackers in the system. The results are illustrated and discussed below.

In the following simulation, we consider $n_h = 100$, $p_f = 0.8$, $c = 0.1$, $C_{i0} = 5$ and $a = 100$. We assume that the cooperation level for users other than i is uniformly distributed and in the range of $[0, 1]$.

Factors' weights. We first show how the weights of the anonymity and cost factors influence i 's strategies. Figure 1 shows $U_i(\cdot)$ as a function of i ' strategy x_i , when the weight attributed to cost varies from 0 to 1. Figure 1b represents the projection of Figure 1a's surface onto the x_i axis, for eleven selected values of ρ_{iC} (from 0 to 1 in 1/10 steps); for each such projection π , Figure 1c plots the value of x_i which maximises π , which attempts to visualise the process of choosing the strategy for i .

Observe that as ρ_{iC} increases from 0 to 1, the equilibrium points x_i decreases: a bias towards anonymity leads to a higher cooperation level for user i .

Number of malicious users. We perform a similar analysis as above (but due to lack of space, in the rest of the paper we put the figures in the Appendix). This confirms that more malicious users result in smaller utilities for i , as i 's anonymity payoffs decrease substantially. In particular, when n_m is equal to 10 and 40 respectively, the maximum

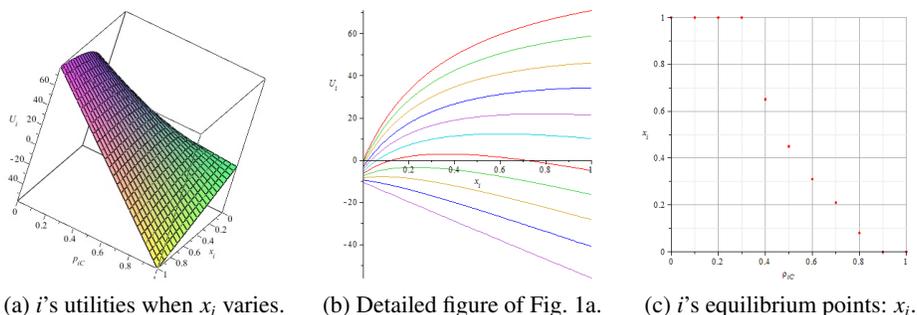


Fig. 1: $U_i(\cdot)$ as x_i and ρ_{iC} vary; ($n_m = 10, \tau = 50$).

utility occurs at $x_i = 0.5$, and $x_i = 0.68$ respectively. The values of x_i on equilibrium points are in the range of $[0.4, 0.7]$. Thus the number of malicious users in the system has a minor impact in encouraging or dissuading honest users to behave cooperatively.

Light traffic vs heavy traffic. Regarding the influence of network traffic, in our simulations as the number τ of messages increases from 5 to 80, the utility $U_i(\cdot)$ decreases considerably, as i incurs a heftier cost. The impact on the value of x_i at the equilibrium points is also significant, covering the interval $[0.35, 1]$. Thus, light traffic encourages the honest users to behave cooperatively more often, whilst heavy traffic pushes them towards selfishness.

Cooperation levels of the honest users other than user i . Here we let \bar{x}_i , the average cooperation level of the users other than i , vary from 0 to 1. We find that when the average \bar{x}_i is small, i will tend to behave selfishly to gain more payoff. When instead \bar{x}_i increases, the values of x_i on equilibrium points increase as well. Thus cooperative behaviour of the honest users encourages more cooperative behaviour.

In conclusion, we see that when a user, interested in both anonymity and cost, wants to optimise her utility, she needs to adapt her level of cooperation constantly, as the network topology (e.g., the number of cooperating users and attackers), the traffic level and her own choice of weight factor vary. As cost tend to be a very tangible a value, we can reasonably conclude that it will be a prominent factor for most users. It is therefore very important for anonymity systems to contemplate incentives mechanisms designed to offer tangible benefits to cooperative users. The next section is dedicated to the analysis of the effectiveness of one such mechanism.

5 Adding Incentives: the Analysis of Gold-star Mechanism

The gold-star mechanism was introduced in Tor [23] to encourage users to act as cooperative relays, and thus enhance the service performance for well-behaved forwarders. We now turn to the *gold-star* incentive mechanism [23] in CROWDS. A request from a user carrying a ‘gold star’ is given higher priority by other users, i.e., it is always relayed

ahead of other traffic. The assignment of gold-star status, in the context of CROWDS, is ruled by the policy “*to have satisfactory cooperation for R times out of the last V measurements.*” In accordance with our game in the setting of CROWDS, defined in §4, we assume a measurement is made in each session, and therefore a user obtain a gold-star status if and only if she cooperated in R sessions out of the last V sessions. Let $r = R/V$ be the above ratio. Then a user i will be awarded the gold-star if x_i is greater than or equal to r . We assume that each cooperative user will give priority to gold-star messages even if she is not a gold-star user. There exist mechanisms in the literature to allow anonymous users to accurately and securely report their interactions with their neighbours, whose description is beyond the scope of this paper. Such mechanism will help enforce the gold-star mechanism. Finally, as in its original proposal [23], we assume that the gold-star status are publicly known.

5.1 The Anonymity Analysis with Gold-star Mechanism

In presence of gold-star mechanism, attackers have an extra information about the initiator due the fact that gold-star status are public. Therefore, we use the anonymity metric encompassing extra knowledge via the conditional probability $P(a_i | o_j, s_l)$, as expressed in (5). Here $s_l \in \{s_1, s_2\}$, where $l = 1$ when the message is a gold-star one, and $l = 2$ otherwise.

The correlation between a message status and its initiator, that is the probability $P(s_l | a_i)$ is as follows.

$$P(s_1 | a_i) = \begin{cases} 1 & x_i \geq r, \\ 0 & x_i < r, \end{cases} \quad P(s_2 | a_i) = \begin{cases} 0 & x_i \geq r, \\ 1 & x_i < r, \end{cases} \quad (8)$$

Now since for each initiator O and S are independent, from [14], we have

$$P(a_i | o_j, s_l) = P(a_i | o_j) \frac{P(s_l | a_i)}{P(s_l | o_j)}.$$

If all honest users are equally likely to initiate a transaction, the probability $P(a_i | o_j, s_l)$ can be rewritten as follows.

Proposition 6.

$$P(a_i | o_j, s_1) = \begin{cases} \frac{P(o_j | a_i)}{\sum_{x_k \geq r} P(o_j | a_k)} & x_i \geq r, \\ 0 & x_i < r, \end{cases} \quad P(a_i | o_j, s_2) = \begin{cases} 0 & x_i \geq r, \\ \frac{P(o_j | a_i)}{\sum_{x_k < r} P(o_j | a_k)} & x_i < r, \end{cases}$$

Now we can prove that the presence of gold-star mechanism reduces the anonymity level of the network. The following indeed holds.

Corollary 3. $P(a_i | o_j, s_l) \geq P(a_i | o_j)$.

We define GS as the set of users j who have gold-star, i.e., $x_j \geq r$, and GS^c as the complement set of those who do not, $x_j < r$. If user i is the only element in either set, then there is no anonymity guaranteed for i , given malicious users are on the path.

We again define the anonymity payoffs as $A_i(-) = (1 - P(a_i | o_i, s_i))a$, the anonymity payoffs for unincentivised CROWDS and gold-star CROWDS are respectively evaluated as

Proposition 7. – *Unincentivised CROWDS (cf. Proposition 4):*

$$A_i(-) = \left(1 - \frac{\frac{n_m}{\eta_i} + \frac{n_m p_f}{\eta_i(\eta_i - \zeta_i p_f)}}{\frac{n_m}{\eta_i} + \frac{n_m p_f}{\eta_i(\eta_i - \zeta_i p_f)} + \sum_{k \neq i} \frac{n_m x_i p_f}{\eta_k(\eta_k - \zeta_k p_f)}} \right) a$$

– *Gold-star CROWDS:*

$$A_i(-) = \left(1 - \frac{\frac{n_m}{\eta_i} + \frac{n_m p_f}{\eta_i(\eta_i - \zeta_i p_f)}}{\frac{n_m}{\eta_i} + \frac{n_m p_f}{\eta_i(\eta_i - \zeta_i p_f)} + \sum_{k \neq i, k \in \phi(i)} \frac{n_m x_i p_f}{\eta_k(\eta_k - \zeta_k p_f)}} \right) a$$

where $\phi(i) = GS$ if $i \in GS$, $\phi(i) = GS^c$ otherwise.

For the unincentivised CROWDS, since $\frac{\partial A_i(-)}{\partial x_i} \geq 0$, behaving cooperatively will bring i maximum anonymity payoffs. However in gold-star CROWDS, $A_i(-)$ also depends on the number of users in the set which i belongs to. More users in such set leads to better anonymity provided for i . When CROWDS starts out with a small number of gold star relays, i has to behave selfishly more often in order not to be rewarded gold-star, and hence gains more anonymity payoffs.

$A_i(-)$ is an increasing function depending on x_i in the following two ranges:

- if $0 \leq x_i < r$, then $\frac{\partial A_i(-)}{\partial x_i} \geq 0$;
- if $r \leq x_i \leq 1$, then $\frac{\partial A_i(-)}{\partial x_i} \geq 0$.

We observe that $A_i(-)$ is a discontinuous function, with a discontinuity at $x_i = r$. Thus maximum points in the two ranges above occur at the extremes, $x_i = r$, and $x_i = 1$, respectively, and the equilibrium behaviour of i will ultimately depend on which of these is larger. If $A_i(x_{-i}; r) \geq A_i(x_{-i}; 1)$, then i will behave according to $x_i = r$ to reach the tipping point and gain the gold-star. If instead $A_i(x_{-i}; r) \leq A_i(x_{-i}; 1)$, then the dominant strategy for i is cooperative.

5.2 The Performance Analysis

We will use $P_i(x_{-i}; 1)$ and $P_i(x_{-i}; 0)$ to denote i 's performance payoffs when she behaves cooperative or selfish, respectively. Thus, the expected performance payoff for i can be evaluated as

$$P_i(-) = P_i(x_{-i}; x_i) = x_i P_i(x_{-i}; 1) + (1 - x_i) P_i(x_{-i}; 0) \quad (9)$$

Following [23], the factor P_i here is interpreted as *forwarding time* T_i : the shorter the forwarding time, the better the system performance. Thus we assume $P_i(\cdot) = (-T_i)p$ where p represents the benefit of each unit of performance.

For the reader's convenience, we summarise here the notation and names we shall be using in the analysis in the rest of the section.

- b : the size of messages sent by initiators;
- $f_j^{\mathcal{C}}$: the number of messages waiting at position j in the forwarding path, when the forwarder at j has strategy cooperative ;
- $f_j^{\mathcal{C}-\mathcal{S}}$: as above, but excepting the messages sent by selfish users;
- n_{ri} : the number of forwarders on the path i initiated;
- Q : the bandwidth of each user. Here we assume each user has the same bandwidth.

The total forwarding time T_i for user i is equal to the sum of the forwarding times of all $n_{ri} + 1$ nodes in user i 's path. We start by evaluating the expected forwarding time for cooperative users. The first relay is i herself and the forwarding time t_1 is

$$t_1 = \frac{b(f_1^{\mathcal{C}} + 1)}{Q}. \quad (10)$$

The forwarding time of cooperative users at position j th can be computed the same as Eq. 10. The cooperative users appear in the path with the probability $p(j, \mathcal{C}) = 1$ in that selfish users will not forward the messages initiated by the users other than themselves, thus

$$t_j = t_j(\mathcal{C}) p(j, \mathcal{C}) = \frac{b(f_j^{\mathcal{C}} + 1)}{Q} \times 1.$$

Given that the total forwarding time T_i is equal to $t_1 + \sum_{j=2}^{n_{ri}+1} t_j$, we can express the performance payoffs of cooperative users as follows, where we denote by \emptyset the 'no-incentive' mechanism

$$P_{\emptyset i}(x_{-i}; 1) = -\left(t_1 + \sum_{j=2}^{n_{ri}+1} t_j\right) \cdot p = -\left[\frac{b(f_1^{\mathcal{C}} + 1)}{Q} + \sum_{j=2}^{n_{ri}+1} \frac{b(f_j^{\mathcal{C}} + 1)}{Q}\right] p \quad (11)$$

The payoff function of the selfish user strategy can be evaluated along similar lines; the resulting formula is shown in (12).

$$P_{\emptyset i}(x_{-i}; 0) = -\left(\frac{b}{Q} + \sum_{j=2}^{n_{ri}} \left(\frac{\zeta_i - 1}{\zeta_i} \frac{b(f_j^{\mathcal{C}} + 1)}{Q} + \frac{1}{\zeta_i} \frac{b}{Q}\right) + \frac{\eta_i - 1}{\eta_i} \frac{b(f_j^{\mathcal{C}} + 1)}{Q} + \frac{1}{\eta_i} \frac{b}{Q}\right) p \quad (12)$$

We now turn to the *gold-star* incentive mechanism [23] in CROWDS. The development is similar to that in the above computations. The messages marked with a gold star, sent by cooperative users, have higher priority. There are then $b f_j^{\mathcal{C}-\mathcal{S}}$ KB rather than $b f_j^{\mathcal{C}}$ KB before i 's requests at the j th position of the path. Let \star denote the gold-star mechanism, we then evaluate $P_{\star i}(x_{-i}; 1)$ as

$$P_{\star i}(x_{-i}; 1) = -\left(t_1 + \sum_{j=2}^{n_{ri}+1} t_j\right) p = -\sum_{j=1}^{n_{ri}+1} \frac{b(f_j^{\mathcal{C}-\mathcal{S}} + 1)}{Q} p \quad (13)$$

The performance payoff of selfish strategy is shown below in Eq. (14).

$$P_{\star i}(x_{-i}; 0) = - \left(\frac{b}{Q} + \sum_{j=2}^{n_{ri}} \left(\frac{\zeta_i - 1}{\zeta_i} \frac{bf_j^c + b}{Q} + \frac{1}{\zeta_i} \frac{b}{Q} \right) + \frac{\eta_i - 1}{\eta_i} \frac{bf_j^c + b}{Q} + \frac{1}{\eta_i} \frac{b}{Q} \right) p \quad (14)$$

Since the utility $P_i(\cdot)$ of a fixed user depends on factors that may change very often at different forwarding paths, we instead consider the average payoffs $P_i(x_{-i}; 1)$ and $P_i(x_{-i}; 0)$ of cooperative and selfish users, respectively for each forwarding path. We define $\overline{f^c}$ (resp. $\overline{f^{c-s}}$) as the average number of messages (resp. gold-star messages) waiting for a cooperative user. We also define the average n_{ri} as n_r .

According to Eq. (9), we have the performance payoffs for i in unincentivised CROWDS and gold-star CROWDS as

Proposition 8.

$$\begin{aligned} P_{\emptyset i}(x_{-i}; x_i) &= - \frac{bp\overline{f^c}x_i}{Q} \left(n_r \left(1 - \frac{\zeta_i - 1}{\zeta_i} \right) + \frac{\zeta_i - 1}{\zeta_i} + \frac{1}{\eta_i} \right) \\ &\quad - \frac{bp}{Q} \left(n_r + 1 + \overline{f^c} \left(\frac{(n_r - 1)(\zeta_i - 1)}{\zeta_i} + \frac{\eta_i - 1}{\eta_i} \right) \right) \\ P_{\star i}(x_{-i}; x_i) &= \frac{bp x_i}{Q} \left(\overline{f^c} \left(\frac{(n_r - 1)(\zeta_i - 1)}{\zeta_i} + \frac{\eta_i - 1}{\eta_i} \right) - \overline{f^{c-s}}(n_r + 1) \right) \\ &\quad - \frac{bp}{Q} \left(n_r + 1 + \overline{f^c} \left(\frac{(n_r - 1)(\zeta_i - 1)}{\zeta_i} + \frac{\eta_i - 1}{\eta_i} \right) \right) \end{aligned}$$

Let α define the ratio $\overline{f^{c-s}}/\overline{f^c}$. It actually represents the percentage of users who have gold star among the all honest users, $0 \leq \alpha \leq 1$ in that if α is relatively small, then it reflects not many users are rewarded the gold-star. Then we study the x_i derivative of the performance payoffs, we have

Proposition 9. – *Unincentivised CROWDS*

$$\frac{\partial P_{\emptyset i}(x_{-i}; x_i)}{\partial x_i} \leq 0$$

– *Gold-star CROWDS*: if $\alpha \leq \frac{n_r - 1}{n_r + 1} \frac{\zeta_i - 1}{\zeta_i} + \frac{1}{n_r + 1} \frac{\eta_i - 1}{\eta_i}$ holds, then

$$\frac{\partial P_{\star i}(x_{-i}; x_i)}{\partial x_i} \geq 0.$$

5.3 Balancing between Performance and Anonymity

By applying our game-theoretic model to the gold-star CROWDS, we consider anonymity and performance factors in this section. We substitute the anonymity and performance payoff equations Proposition 7 and 8 to our utility function, we obtain:

– Unincentivised CROWDS:

$$U_i(-) = \rho_{iA} \left(1 - \frac{\frac{n_m}{\eta_i} + \frac{n_m p_f}{\eta_i(\eta_i - \zeta_i p_f)}}{\frac{n_m}{\eta_i} + \frac{n_m p_f}{\eta_i(\eta_i - \zeta_i p_f)} + \sum_{k \neq i} \frac{n_m x_k p_f}{\eta_k(\eta_k - \zeta_k p_f)}} \right) a$$

$$+ \rho_{iP} \left(-\frac{b p f^c x_i}{Q} \left(n_r (1 - \frac{\zeta_i - 1}{\zeta_i}) + \frac{\zeta_i - 1}{\zeta_i} + \frac{1}{\eta_i} \right) \right.$$

$$\left. - \frac{b p}{Q} \left(n_r + 1 + f^c \left(\frac{(n_r - 1)(\zeta_i - 1)}{\zeta_i} + \frac{\eta_i - 1}{\eta_i} \right) \right) \right)$$

– Gold-star CROWDS:

$$U_i(-) = \rho_{iA} \left(1 - \frac{\frac{n_m}{\eta_i} + \frac{n_m p_f}{\eta_i(\eta_i - \zeta_i p_f)}}{\frac{n_m}{\eta_i} + \frac{n_m p_f}{\eta_i(\eta_i - \zeta_i p_f)} + \sum_{k \neq i, k \in \phi} \frac{n_m x_k p_f}{\eta_k(\eta_k - \zeta_k p_f)}} \right) a$$

$$+ \rho_{iP} \left(\frac{b p x_i}{Q} \left(f^c \left(\frac{(n_r - 1)(\zeta_i - 1)}{\zeta_i} + \frac{\eta_i - 1}{\eta_i} \right) - f^{c-s} (n_r + 1) \right) \right.$$

$$\left. - \frac{b p}{Q} \left(n_r + 1 + f^c \left(\frac{(n_r - 1)(\zeta_i - 1)}{\zeta_i} + \frac{\eta_i - 1}{\eta_i} \right) \right) \right)$$

where $\phi(i) = GS$ if $i \in GS$, $\phi(i) = GS^c$ otherwise.

From the above equations, we derive the following proposition.

Proposition 10. *In gold-star CROWDS, if $\alpha \leq \frac{n_r - 1}{n_r + 1} \frac{\zeta_i - 1}{\zeta_i} + \frac{1}{n_r + 1} \frac{\eta_i - 1}{\eta_i}$ holds, then*

$$\frac{\partial U_i(-)}{\partial x_i} \geq 0 \text{ when } x_i \in [0, r); \quad \frac{\partial U_i(-)}{\partial x_i} \geq 0 \text{ when } x_i \in [r, 1],$$

and function $U_i(-)$ is discontinuous at the point $x_i = r$.

We run simulations to illustrate the equilibrium points of i strategies in different situations. We consider $n_h = 100$, $n_r = 3$, $b = 100$, $Q = 500$, $a = 100$ and $p = 1$. We assume the cooperation level for users j other than i is uniformly distributed and in the range of $[0, 1]$. We start with the case of **unincentivised CROWDS**.

Factors' weights. As we did in §4.4, we first show how the weights of the anonymity and performance factors influence i 's strategies. In our simulations, as the weight attributed to performance varies from 0 to 1, $U_i(-)$ varies as a function of i 's strategy x_i . As ρ_{iP} keeps increasing from 0 to 1, the equilibrium points x_i decrease. Higher weight towards anonymity factor leads to higher cooperation level of user i .

Number of malicious users. More malicious users result in smaller anonymity but greater performance payoffs for i . In particular, when n_m varies from 2 and 80, the values of x_i on equilibrium points are in the range of $[0.2, 0.43]$. Thus the number of malicious users has a minor impact in encouraging or dissuading cooperation behaviours of honest users.

Light traffic vs heavy traffic. In our simulations, as the average number $\overline{f^c}$ of messages waiting at the forwarders increases from 0 to 10, the utility $U_i(\cdot)$ decreases, so that i incurs more delivery time. The values of x_i on equilibrium points decrease as well, from 1 to 0.18. Thus, like before, light traffic encourages more frequent cooperation by honest users, while heavy traffic suggests them selfishness.

Cooperation levels of the honest users other than user i . Regarding the influence of users' behaviours, we find that when the average \bar{x}_j is small, i will be willing to behave cooperatively more often to gain more payoff. When \bar{x}_j increases, the values of x_i on equilibrium points decrease until 0.37. Thus, cooperative behaviours of the honest users other than i do not encourage more cooperative behaviours of i .

Moving to analyse **Gold-star Crowds**, we consider the following parameter settings: $\rho_{iP} = 0.5$, $n_m = 10$, $\overline{f^c} = 3$, $\overline{f^{c-s}} = (1-r)\overline{f^c}$ in our simulations. When r varies as 0.8, 0.7 and 0.6 respectively, the equilibrium points for i are $x_i = 0.8$, $x_i = 0.7$ and $x = 1$, respectively. They are better than those in unincentivised Crowds.

Note that the first two x_i of equilibrium points are exactly the values of rule r . We find that before x_i increases to r , both the anonymity and performance payoffs increase as x_i increases. Then, when x_i exceeds r , the performance payoff increases further, whilst the anonymity payoff depends on the number of gold-star users. In this simulation, when i gets the gold-star, the anonymity payoff of i is smaller than that when i has not. This is because, since x_j is uniformly distributed and $r = 0.8$, the number of gold-star users is smaller than the number of users without gold-star. Thus by balancing with performance payoffs, i chooses her strategy as $x_i = r$. The third x_i is 1, because from r to 1 the performance payoff keeps increasing while the the anonymity payoff increases as well. On the point $x_i = 1$, the balanced payoffs ($U_i = 38.9139$) of performance and anonymity have exceed the previous maximum payoffs ($U_i(\cdot) = 34.6699$) where $x_i = 0.6$.

We do simulations by varying n_m , $\overline{f^c}$, and find the shapes of the lines are the same as those of the above simulations. Therefore, as we proved in Proposition 10, i 's strategy depends on the comparison of $U_i(\cdot)$ on $x_i = r$ and $x_i = 1$.

6 Related Work

The anonymity systems and protocols are typically based on a suitable infrastructure for forwarding messages. For instance, the Crowds protocol [24] requires a set of users or peers willing to route each others' requests. In other cases where they do not directly require 'forwarders' – as e.g. for single-hop web proxies like the *Anonymizer* protocol [3] – they rely on the obfuscation of network traffic provided by the activity of a (large) set of users. The number of users who forward requests or join in the infrastructure determines the anonymity degree of the systems. A point in case is the Tor

protocol [9]. Although Tor has built a significant community of volunteer forwarders, if at any time the user-to-relay ratio becomes too small, then all users will be affected, and will receive a lower-security service [23].

To address these problems, researchers in anonymity networks have considered mechanisms to incentivise cooperation [23, 16]. The ‘*gold star*’ mechanism in Tor [23] encourages users to act as cooperative relays, by enhancing the service performance for well-behaved forwarders. Indeed, relays which provide a good service to other users get the gold-star reward, and messages sent by gold-star holders are given higher relay priority, i.e. they are always relayed ahead of other traffic. Tor’s management algorithms routinely scans existing directory authorities to actively measure each user’s performance, and only grant the gold star where appropriate. Users of BRAIDS [16] anonymously ‘pay’ Tor relays with generic tickets according to the three hierarchical service classes. This allows the users to earn credits which they can redeem against improved traffic performance in Tor. Some mechanisms were proposed to encourage peers to act cooperatively in P2P systems. These include, e.g., payment schemes, which charge for anonymity services and/or reward good user behaviour; and reputation schemes, where users with higher reputation get higher-quality service. PAR [2] and XPay [7] are payment mechanisms: they produce monetary incentives by using e-cash and an online bank. Reputation schemes [8, 27, 28] use interaction histories to develop trust levels for members in the systems; this encourages trustworthy behaviour and incentives future cooperative behaviours. Some reputation schemes are however incompatible with anonymity systems, as relays cannot always link interactions to users.

In this paper we model user behaviours in the anonymity systems as non-cooperative games. In the context of network security, game-theoretic models have primarily been used to address problems related to free-riding in P2P systems [18] and distributed intrusion detection [5, 17, 26]. In [18], game theory has been used to characterise peer selfishness and provide incentives for peers to contribute their upload capacities. The work closest to ours is [1], where the authors study incentive systems for four types of users and in doing so lay the foundations for a game-theory approach to modelling anonymity infrastructures. Their model is based on mix-nets [6]. Although quite general, such model cannot accommodate the evaluation of specialised utility functions in the context of specific anonymity protocols. Each player in *loc. cit.* belongs to one of only four types (*viz.*, user, honest node, dishonest node and sender), and in each type all players behave uniformly. Finally, [1] assume that traffic is distributed uniformly across nodes, which clearly may not be a realistic assumption: e.g., in reputation-based networks users are obviously more likely to ask relays with high reputation as forwarders for the messages. In such cases, the anonymity degree differ for each user [25].

7 Conclusion

The effectiveness of anonymity networks depends heavily on the number of cooperative users. In this paper, we investigated the incentives for users to behave cooperative or selfish in such networks. We proposed a game theoretic framework and used it to analyse users’ behaviours and also predict what strategies users will choose under different circumstances and according to their exact balance of preferences among factors

such as anonymity, performance (message delivery time) and cost. To allow to trade-off against each others quantities as different as cost (measured in, say, dollars) and anonymity (measured in the interval $[0,1]$), the model uses multiplicative parameters (*viz.*, a and c) to map them to a common or standard scale. Significantly, we also used the model to assess the effectiveness of the gold-star incentive mechanism.

We studied the phenomenon that in the original Crowds protocol users have little incentive to act cooperatively beyond the minimum required to remain probabilistically anonymous, as cooperation incurs a cost which reflects in the user suffering a utility loss. We then investigated the effectiveness of gold-star mechanism when implemented in Crowds. We showed that the gold-star mechanism does create incentives for users to cooperate exactly when the performance incentives cover the cost of forwarding other users' messages. Depending on the amount of performance incentives, users will be willing to be cooperative all the time, or they will choose the Nash equilibrium mixed strategy, which gains them maximum utility. We observed that the mechanism can become de-anonymising when there are not enough gold-star users, as gold-star messages then carry a strong clue about their originators. To factor this in, our analysis used an anonymity measure which takes into account the attackers' extra knowledge, *i.e.*, the a priori knowledge they may acquire besides the protocol. In order for the mechanism to remain effective, the system must therefore enforce a minimum number of gold-star users in the system, and relax the condition for obtaining gold-star status when the threshold is not met. Observe that although the gold-star mechanism was conceived for Tor, in order to keep the paper self-contained, here we have formulated its concepts and mechanisms on Crowds. We expect no major difficulties to translate the present results back to Tor, by applying our game model to the relays, which in fact are the entities the gold-star mechanism was designed for. Also, we plan to validate our model of Tor by comparing the predictions made through it with existing results in the literature (*viz.*, the simulations in the original gold-star paper [23]). We believe that by restricting to relays, we can apply the game-theoretic framework to Tor, and the performance analysis can be translated back. We are currently working on mixed anonymity/performance/cost utilities in Tor, which appears to be more complex.

A cost model alternative to bandwidth is a relay's liability for the traffic emerging from it. This generates interesting issues. For instance it may create a deficit of exit nodes, as relay do not drop messages to save bandwidth, but keep relaying them to avoid liability for their delivery. This is tantamount to a user fiddling with the forwarding parameter p_f , and goes beyond the cooperative vs selfish choice. Our investigation in this paper focused on the cooperative/selfish forwarding behaviours in Crowds, where p_f is fixed and equal for all users. We leave the study of the important *liability* payoff for future work.

Modelling performance payoffs is useful in real-world scenarios, in that it allows researchers to make good and useable predictions with no or only minor resort to simulations. We believe that this is a significant contribution, as simulations can be taxing in terms of computational power as well as time.

Differently from previous work, our utility functions are not limited to anonymity aspects, but are composed of independently-configurable factors which allow us to model different types of users as well as adapt the model to specific applications. This

adds a good deal of flexibility to our game model. In particular, it makes the model more suitable for *asymmetric* anonymity systems, i.e., systems where users differ from each other by means of different payoffs and utility functions. Also, we employ our model to study the *Nash strategies* and *dominant strategies* for users in the game. We are not aware of previous work in this line.

In future work, we plan to refine the model and, more specifically, adapt the techniques presented here to *cooperative games* in the presence of irrational players and more complex utilities. This will allow us to take into account more kinds of attack, and model the fact that attacks targeted to specific users may definitely affect their utilities. For instance, a ‘denial-of-service’ attack will impact adversely the effectiveness of reputation-based incentive mechanisms, whilst the gold-star scheme will suffer from ‘intersection attacks’ on anonymity. We also plan to compare different anonymity systems, such as, CROWDS, onion routing, Mix-net in our model, to study which incentive schemes are better suited to each of them.

References

1. A. Acquisti, R. Dingledine, and P. F. Syverson. On the economics of anonymity. In R. N. Wright, editor, *Financial Cryptography*, volume 2742 of *Lecture Notes in Computer Science*, pages 84–102. Springer, 2003.
2. E. Androulaki, M. Raykova, S. Srivatsan, A. Stavrou, and S. M. Bellovin. Par: Payment for anonymous routing. In N. Borisov and I. Goldberg, editors, *Privacy Enhancing Technologies*, volume 5134 of *Lecture Notes in Computer Science*, pages 219–236. Springer, 2008.
3. The anonymizer. Available at <http://www.anonymizer.com>.
4. N. Borisov, G. Danezis, P. Mittal, and P. Tabriz. Denial of service or denial of security? In *Proceedings of the 14th ACM conference on Computer and communications security, CCS '07*, pages 92–102, New York, NY, USA, 2007. ACM.
5. R. Bye, K. Luther, S. A. Çamtepe, T. Alpcan, S. Albayrak, and B. Yener. Decentralized detector generation in cooperative intrusion detection systems. In T. Masuzawa and S. Tixeuil, editors, *SSS*, volume 4838 of *Lecture Notes in Computer Science*, pages 37–51. Springer, 2007.
6. D. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Commun. ACM*, 24(2):84–88, 1981.
7. Y. Chen, R. Sion, and B. Carbunar. Xpay: practical anonymous payments for tor routing and other networked services. In E. Al-Shaer and S. Paraboschi, editors, *WPES*, pages 41–50. ACM, 2009.
8. E. Damiani, S. D. C. di Vimercati, S. Paraboschi, P. Samarati, and F. Violante. A reputation-based approach for choosing reliable resources in peer-to-peer networks. In V. Atluri, editor, *ACM Conference on Computer and Communications Security*, pages 207–216. ACM, 2002.
9. R. Dingledine, N. Mathewson, and P. F. Syverson. Tor: The second-generation onion router. In *USENIX Security Symposium*, pages 303–320. USENIX, 2004.
10. R. Dingledine, V. Shmatikov, and P. Syverson. Synchronous batching: From cascades to free routes. pages 186–206, 2004.
11. M. J. Freedman and R. Morris. Tarzan: a peer-to-peer anonymizing network layer. In *ACM Conference on Computer and Communications Security*, pages 193–206, 2002.
12. D. Fudenberg and J. Tirole. *Game Theory*. MIT press, 1991.

13. J. Y. Halpern and K. R. O'Neill. Anonymity and information hiding in multiagent systems. *Journal of Computer Security*, 13(3):483–512, 2005.
14. S. Hamadou, C. Palamidessi, V. Sassone, and E. ElSalamouny. Probable innocence in the presence of independent knowledge. In P. Degano and J. D. Guttman, editors, *Formal Aspects in Security and Trust*, volume 5983 of *Lecture Notes in Computer Science*, pages 141–156. Springer, 2009.
15. N. Hopper, E. Y. Vasserman, and E. Chan-Tin. How much anonymity does network latency leak? *ACM Trans. Inf. Syst. Secur.*, 13(2), 2010.
16. R. Jansen, N. Hopper, and Y. Kim. Recruiting new tor relays with braids. In E. Al-Shaer, A. D. Keromytis, and V. Shmatikov, editors, *ACM Conference on Computer and Communications Security*, pages 319–328. ACM, 2010.
17. Y. Liu, C. Comaniciu, and H. Man. Modelling misbehaviour in ad hoc networks: a game theoretic approach for intrusion detection. *IJSN*, 1(3/4):243–254, 2006.
18. R. Ma, S. Lee, J. Lui, and D. Yau. A game theoretic approach to provide incentive and service differentiation in P2P networks. *ACM SIGMETRICS Performance Evaluation Review*, 32(1):189–198, 2004.
19. J. McLachlan, A. Tran, N. Hopper, and Y. Kim. Scalable onion routing with Torsk. In E. Al-Shaer, S. Jha, and A. D. Keromytis, editors, *ACM Conference on Computer and Communications Security*, pages 590–599. ACM, 2009.
20. S. J. Murdoch and G. Danezis. Low-cost traffic analysis of tor. In *IEEE Symposium on Security and Privacy*, pages 183–195. IEEE Computer Society, 2005.
21. A. Nambiar and M. Wright. Salsa: a structured approach to large-scale anonymity. In A. Juels, R. N. Wright, and S. D. C. di Vimercati, editors, *ACM Conference on Computer and Communications Security*, pages 17–26. ACM, 2006.
22. C. A. Neff. A verifiable secret shuffle and its application to e-voting. In *ACM Conference on Computer and Communications Security*, pages 116–125, 2001.
23. T.-W. Ngan, R. Dingledine, and D. S. Wallach. Building incentives into Tor. In R. Sion, editor, *Financial Cryptography*, volume 6052 of *Lecture Notes in Computer Science*, pages 238–256. Springer, 2010.
24. M. K. Reiter and A. D. Rubin. Crowds: Anonymity for web transactions. *ACM Trans. Inf. Syst. Secur.*, 1(1):66–92, 1998.
25. V. Sassone, S. Hamadou, and M. Yang. Trust in anonymity networks. In P. Gastin and F. Laroussinie, editors, *CONCUR*, volume 6269 of *Lecture Notes in Computer Science*, pages 48–70. Springer, 2010.
26. K. wei Lye and J. M. Wing. Game strategies in network security. *Int. J. Inf. Sec.*, 4(1-2):71–86, 2005.
27. L. Xiong and L. Liu. Peertrust: Supporting reputation-based trust for peer-to-peer electronic communities. *IEEE Trans. Knowl. Data Eng.*, 16(7):843–857, 2004.
28. R. Zhou and K. Hwang. Powertrust: A robust and scalable reputation system for trusted peer-to-peer computing. *IEEE Trans. Parallel Distrib. Syst.*, 18(4):460–473, 2007.

A The CROWDS protocol

Crowds is a protocol proposed by Reiter and Rubin in [24] to allow Internet users to perform anonymous web transactions by protecting their identities as originators of messages. The central idea to ensure anonymity is that the originator forwards the message to another, randomly-selected user, which in turn forwards the message to a third user, and so on until the message reaches its destination (the end server). This routing process ensures that, even when a user is detected sending a message, there is a substantial probability that she is simply forwarding it on behalf of somebody else.

More specifically, a crowd consists of a *fixed* number of users participating in the protocol. Some members (users) of the crowd may be corrupt (the *attackers*), and they collaborate in order to discover the originator's identity. The purpose of the protocol is to protect the identity of the message originator from the attackers. When an *originator*—also known as *initiator*—wants to communicate with a server, she creates a random *path* between herself and the server through the crowd by the following process.

- *Initial step*: the initiator selects randomly a member of the crowd (possibly herself) and forwards the request to her. We refer to the latter user as the *forwarder*.
- *Forwarding steps*: a forwarder, upon receiving a request, flips a *biased* coin. With probability $1 - p_f$ she delivers the request to the end server. With probability p_f she selects randomly a new forwarder (possibly herself) and forwards the request to her. The new forwarder repeats the same forwarding process.

The response from the server to the originator follows the same path in the opposite direction. Users (including corrupt users) are assumed to only have access to messages routed through them, so that each user only knows the identities of her immediate predecessor and successor in the path, as well as the server.

B Proof in §4

B.1 Proof of Proposition 2

Proof. Let k denote the position occupied by the first honest user preceding an attacker on the path, with the initiator occupying position zero. Let $P(o_j | a_i)_{(k)}$ denote the probability that user j is detected exactly at position k . Only the initiator can be detected at position zero, and the probability that this happens is equal to the overall probability that the initiator chooses a corrupt member as a forwarder. Therefore

$$P(o_j | a_i)_{(0)} = \begin{cases} \frac{n_m}{n_i} & i = j \\ 0 & i \neq j \end{cases}$$

Now the probability that j is detected at position $k > 0$ is given by

- the probability that she decides to forward k times and picks $k - 1$ honest users, i.e., $p_f^{k-1} (\frac{x_i}{n_i})^{k-1}$,
- times the probability of choosing j as the k th forwarder and j forwards the message with probability x_j i.e., $\frac{x_j}{n_i}$,

– times the probability that she picks any attacker at stage $k + 1$, i.e., $\frac{n_m}{\eta_i} p_f$.

Therefore

$$\forall k \geq 1, P(o_j | a_i)_{(k)} = p_f^k \left(\frac{\zeta_i}{\eta_i}\right)^{k-1} \frac{x_j}{\eta_i} \frac{n_m}{\eta_i}$$

and hence

$$P(o_j | a_i) = \begin{cases} \frac{n_m}{\eta_i} + \frac{n_m p_f}{\eta_i(\eta_i - \zeta_i p_f)} & i = j \\ \frac{n_m x_j p_f}{\eta_i(\eta_i - \zeta_i p_f)} & i \neq j \end{cases}$$

B.2 Proof of Equation 6

Proof. We can split the computation into three parts: the first, the middle and the last positions of the paths. For the first position, initiator i will appear on her own path with probability 1; For the inner positions, there are $p_f/(1 - p_f)$ positions which can only be occupied by either i or other honest users. Because user i chooses them with uniform probability, the average number of times i appears on the inner positions is evaluated as

$$\frac{p_f/(1 - p_f) \cdot 1}{\zeta_i};$$

As for the last position, the malicious users can occupy it as well. Since the malicious users will forward any messages with probability one, thus we have

$$\frac{1 \cdot 1}{\eta_i}.$$

This proves the formula.

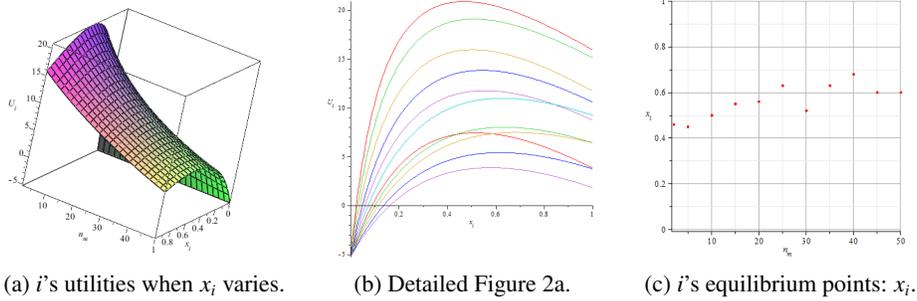


Fig. 2: $U_i(\cdot)$ as n_m and x_i vary; ($\rho_{iC} = 0.5, \tau = 50$).

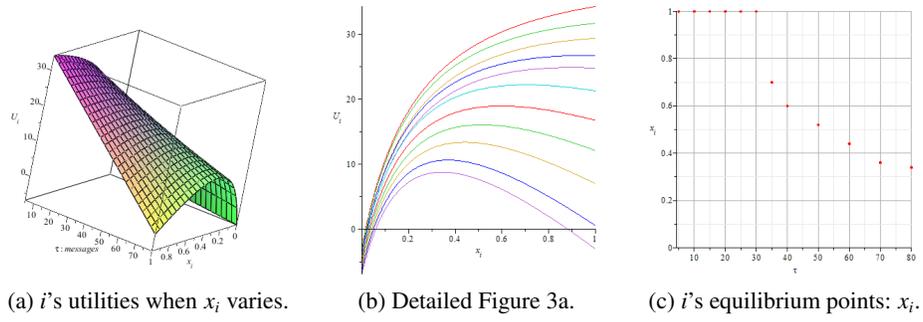


Fig. 3: $U_i(-)$ as τ and x_i vary; ($\rho_{iC} = 0.5$, $n_m = 10$).

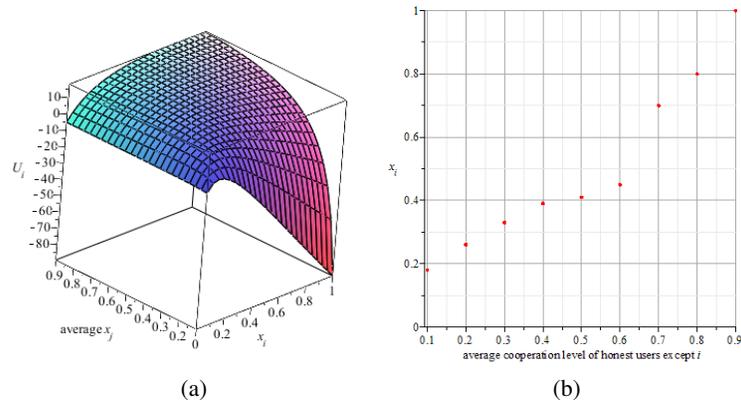


Fig. 4: $U_i(-)$ as \bar{x}_j and x_i vary; ($n_m = 10$, $\tau = 50$ and $\rho_{iC} = 0.5$).

C Figures in §4.4

See Figures 2, 3 and 4.

D Proof of Proposition 6

Proof.

$$\begin{aligned}
 P(a_i | o_j \wedge s) &= P(a_i | o_j) \frac{P(s | a_i)}{P(s | o_j)} \\
 &= P(a_i | o_j) \frac{P(s | a_i)}{\frac{\sum_k P(o_j \wedge s | a_k) P(a_k)}{P(o_j)}} \\
 &= \frac{P(o_j | a_i) P(a_i)}{P(o_j)} \frac{P(s | a_i) P(o_j)}{\sum_k P(o_j \wedge s | a_k) P(a_k)}
 \end{aligned}$$

If the honest members are equally likely to initiate a transaction in CROWDS, then

$$\begin{aligned}
 &= P(o_j | a_i) \frac{P(s | a_i)}{\sum_k P(o_j \wedge s | a_k)} \\
 &= P(o_j | a_i) \frac{P(s | a_i)}{\sum_k P(o_j | a_k) P(s | a_k)}
 \end{aligned}$$

By substituting Proposition 8 to the above equation, we obtain $P(a_i | o_j \wedge s)$.

E Figures in §5.3

See Figures 5–9.

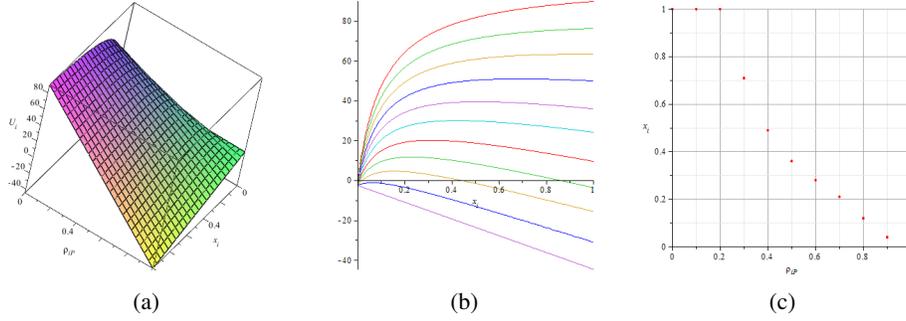


Fig. 5: $U_i(-)$ as ρ_{iP} and x_i vary; ($n_m = 10, \overline{f^c} = 3$).

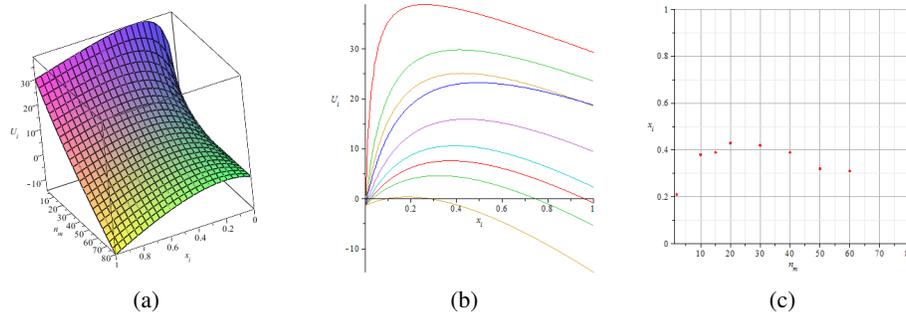


Fig. 6: $U_i(-)$ as n_m and x_i vary; ($\rho_{iP} = 0.5, \overline{f^c} = 3$).

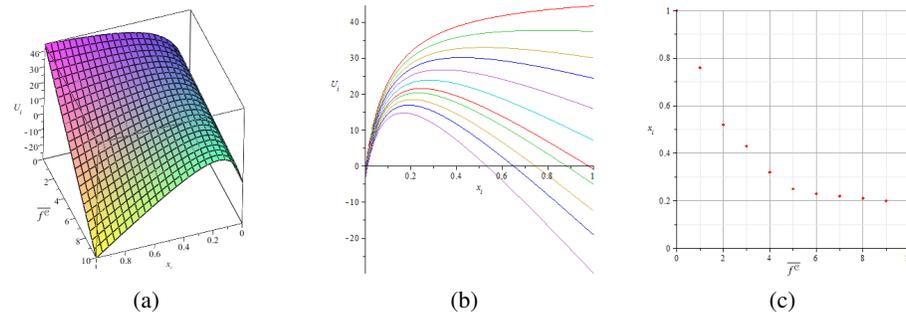


Fig. 7: $U_i(-)$ as $\overline{f^c}$ and x_i vary; ($\rho_{iP} = 0.5, n_m = 10$).

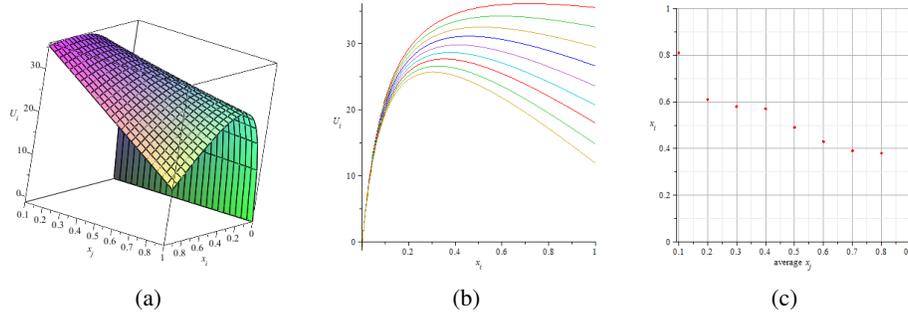


Fig. 8: $U_i(\cdot)$ as \bar{x}_j and x_i vary; ($\rho_{iP} = 0.5$, $n_m = 10$, $\overline{f^c} = 3$).

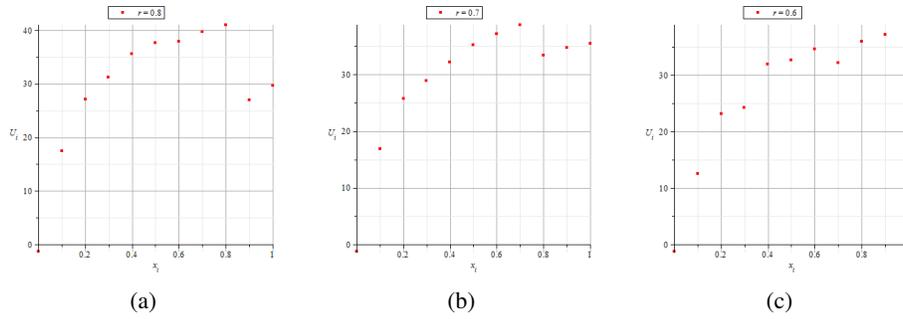


Fig. 9: $U_i(\cdot)$ as r and x_i vary; ($\rho_{iP} = 0.5$, $n_m = 10$, $\overline{f^c} = 3$).