MŰEGYETEM 1782

Budapest University of Technology and Economics
Faculty of Electrical Engineering and Informatics
Department of Networked Systems and Services

Anonymity and Altruism: Game-Theoretical Analysis of Fuzzy Message Detection

Scientific Students' Association Report

Author:

Marcell Frank

Advisor:

Dr. Gergely Biczók

2022

Contents

Abstract

Fuzzy Message Detection (FMD) provides a novel solution to the long-standing problem of costly hardware usage for anonymous communications. With FMD, users can outsource the reception and identification of their incoming messages efficiently to an untrusted third-party server, which forwards them when appropriate. To avoid malicious actors uncovering the relationship between the participants, users can set a false-positive rate on the server, i.e., how much cover traffic it should forward along with intended messages protecting recipient anonymity. Of course, this mechanism causes unwanted messages to appear on the user's end, thereby creating a need to balance one's own bandwidth cost with the anonymity of other users: essentially, the welfare of others.

In this report, we show that finding this balance is possible via careful modeling and by applying the perspective and tools of game theory. We show that fully selfish behavior renders FMD useless; the system needs at least a few altruistic users to operate properly. In fact, the question becomes: what is the desired number and individual characteristics of altruists in FMD-enabled systems? We implement an improved greedy algorithm running on multiple cores finding epsilon-Nash equilibria, i.e., efficient false-positive rate allocations, in real-world FMD-enabled systems modelled as graphs, and analyze both the amount and quality of altruistic nodes need to maintain recipient anonymity. FMD is a promising privacy-enhancing technology, foreseen to be utilized in many systems in the near-future: our results can serve as guidelines for system designers, reminding them to engineer proper individual economic incentives into their system to ensure the desired anonymity.

Chapter 1

Introduction

Metadata leakage is a serious problem when it comes to private messaging. People looking to ensure their own and their partners' confidentiality must go out of their way to set up a system where not only is it hard to eavesdrop on them, but a malicious third-party could hardly infer anything of value from their conversation. All these methods require extensive communication, knowledge regarding privacy-enhancing systems and extra computational power and bandwidth. A common solution to this problem is a trusted third party, a server can provide this service for example, where messages are stored, then forwarded at the appropriate time. It is often necessary to include the processing power of a third party in public key encryption, because in order to determine which messages are meant for the user, you have to examine every ciphertext (message identifier), and the download process imposes a considerable bandwidth cost. Outsourcing is easy, you only download the selected messages; however, this exposes the exact number of messages someone receives. Even a potentially simple system has many risk factors associated with a privacy breach, upon which, for example, someone's usage patterns and communication graph could be revealed. One of the most dangerous threats is hostile actors taking over the trusted server, and extracting massive amount of metadata. Fuzzy Message Detection (FMD) [4, 16] provides an easily implementable, easily understandable solution to the many problems outlined in these systems. Its key concept, an individually adjustable false-positive rate for every user, potentially makes the chosen system heavily resilient against recipient linkability, and provides sufficient relationship anonymity. With key generation creating little hardware overhead, FMD's applications can range from privacy-oriented blockchain transactions to messaging applications.

Problem statement. The concept of user-set false-positive rates enables a more streamlined user experience. People who do not frequent the system need not suffer much extra bandwidth cost imposed on them upon each message access, while core users or other message nodes can set this value relatively high in order to better mask their own and their partners' traffic. The problem, however, is intent. In an environment where every user is left to its own devices (meaning no cooperation), it has every reason to decrease its own cost, and zero to increase it (provided he is not altruistic, more on that later). This comes from the fact that your own privacy benefit is gained by other users downloading your messages, masking your interaction with the network; therefore, the user is free to tell the server that he only needs his own messages, since the privacy losses of other users does not affect his own utility. Intuitively, this system rapidly degrades to the point where no privacy is achieved, since no-one downloads anything besides his/her own messages. In this worst-case scenario, every user's false-positive rate is set to 0, and the system provides no protection whatsoever. This system, without altruism [3, 10], is inherently flawed. The only question is, under what circumstances does the user base achieve an outcome, where the collective welfare of participants of the system is higher than the worst-case scenario.

Our contribution. Our goal is to analyze this system from a game-theoretic perspective, prove that the system is only stable when a certain amount of altruism is present, and find that stable setup. Here, stable setup means a distribution of false-positive rates between the altruistic users, where we reach a point at which no user finds it rational to further adjust their own fp-rate. Our hypothesis is that these cases present a much better outlook for the system, with significantly better social welfare equilibria. We prove it by building a user model implementing a type of altruism, then running a best-response dynamics (BRD) algorithm, which outputs, depending on the setting, the Nash-equilibrium [12], or the social optimum of the given dataset, with the necessary fp-rate distribution of the users. With these, equilibrium descriptive properties, such as the Price of Anarchy [9] or Price of Stability [2] can be expressed. The system can be tested with various altruism settings, changeable message and privacy breach costs, and with adjustable uniform starter fp-rates to make convergence easier. Furthermore we give a brief overview of other possible avenues regarding the upgrade of the model, or applying it to different privacy-oriented systems.

Structure of this paper. In the first part of Chapter 2, we give a brief overview of Fuzzy Message Detection, then introduce the relevant game theory definitions. In the second part we present papers directly relevant to FMD, while also presenting others with relevancy to the problems in this article. In Chapter 3 we present the the game model used to solve the problems outlined in this paper. Chapter 4 presents the technical implementation of the best response dynamics, and presents its results. To wrap it up, Chapter 5 gives a brief summary of the results, while noting the limitations of the used model, and introducing concepts for future works.

Chapter 2

Background and related work

2.1 Background

2.1.1 Game theory basics

In this section we aim to provide a brief overview of the game-theoretic terms used in the paper. Game theory is by definition the mathematical modelling of rational agents, where the environment (in two-player games) is typically modelled by a payoff matrix, in which agents optimize their strategies to gain the highest utility, or in the case of this paper, the lowest possible cost. In a payoff matrix, one axis represents the available strategies of player 1 while the other represents that of player 2's. Every cell contains the payoff they both receive for their chosen actions. In other cases, an agent's utility or cost is expressed by an objective function, which takes into account the actions of all other players and the net utility of the action itself. For the objective function of this game, see Section 3.2. Incurring higher costs means having a lower overall utility, while achieving higher utility means decreasing costs. A Nash equilibrium occurs when none of the players can unilaterally deviate from their chosen strategy without incurring higher cost/lower utility. There are many types of equilibria,the one discussed in this paper is a pure Nash-equilibrium (PNE), meaning every player chooses her strategy with a probability of 1. This is one of the strongest concepts amongst all NE's, but is the hardest to compute.

Definition 1 (Pure Nash Equilibrium)**.** A strategy profile is a Pure Nash Equilibrium (PNE) [15] if there is no player with a unilateral deviation which would increase his payoff.

$$C_i(s) \leq C_i(s_i^{'}, s_-i) \tag{2.1}$$

Where i is the index of the player, $s_i^{'}$ is a unilateral deviation, and s_-i is the strategy of every other player.

Definition 2 (Social welfare)**.** The aggregate of every player's utility combined.

Definition 3 (Social optimum)**.** The player's strategy layout where the social welfare is the highest/social cost the lowest. Note that this is not necessarily an equilibrium.

Definition 4 (Price of Anarchy)**.** A popular measure of the inefficiency of equilibria [9], Price of Anarchy (PoA) measures the worst outcome over the social optimum. Its value depends on the player's utility function and the type of equilibrium used. A PoA close to 1 means that uncoordinated selfish players achieve a near-optimal system state without central control. S is the strategy space of the game.

$$\frac{\text{Highest cost equilibrium(S)}}{\text{Social optimum(S)}}$$

Definition 5 (Price of Stability)**.** Similar to the Price of Anarchy, Price of Stability (PoS) [2] provides a distinction between games where all equilibria are inefficient, and those where only some of them are. In a game with multiple equilibria its value is at least as close to 1 as the PoA or even closer. It measures the ratio between the best equilibrium and the social optimum. S is the strategy space of the game.

$$\frac{\text{Lowest cost equilibrium(S)}}{\text{Social optimum(S)}}$$

Definition 6 (Potential game)**.** A potential game [17] is one for which there exists a potential function Φ with the property that, for every unilateral deviation by some player, the change in the potential function value equals the change in the deviator's cost. Formally,

$$\Phi(s_i^{'}, s_-i) - \Phi(s) = C_i(s_i^{'}, s_-i) - C_i(s) \tag{2.2}$$

for every outcome s, player i, and unilateral deviation $s_i^{'}$

Definition 7 (ε-PNE)**.** A relaxation of the original PNE, with $\varepsilon \in [0, 1]$, this equilibrium is easier to achieve [15].

$$C_i(s) \cdot (1 - \varepsilon) \leq C_i(s_i^{'}, s_-i) \tag{2.3}$$

Definition 8 (ε-Best-response-dynamics (Maximum Gain))**.** An algorithm where in each iteration we select the player with a strategy that can obtain the largest cost decrease while other players strategy remains unchanged. Update his strategy and repeat this process until we reach ε-PNE. [15]

$$C_i(s) - \min_{s_i^{'} \in S_i} C_i(s_i^{'}, s_-i) \tag{2.4}$$

where $s_i^{'}$ is a best response to s_-i, and player i's strategy is updated to $(s_i^{'}, s_-i)$

Theorem 1 (Nash's existence theorem)**.** Every game with a finite number of players in which each player can choose from finitely many pure strategies has at least one Nash equilibrium, which might be a pure strategy for each player or might be a probability distribution over strategies for each player.

It is important to note that these situations may not always result in the highest social welfare (see prisoners dilemma for example). This inefficiency is captured by many factors, two of the most prominent are the Price of Stability(PoS), and the Price of Anarchy(PoA).

Theorem 2. Every potential game has at least one PNE.

Theorem 3. In a potential game, from an arbitrary initial outcome, best-response dynamics converges to a PNE.

Proof. In every iteration of best-response dynamics, the deviator's cost strictly decreases. As seen in its definition, the potential function strictly decreases,as players are optimizing their strategy, i.e, decreasing their costs. Because of this, no cycles are possible, and since the game is finite by assumption, best-response dynamics eventually converges at a PNE. □

The fact that applying BRD to a potential game will always yield a NE motivated heavily the development of the initial algorithm, however, with the addition of an explicit altruism modifier the objective function would no longer constitute a potential function. The players' actions would affect the potential function not only in their bandwidth related costs, but in the privacy property of other players as well, which would change according to the the new strategy set by the player, causing implicit changes in the function.

2.1.2 Fuzzy Message Detection basics

Fuzzy Message Detection(FMD) [4] is a cryptographic protocol similar to public key encryption, with the modification that every user shares their so-called "detection key". Each user's detection key is made by an algorithm that takes in their private key and their fp-rate, which could be then, for example, be forwarded to the server. The server then takes this key and a chosen ciphertext, and inputs a match if the holder of the key is its presumed intended destination. Of course, the server will never know with absolute certainty, and may send the message out to several other potential participants, hence the name: false-positive. FMD has three main properties:

- Every message must reach their intended target.
- Targets should also receive additional messages proportional to their desired false-positive rate.
- Only the recipient should be able to distinguish between honest messages and junk messages.

Another important aspect is that each message always decodes to a random sample from an array of plaintexts. This means that by random chance any message could be translated to something human readable and comprehensible, when deciphered by a random user's detection key, but this does not mean that the user is its intended target. This additional layer of ambiguity is called ambiguous encryption (AMB-PKE) and is formalized in the original paper [4].

Over the course of this paper, this system is implemented in a way that users can only select false positive rates in the range 2^{-n} where n ranges from 1 to 10. FMD is easily implemented in these cases, as explained in the original paper, also it was convenient to search in these ranges as the follow-up paper [16] has already dealt with many aspects of this system around these values.

2.2 Related work

Many of the game theory related concepts were taken from other articles, papers, and books. Price of Anarchy is a term coined in [9], while the Price of Stability is mostly known from [2]. While these are mainly used to describe network games, they can be handily used in this paper as well, to express differences between the results of different setups; e.g., how they change with increasing altruism. Regarding the definition of potential games and best-response dynamics, further information can be found in [15] and [17].

In the earlier iterations of the algorithm, an altruism model based on the whole social welfare was implemented based on [3], where the concept of of α-selfishness describes an equilibrium in which appending the social welfare to the payout of every player potentially motivates them to reach the social optimum. The fraction of the social welfare needed to be added describes the selfishness-level of the game. A game which cannot reach an equilibrium in the social optimum is ∞-selfish. This could prove to be useful in future works in this area, and can be added as another general descriptor of the achieved equilibrium.

Altruism is a ubiquitous concept, and can be found in economics [18] and evolutionary game theory [6], its goal in our model is to stimulate cooperation. If this model is to be implemented anywhere, its target audience is most likely to be privacy-conscious individuals, who may very well realize the beneficial effects of providing cover traffic. Altruism need not be the only motivation factor, with several systems providing other incentives [7, 8]

The original idea of examining the FMD system through the lens of game theory stems from a recent paper [16]. In it, several statistical tests are performed with real world data on the system, and it is shown that FMD performs weakly in nearly all privacy aspects. Their model was designed on the premise that every user is similarly privacy-conscious, there they assigned random fp-variables to every node. This method severely weakens the system, as its performance relies heavily on the effort made by its biggest users to shield the messaging process, as shown in Section 4.2. While data in itself is solid evidence, a preliminary analysis explained this weakness from a much broader perspective, utilizing game theory. Several questions were left unanswered, and so this paper picks up where the other one left off.

Regarding the balancing of bandwidth cost with the goal of avoiding privacy breach is explored in [19], where both the user and the attacker must weigh-in the costs of allocating bandwidth for their operations, either to choose a lesser-known but safer tor circuit, or to allocate it amongst different malicious nodes in order to attract traffic through them. The gPath algorithm successfully reduces the probability of an attack while inducing little extra bandwidth cost, by introducing additional uncertainty to the node selection process.

As seen later in the article, the system stabilises when the most active users set their fp-rate exceptionally high, while everyone else disables their own. This free-riding problem is introduced in [11]. [1] explains how the small amount of active uploaders in the Gnutella P2P system causes service degradation, with the influx of new users causing them to eventually be outside the search horizon, with no way of reaching the uploaded files. The paper argues that the eventual decline of uploaders ratio makes the system not only less efficient, but less private as well, with a complete shutdown possible by copyright authorities. Similar to Gnutella, Napster also experienced free-rider problems, and [7] takes a look at possible payment solutions without relying on altruism as we do. They employ best-response dynamics with Q-learning to simulate user behaviour.

Chapter 3

Modeling

3.1 Central distribution

The very first step towards determining the desired outcome of the model was constructing a mechanism that would centrally set the false positive rate of every user, then compare them and select the best one with regards to social welfare. This can be interpreted as a mechanism for faster convergence of the best-response dynamic when examining social optima. Based on this, discrete values have been established: 2^6 for the EU mail server and 2^7 for the college one. Due to the low refinement of the process function, actually setting every fp exponent to -6 or -7 is not the most efficient strategy, as shown in Chapter 4, where some users slide into the next exponent category. The social cost in the case of centrally set false positive rates are easy to compute, but not every system converges to them easily. Do note that this is not a game, as players have no moves or strategy here. Figure 3.1 show the college messages dataset values, in the case of no altruism applied. Compare the recommended fp exponent rate with the direct result of running a social optimum run in section 4, Figure 4.1.

3.2 The game

The game presented and modeled in this paper is a tuple (ν, ϵ, μ) where the set of players is $\nu = 1, ..., U$ and their actions are $\epsilon = p(1), ..., p(U)$ where $p(u) \in 0, [2^{-x}]$, where x is the set of whole negative integers from -1 to -10. Their utility functions are $\mu = \varphi(p(1), ..., p(U))$ such that for every $1 \leq u \leq U$:

$$\varphi_u = -L \cdot (1 - (1 - \alpha_u)^{in(u)}) - f \cdot (in(u) + p(u) \cdot (M - in(u))). \tag{3.1}$$

L is the utility loss upon a violation of a privacy requirement, such as recipient unlinkability, and is multiplied by the probability of user linkage, where $\alpha_u = \prod(1 - p(v)), v \in \nu \setminus u$. in(u) is total incoming honest messages for player u. f is the bandwidth cost for a single message, and M is the total amount of messages sent by every user. For ease of comprehension the first expression of the equation will be called C_i^{priv},for the expected value of a privacy breach externality,and the second C_i^{BW}, for the users own cost of downloading its messages. Upon closer inspection it can be seen that the privacy part is independent of the users own actions, while the message bandwidth cost is fully dependent on it. Section 3.4 shows that this leads to every player setting their strategy p(u)=0. Altruism is indeed

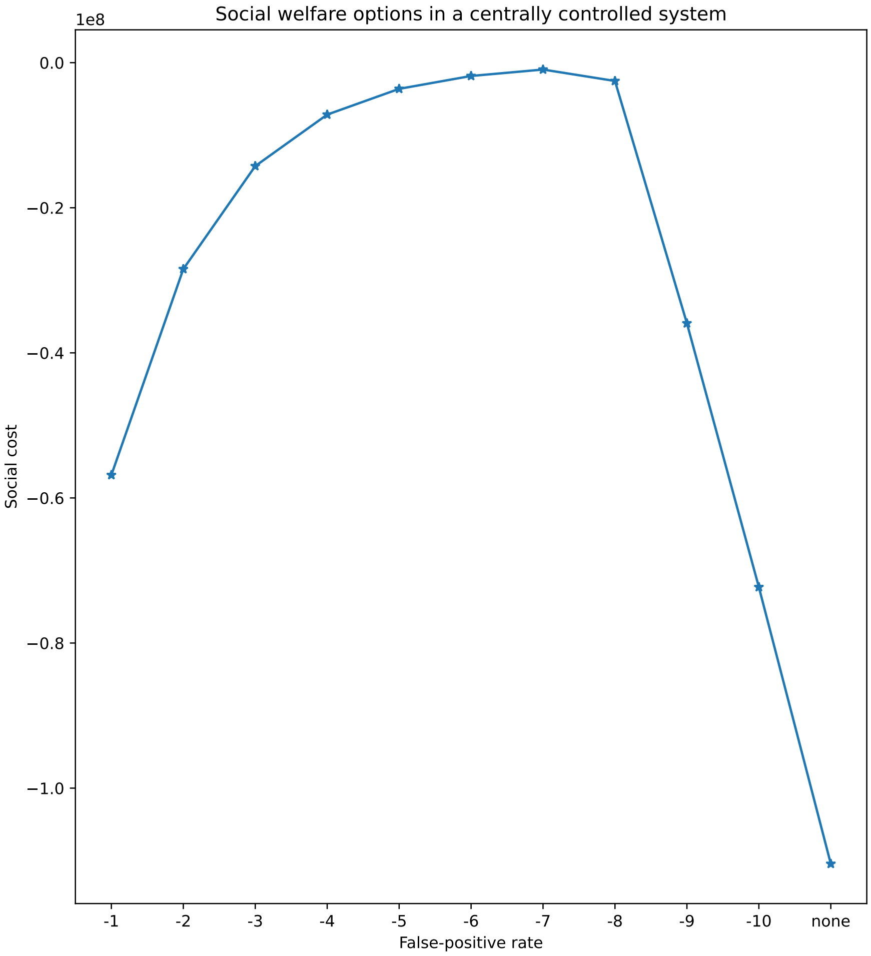

Figure 3.1: [College] Social welfare in a central distribution with no altruism coefficient. Note how applying zero protection is worse than applying too much.

necessary to implement for a higher social welfare, section 3.5 appends a third expression to the objective function, in order to effectively model it.

3.3 The graphs

A system of parties where every user can send a message to everyone else can be sufficiently modelled by a directed graph where we allow parallel edges to be made. For this purpose a python module called networkx[1] is used.

The 2 main datasets [14][13] used in this paper are identical to the ones used in [16], which were chosen on the premise that they could benefit from implementing FMD. One

[1] `https://networkx.org`

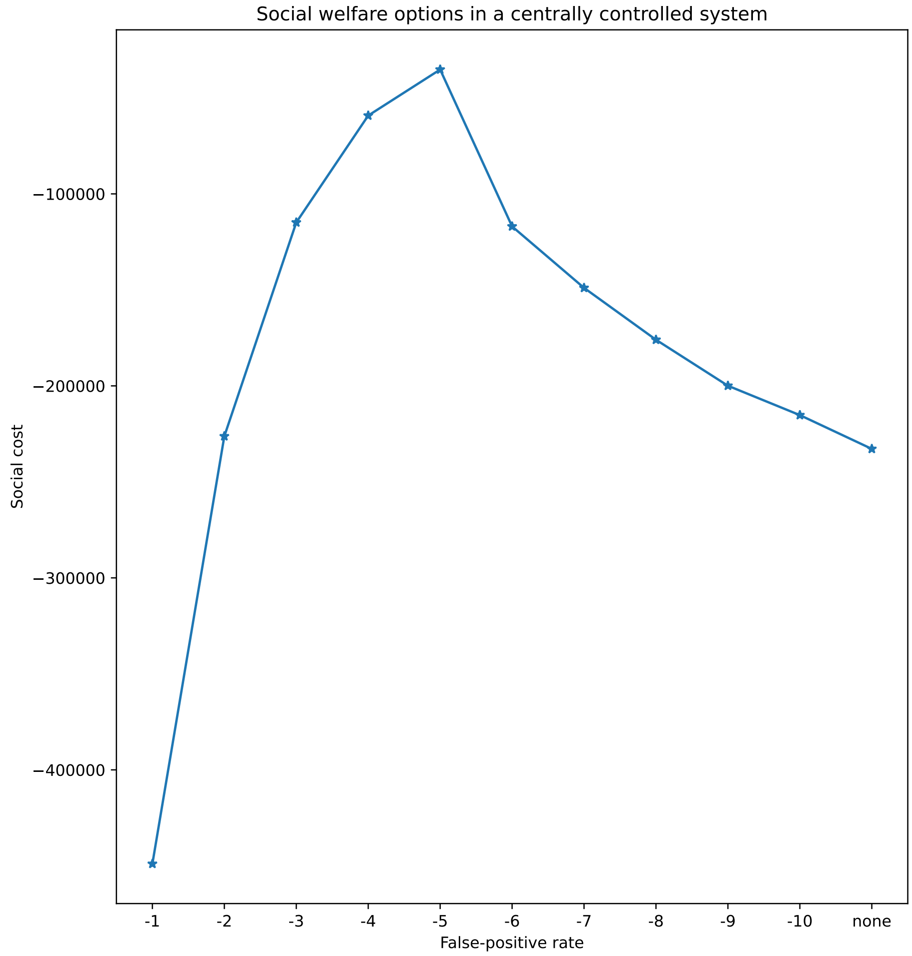

Figure 3.2: [Binomial] Social welfare in a central distribution with no altruism coefficient. Note how applying zero protection is better than applying too much.

is an instant messaging network from the University of California, while the other contains emails from various European research institutions.

During testing, two artificial graphs were made, one of them is made by connecting the nodes based on scale-free graph properties, and the other by binomial properties. In order to achieve the necessary parallel edge density, we ran the generation many times over-and over, this caused the graphs to lose some of their original properties, but proving to be interesting benchmarks nonetheless. Their main purpose was to provide smaller examples where various ideas had been tested and explored. They also proved to be interesting benchmarks for showcasing the problem of shared resources.

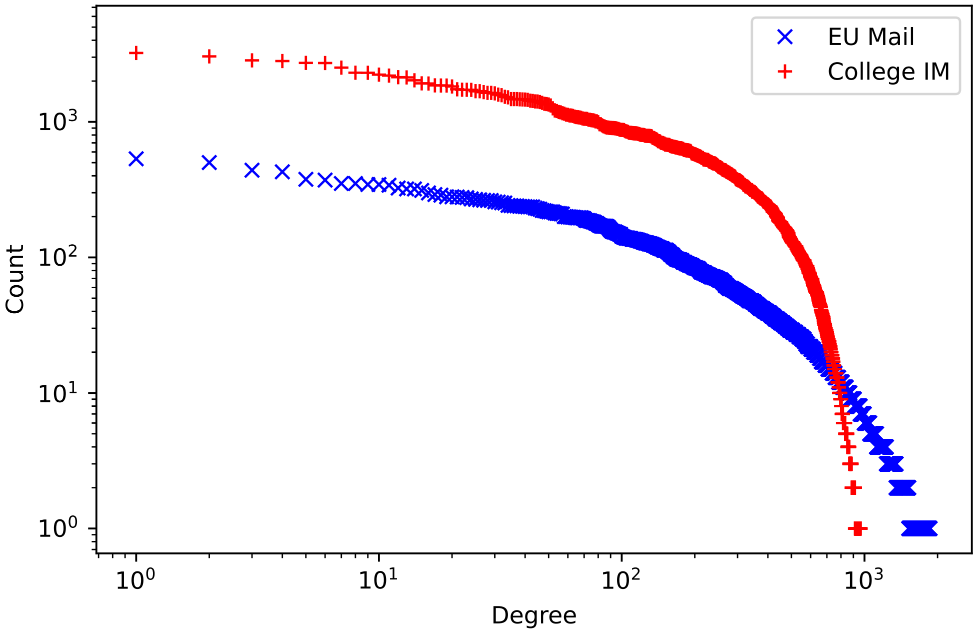

Figure 3.3: The distribution of messages. Altruism types are decided based on incoming messages.

3.4 Selfish Nash equilibrium

Theorem 4. Selecting the strategy of p(u)=0 is the only NE in the case of no altruism for every $u \in \nu$.

Proof. To prove it by contradiction, lets assume that there is a strategy $s_i \in S$ where $p(i) \neq 0$, and the game has reached an equilibrium, meaning no rational player can increase its own utility by unilaterally deviating, and changing its strategy. This is no true NE since, any player with non-0 false-positive rate has a strictly dominant strategy of setting it 0, as the privacy related cost does not matter from the players own standpoint. This causes the setup of only p(i)=0, for every $i \in \nu$, to be the only Nash equilibrium. □

As seen in figure 3.6, the lack of altruism causes an eventual decline to the absolute worst social welfare status. Observe figure 3.7 as through selfish cost decreases each player ultimately suffers almost 4x times the original cost in the end.

3.5 Altruism

Definition 9 (Altruistic player)**.** A player is said to be altruistic if its payoff/utility is directly affected by the welfare of others.

A key element to successfully run the system is the concept of altruism. Many iterations were made in implementing it, starting with no actual altruism present, rather just players participating in the game and players who don't. Non-participants automatically set their fp-value to zero, acting as "dead weight" to the system, while still suffering the privacy breach cost of it. In the context of the model, this means that every agent/player who

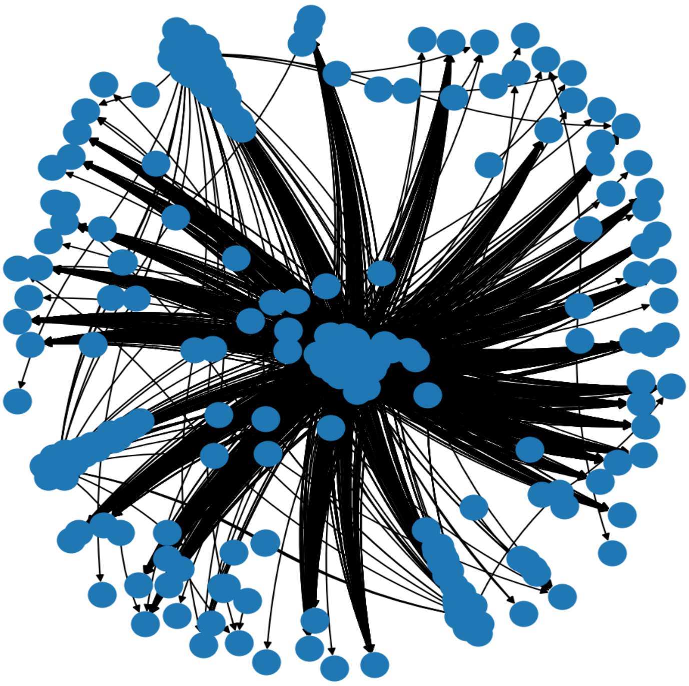

Figure 3.4: Layout of the graph created by binomial edge connection methods. Each node represents a user.

participates in the user selection is altruistic, since in this case, their own utility does not depend on the welfare of others, and setting their own fp-rate to 0 is the only rational choice. However, by abiding to the system rules, and only changing the exponent when asked to, and making no further change upon finalization, they partake in the group effort, thus they are altruistic. This implementation comes with a few shortcuts in designing the algorithm, and its model would prove to be too simplistic.

The final iteration saw the formal implementation of altruism/selfishness to the game: those who would became an altruistic actor in the system received a selfish coefficient more than 0. Other players could still be participating in the system, however they had a selfish rating of 0 meaning other users privacy loss would not affect their utility.This player type was not utilized in the results below. A selfish rate of 1 means others privacy loss is fully factored into the users cost. This plays an important role in finding the NE of the game, as we expect users with altruism factored in to show some convergence towards the social optimum. It is very important to note that rational players will not choose an action that would decrease their utility. In a sense, their willingness to cooperate is expressed in their altruism coefficient.

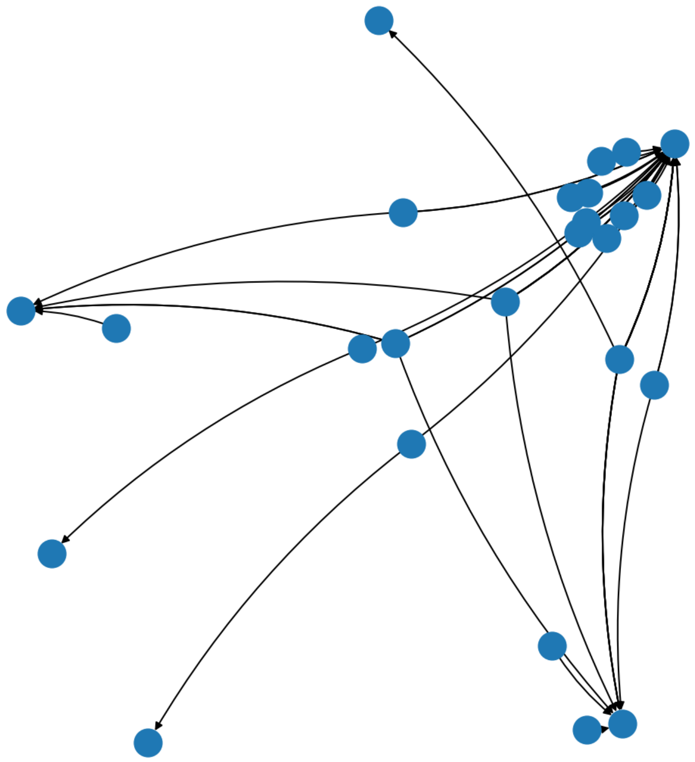

Figure 3.5: Layout of the graph created by the power law connection method. Each node represents a user.

In our altruism definition below, the payoff function is modified with an extra expression, the summed privacy cost of the other users, which is modified by the altruism constant, with $\alpha \in [0, 1]$. With the increase of α players are less likely to pick lower false-positive values, as it would hurt the anonymity of the system. Intuitively, players with large amount of incoming or outgoing messages will have a larger impact on the system, meaning the change of their strategy carries a bigger impact, and is more prone to be penalized by the altruism factor, if the fp-rate is lowered. Therefore it could be expected that the most active users in the system will pick fp-rates closer to the social optimum. However, as seen in section 4 high impact altruistic players must compensate the fp reduction of others, often finding themselves having to set their fp-rate to the maximum available, thereby incurring high bandwidth costs.

$$U_i = C_i^{BW} + C_i^{Priv} + \alpha \cdot \sum_{i \neq j} C_j^{Priv} \tag{3.2}$$

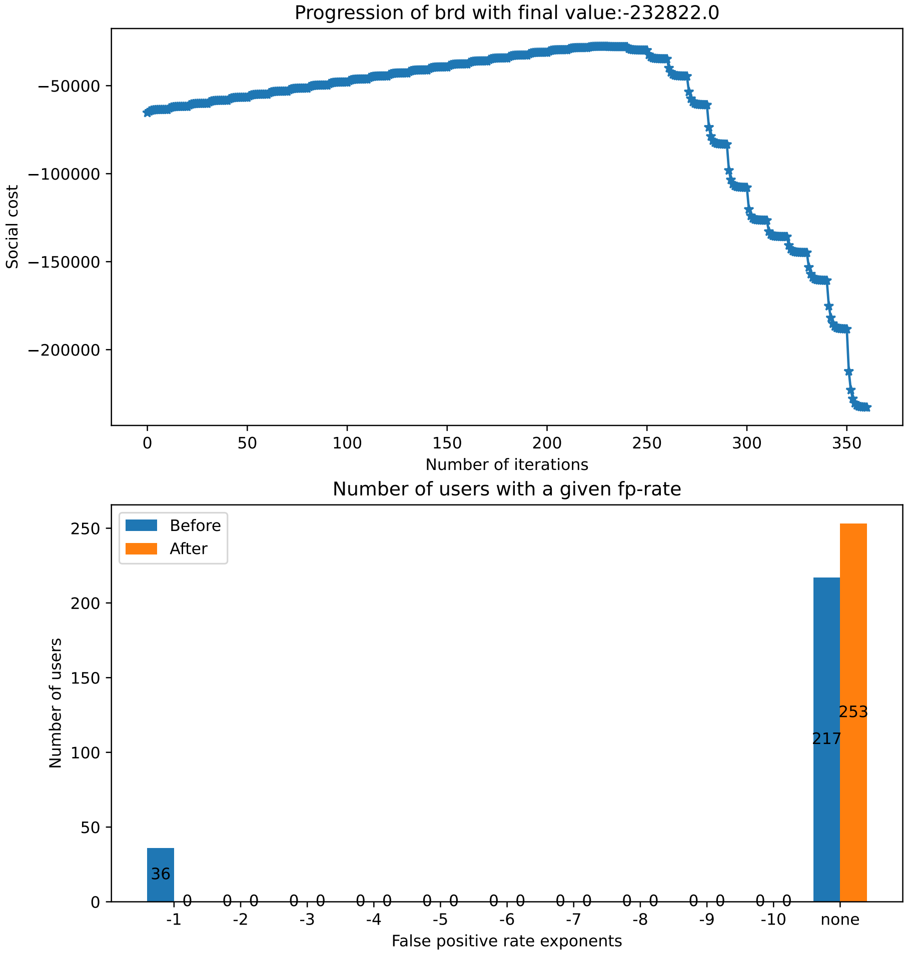

Figure 3.6: [Binom]A game with no altruism on the binomial artificial graph. Note the degradation of social welfare over the iterations, as the algorithm eventually sets every fp rate to 0.

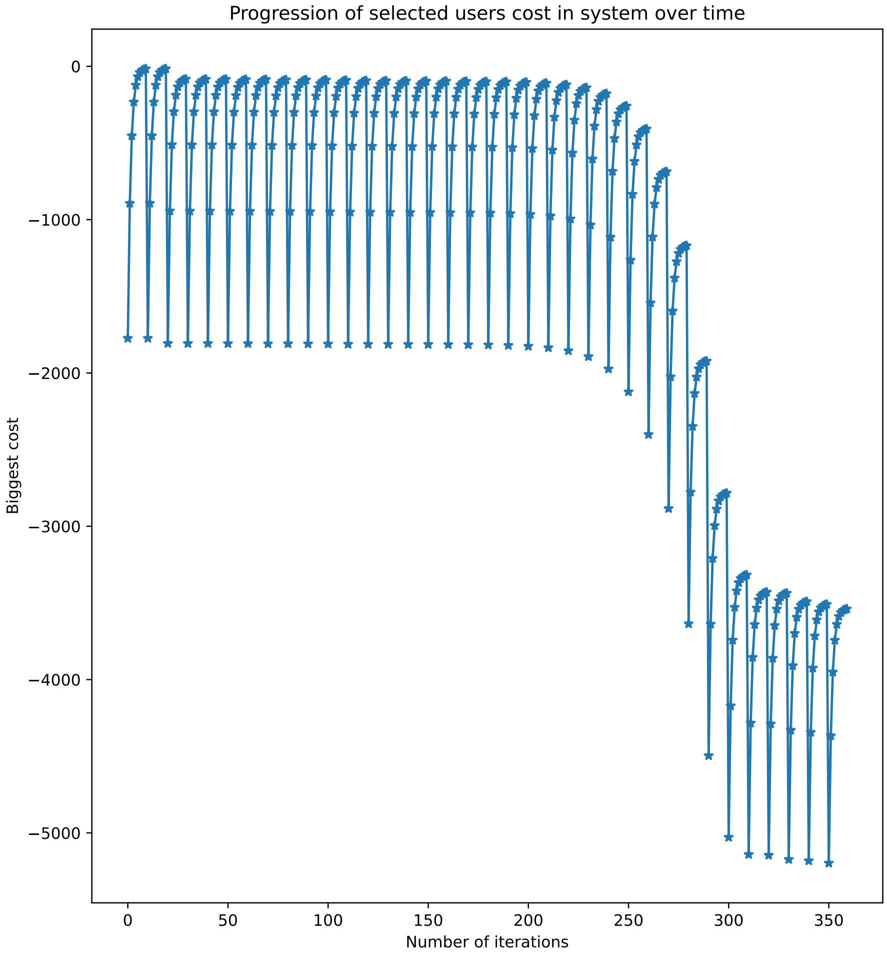

Figure 3.7: [Binom]A game with no altruism on the binomial artificial graph. Note the degradation of individual utility, as the algorithm eventually sets every fp rate to 0.

Chapter 4

Implementation and Results

4.1 Best-response dynamics

From a modeling perspective, its important to note that players take their turns after each other, only one player moves at a time, and each time the social cost is calculated again. Each move is setting its own fp-value one step higher or one step lower, where a step is the increase or decrease of its fp exponent by a value of 1, meaning the process has a learning rate of 1. This starting setup is just a method to make the system converge easier. Neither this setup nor the convergence is real in the sense that they need not take place in the dynamic of a real system. The only true result is the false positve rate distribution of the users. The original objective function in 3.2 presented games with truly one PNE, however, this was that of a selfish equilibrium. The addition of an altruistic factor most probably added local minimums to the potential function, thereby the system could halt at various equilibria, depending on the starting setup. With this the price of anarchy and the price of stability will not be the same, albeit still computable.

The distinction of selecting the player with the largest cost decrease versus any player with an acceptable cost decrease means a significant increase in computing power required, as the complete traversal of the user array is necessary in every iteration. During the calculation process, getting the the user linkage coefficient also means a nested for loop, in the current naive implementation, since each users linkage depends on everybody else's fp-rate. Additionally, altruism regarding the privacy loss of others may only be calculated after this update has taken place. During testing, the process of reaching the Nash-equilibrium proved to be extremely taxing CPU wise, therefore a multi-core solution was needed to speed the process up. This was achieved by using the multiprocessing package of the python language.

It should also be noted that after some considerations, during the calculation of the social optimum and other equilibrium's, inactive users do not participate, as to illustrate the difficulty of optimizing a system with many free-riders. Ideally in a completely optimal scenario, every single user participates, but this may very well be a completely unlikely case. It is expected that in these cases the fp strategy converges harder to the centrally recommended fp-value, as it is no longer necessary for the most active users to overcompensate in their fp-settings.

4.2 Results

In the test environment, both the original datasets, as well as 2 synthetic ones were tested for best/worst case scenarios and social optima. Users were split into groups based on incoming and outgoing messages, where the thresholds was around the single digits as illustrate the findings in [16], where the amount of messages sent or received greatly correlates with privacy loss susceptibility. Therefore users with little impact on the system will hardly ever raise their fp-rates above 0. To find the price of stability/anarchy, the algorithm would start from -1 and -10 starting setups respectively, with the notion that they would arrive at the same strategy distribution. Results showed that they would halt at different setups, although similar in social cost. One of the main differences was the recommended user compositions. The 2 different equilibria would then become the basis for the PoA/PoS, with the higher cost in the former, and the lower cost in the latter.

Figure 4.1 displays the social optimum of the College messaging system dataset. Compare the fp exponent values to figure 3.1, it can be seen that the social optimum could very easily converge around the preset value of central distribution mechanism.

Observe the results of 4.3 laying out the user distribution of graph (image:3.4) over the whole strategy space, grouped by their incoming messages. Those few with a significant number of messages set their fp-rate to the maximum to combat the degradation of social welfare seen in figure 4.2. While the distribution clearly displays the effect of free-riding it is still a significantly better equilibra than the one seen in figure 3.6.

Comparing figure 4.4 with 4.6 the effect altruism is clearly visible. In the case of this graph no user would find it rational to increase its fp-rate even with an altruism coefficient of 0.3. A rate of 0.8 is needed for every active participant to set its exponent to -1. During testing every participant either completely disabled, or had set its fp-rate to the maximum. This causes the setup to have a selfishness level of 0.8 for active users, since they are playing the social optimum at that level by maximising their fp-rate. In the case of the coefficient being 0.3 the PoA/PoS is ≈ 2.28 for example. This means they are far from playing the optimum.

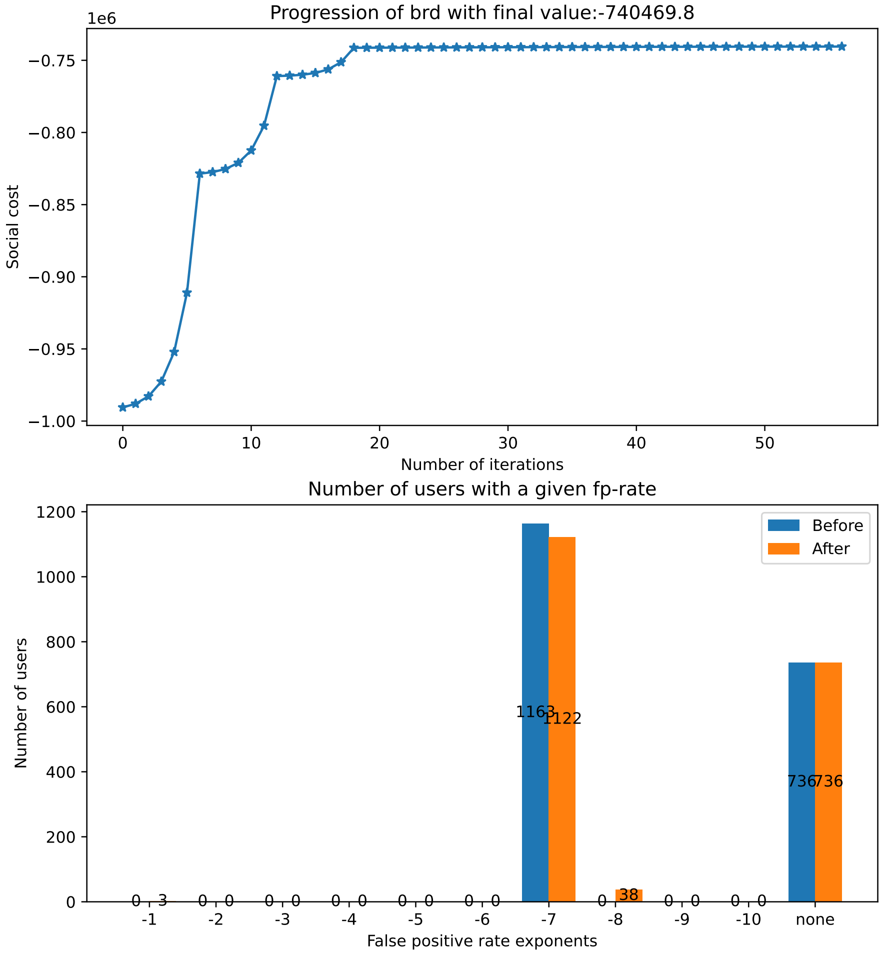

Figure 4.1: [College] Social optimum as a result of applying the brd algorithm with no altruism coefficient. Note how the distribution is similar to the centrally recommended fp exponents

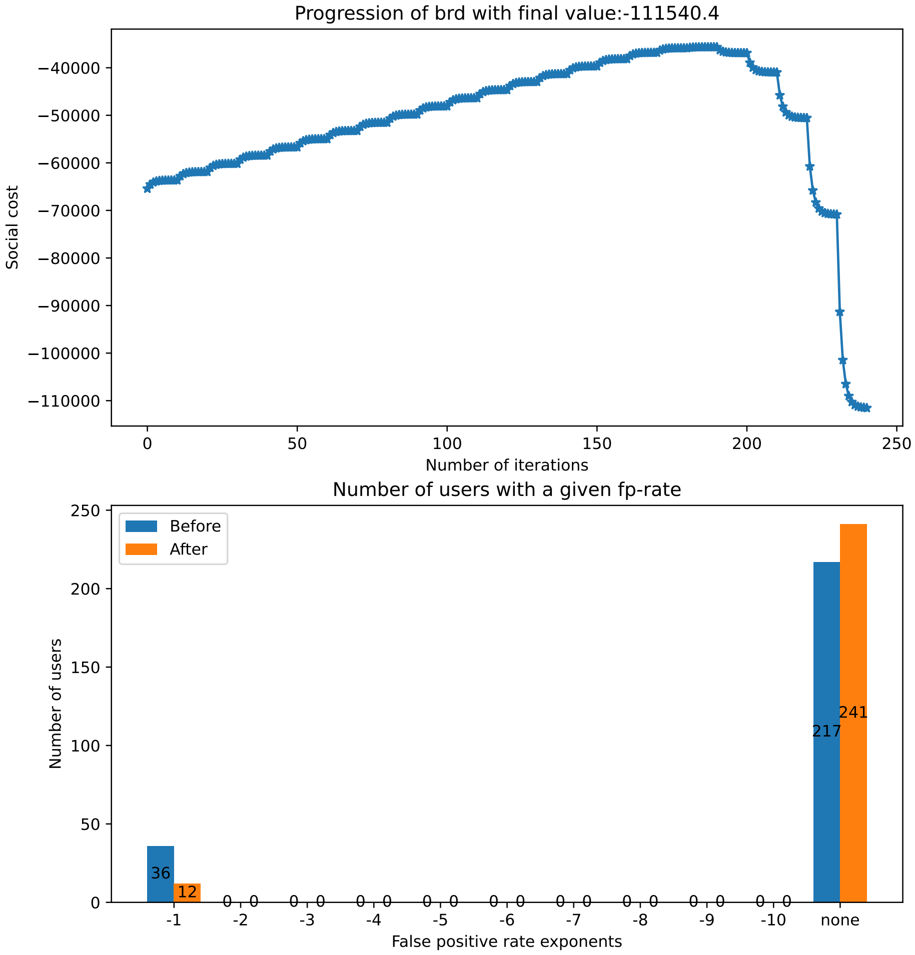

Figure 4.2: [Binom] Social welfare as a result of applying the brd algorithm with 0.6 altruism coefficient. Note the degradation due to selfishness

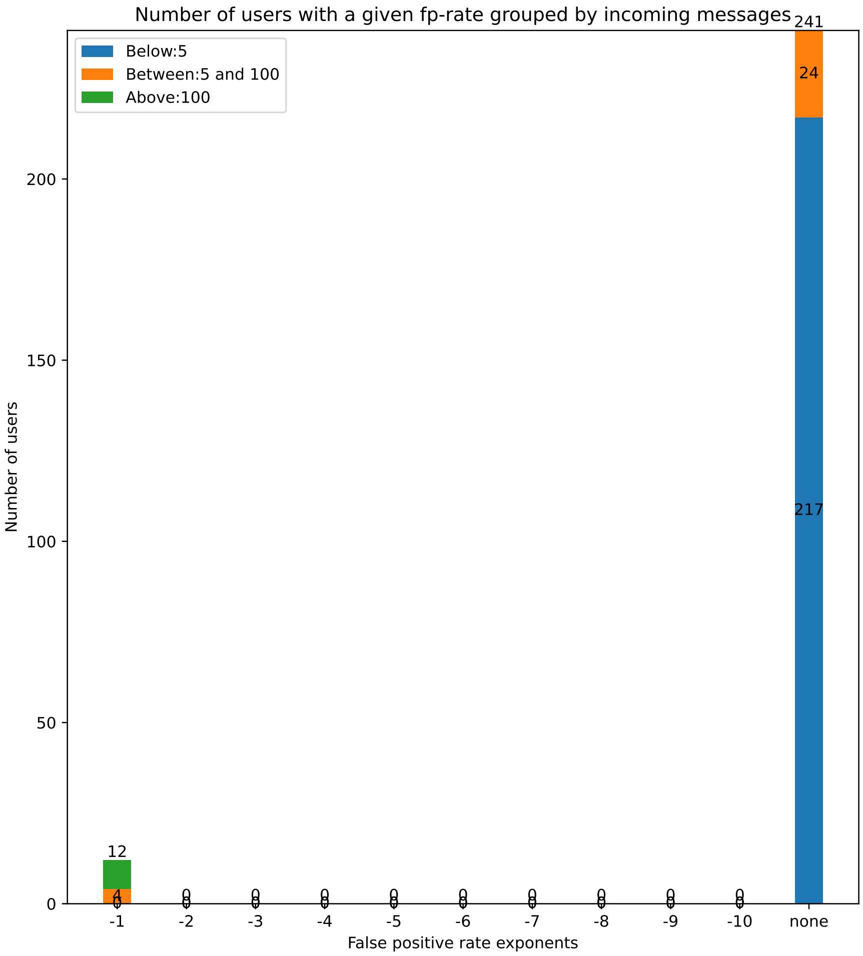

Figure 4.3: [Binom] Distribution of users based on incoming messages as a result of applying the BRD algorithm with 0.6 altruism coefficient. Users with a high amount are exclusively placed at -1, while the rest mostly turns their protection off.

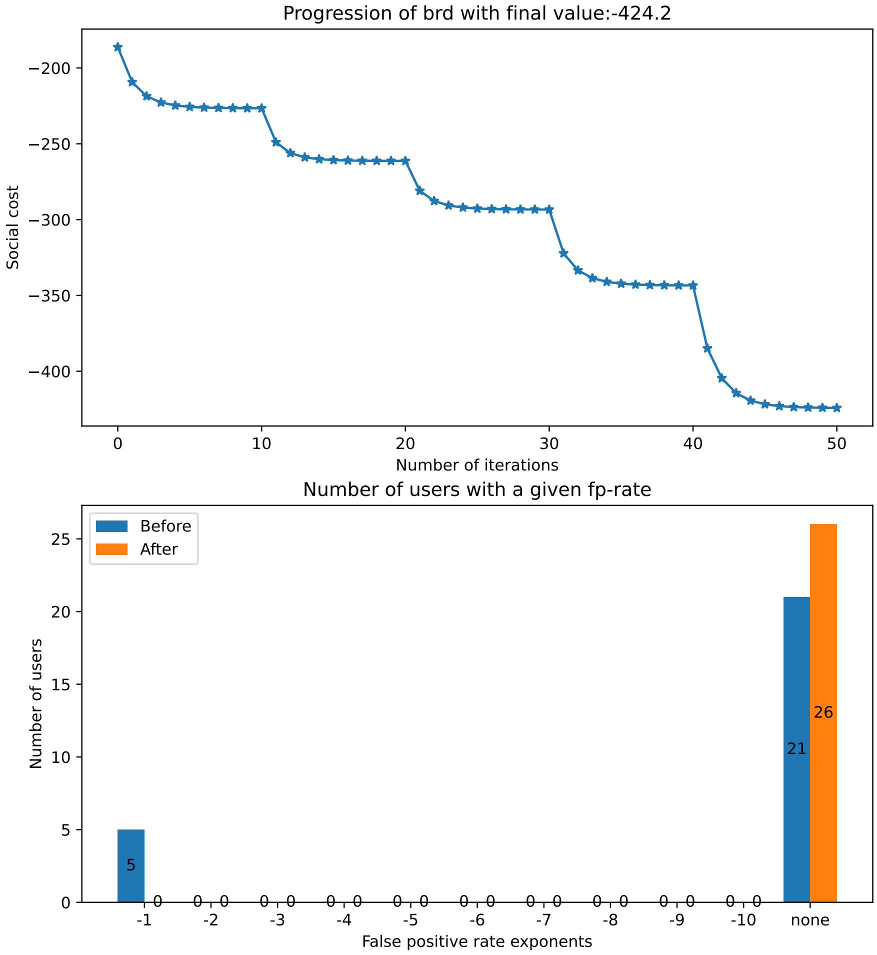

Figure 4.4: [Scale-free] Social welfare as a result of applying the BRD algorithm with 0.3 altruism coefficient. Every active player disables its protection

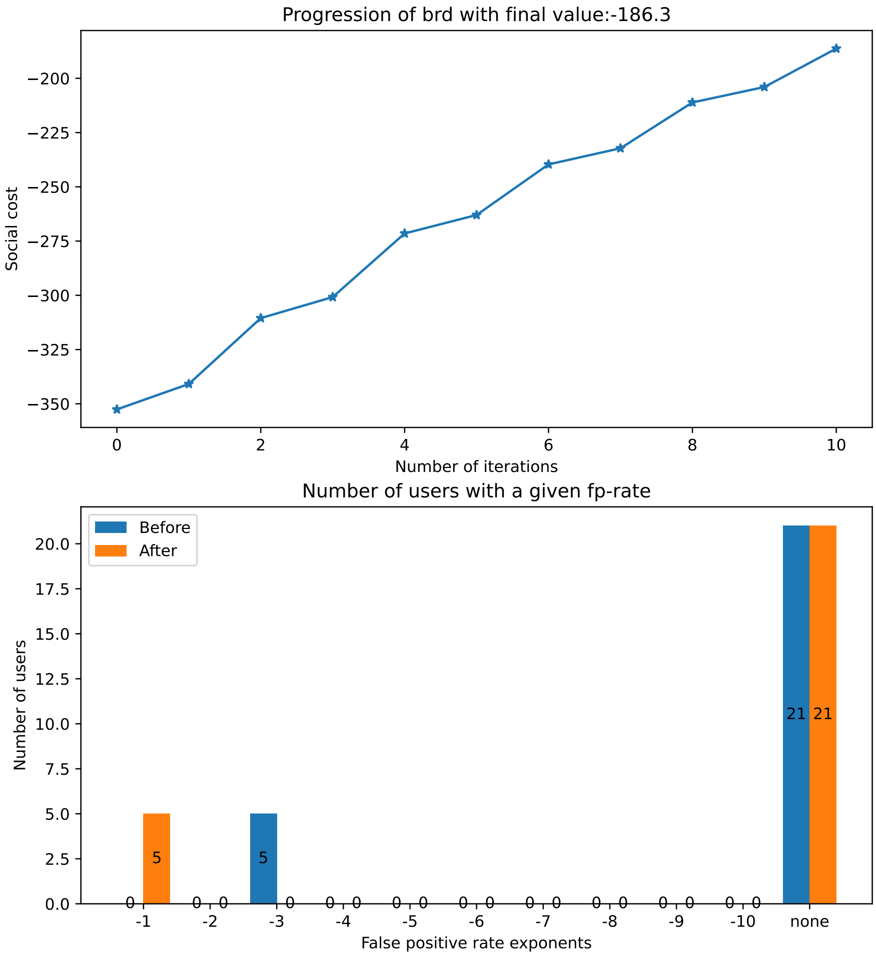

Figure 4.5: [Scale-free] Social optimum as a result of applying the brd algorithm with 0.3 altruism coefficient. Every active player set its exponent rate to -1.Compare the final social welfare value with figure 4.4

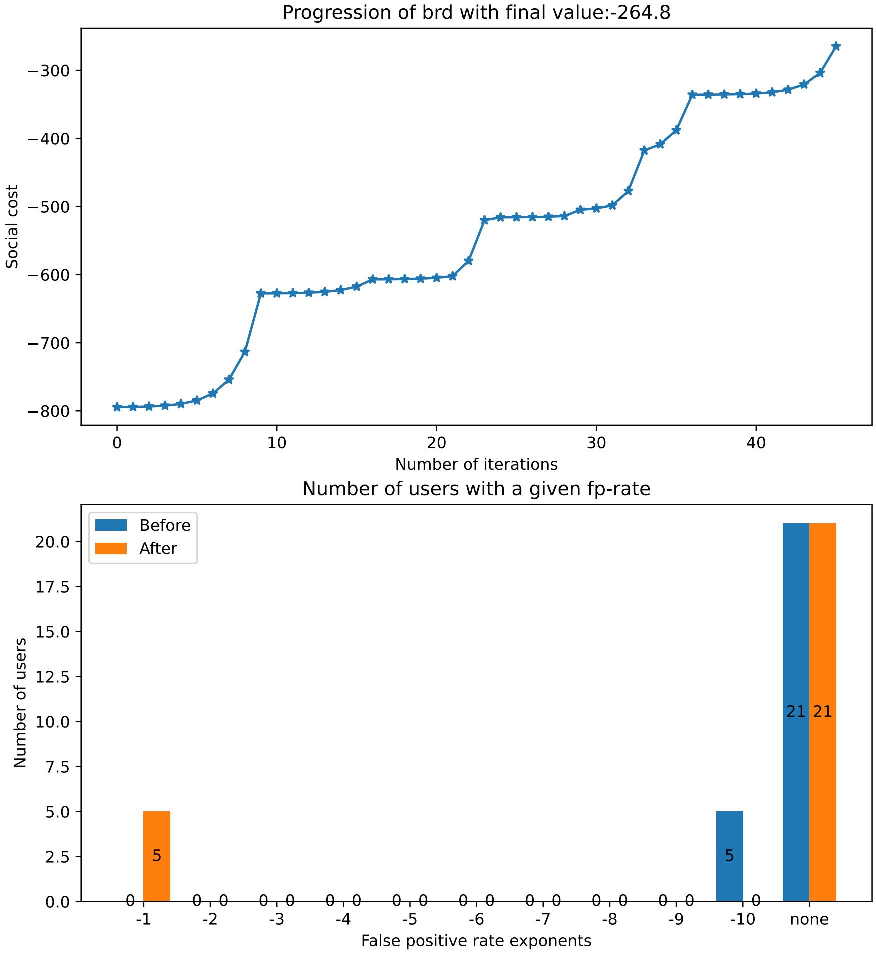

Figure 4.6: [Scale-free] Social welfare as a result of applying the brd algorithm with 0.8 altruism coefficient. Every active player sets its protection the maximum thereby greatly increasing the social welfare

Chapter 5

Conclusion

Having been laid out in [16], mixing in altruism with the FMD game does indeed produce strategies where the equilibrium is more desirable. This is important as the system clearly demonstrates its superiority in many privacy related aspects compared to existing methods/applications, while its ease of use may mean widespread popularity. The classic effects of a shared resource system had been observed, with some graphs/setups clearly displaying the inefficiency of a system with relatively few participants or small incentives. This is something that needs to be addressed in order to avoid the fate of early 21st century p2p networks. With the participants of the system clearly relying on others to protect their privacy, interdependent privacy comes in[5], since the act of modifying your own fp-rating is externality for everyone else, without you suffering any cost or receiving any utility(in case of no altruism). The intended audience for such a system however might be able circumvent its weaknesses more effectively as they could easily be aware of such concepts, and communicate/plan their strategies accordingly.

5.1 Limitations

The main limitation of the model is the fact that it draws from past data, while in reality it should provide real-time recommendations to users regarding false-positive rate. The implementation of 2^{-n} like false-positive rates, while plausible by the original article, could still prove to be too restrictive, with a fine grained fractional model providing more possibilities. Other design shortcuts include the construction of the objective function, which does not take into account every privacy property, nor does it accommodate different user types. Game theory in general assumes that every player is rational, which is seldom the case when it comes to the bounded rationality of human actors. A payment or a reward system is also missing from the model, and while altruism as a concept is enough to prove that the system is stable, in real life it is rarely the only supporting force behind a successful group-effort environment. For any future work based on more complex cases, the running time of the algorithm must significantly decrease. Its upgrade is of key importance, as its current implementation makes many traversals over long lists, some which could be shortened, or be left out with clever algorithm design. While this game does not constitute a traditional congestion or other kind of network game, it is likely that the bounds on computational efficiency established for them carry over to our game as well. This means that any kind of best-response dynamics we implement, its lower bound for the number of iterations required will always be exponential in regards to the number of players/users. Therefore, after the optimization of the algorithm, modern hardware and software solutions are needed to keep the computational time at an acceptable level.

5.2 Generalization

One of the main future avenues would the construction of a temporal detection ambiguity game, either as a standalone, or incorporated into the existing game. The strategy space would be similar, with each player having the ability to adjust its fp-rate. The system could also be viewed as a cooperative game, with a Shapley-value, and a Shapely-Shubik index. Other privacy enhancing systems could perhaps be viewed with enough generalization of the modeling system. This would produce interesting equilibriua comparisons, as well as different PoA/PoS setups. A user group simulation is also an interesting concept that would show how its potential user base would evolve over time, and perhaps yield insight into its potential longevity.

Bibliography

[1] Eytan Adar and Bernardo A Huberman. Free riding on gnutella. *First monday*, 2000.

[2] Elliot Anshelevich, Anirban Dasgupta, Jon Kleinberg, Éva Tardos, Tom Wexler, and Tim Roughgarden. The price of stability for network design with fair cost allocation. *SIAM Journal on Computing*, 38(4):1602–1623, 2008.

[3] Krzysztof Apt and Guido Schaefer. Selfishness level of strategic games. *Annals of Pure and Applied Logic - APAL*, 49, 05 2011.

[4] Gabrielle Beck, Julia Len, Ian Miers, and Matthew Green. Fuzzy message detection. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, pages 1507–1528, 2021.

[5] Gergely Biczók and Pern Hui Chia. Interdependent privacy: Let me share your data. In *International conference on financial cryptography and data security*, pages 338–353. Springer, 2013.

[6] Jeffrey A Fletcher and Martin Zwick. The evolution of altruism: Game theory in multilevel selection and inclusive fitness. *Journal of theoretical biology*, 245(1):26–36, 2007.

[7] Philippe Golle, Kevin Leyton-Brown, and Ilya Mironov. Incentives for sharing in peer-to-peer networks. In *Proceedings of the 3rd ACM conference on Electronic Commerce*, pages 264–267, 2001.

[8] Murat Karakaya, Ibrahim Korpeoglu, and Ozgur Ulusoy. Free riding in peer-to-peer networks. *IEEE Internet Computing*, 13(2):92–98, 2009. DOI: `10.1109/MIC.2009.33`.

[9] Elias Koutsoupias and Christos H. Papadimitriou. Worst-case equilibria. *Comput. Sci. Rev.*, 3:65–69, 1999.

[10] Giuseppe De Marco and Jacqueline Morgan. Slightly Altruistic Equilibria in Normal Form Games. CSEF Working Papers 185, Centre for Studies in Economics and Finance (CSEF), University of Naples, Italy, October 2007. URL `https://ideas.repec.org/p/sef/csefwp/185.html`.

[11] Gerald Marwell and Ruth E Ames. Experiments on the provision of public goods. i. resources, interest, group size, and the free-rider problem. *American Journal of sociology*, 84(6):1335–1360, 1979.

[12] John Nash. Non-cooperative games. *Annals of mathematics*, pages 286–295, 1951.

[13] Pietro Panzarasa, Tore Opsahl, and Kathleen Carley. Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. *JASIST*, 60:911–932, 05 2009. DOI: `10.1002/asi.21015`.

[14] Ashwin Paranjape, Austin R. Benson, and Jure Leskovec. Motifs in temporal networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, WSDM '17, page 601–610, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450346757. DOI: `10.1145/3018661.3018731`. URL `https://doi.org/10.1145/3018661.3018731`.

[15] Tim Roughgarden. *Twenty Lectures on Algorithmic Game Theory.* Cambridge University Press, 2016. DOI: `10.1017/CBO9781316779309`.

[16] István András Seres, Balázs Pejó, and Péter Burcsi. The effect of false positives: Why fuzzy message detection leads to fuzzy privacy guarantees? In *International Conference on Financial Cryptography and Data Security*, pages 123–148. Springer, 2022.

[17] Lloyd S. Shapley and Dov Monderer. Potential games. 1994.

[18] Herbert A Simon. Altruism and economics. *The American Economic Review*, 83(2): 156–161, 1993.

[19] Nan Zhang, Wei Yu, Xinwen Fu, and Sajal K Das. gpath: A game-theoretic path selection algorithm to protect tor's anonymity. In *International Conference on Decision and Game Theory for Security*, pages 58–71. Springer, 2010.