Is Distributed Trust More Trustworthy?

Kurt Nielsen[*]
University of Copenhagen

The Seventh Workshop on the Economics of Information Security, 2008

April 23, 2008

Abstract

We provide a comparative economic analysis of a traditional trusted mediator, e.g. an auction or a consultancy house, and a mediator based on distributed cryptography (threshold trust). The two institutions are compared in a supergame that compares the immediate gain from corruption with future losses if corruption is detected. Corruption with threshold trust requires cooperation among $T+1$ out of N preassigned independent third parties, which results in relative higher detection rates. If all incidents of corruption are detected, traditional trust is the most trustworthy institution. This, follows from the fundamental division problem that gain from corruption is divided among less than honest gain with threshold trust. On the other hand, if the threshold is $T = N-1$, threshold trust is the most trustworthy institution for any detection rate less than 1. In all intermediate situations, determining the most trustworthy institution depends on the institutional setup and payoffs. However, the required cooperation with threshold trust allows a public authority to enhance trust in various ways. Furthermore, conflicting interests may cause the NT-TTP to breakdown after detected corruption, and thereby make the punishment more harsh and the institution more trustworthy.

Keywords

Secure multiparty computation, threshold trust, corruption, supergames.

[*] I would like to thank Ivan B. Damgaard, Jakob I. Pagter, Jens-Martin Bramsen and Philipp Festerling for their helpful comments. This work is sponsored by the Danish Strategic Research Council as part of the research project SIMAP: www.sikkerhed.alexandra.dk/uk/projects/simap.

1 Introduction

We consider the institutional design of a Trusted Third Party (TTP) that is paid to confidentially coordinate private information according to a comprehensive protocol. One example could be a sealed bid auction, where the submitted bids are kept confidential and coordinated as prescribed in the auction rules.

In information economics the existence of a TTP is a basic assumption and a requirement for many economic mechanisms. One of the most well known theoretical results in this field is *The Revelation Principle*, which says that for any mechanism there exists a weakly dominating mechanism where all participants reveals their type to a TTP that tells the participants what to do, see e.g. Gibbard (1973) or Myerson (1979). Economic theory has little to say about the ideal design of the required TTP and where it comes from. However, economic literature provides some insight to the nature of trust, see e.g. James Jr. (2002).

In practice the market provides many different TTPs that in one way or the other handle private information. Common for these is that the TTP is a single entity; a person or an institution. Almost by definition, the information revealed to the TTP is crucial and valuable to others or the TTP itself (e.g. when the commission depends on the turnover). To counteract corruption, the traditional TTP enhances its reputation, for example by enforcing strict procedures. Nevertheless, corruption happens, either independently by an individual or an institution or more organized among individuals or institutions.

In Computer Science the topic of designing TTP institutions has been a central challenge for many years. While traditional cryptography focuses on preserving privacy within a group of individuals with full access to the information, recent contributions is a fundamental break with the idea of placing all trust in a single entity at any time. The discipline of distributed cryptography provides a theoretical solution by the so-called *secure multiparty computation*, which allows a number of parties to jointly perform a computation on private inputs without releasing other information than agreed upon a priori. The seminal ideas go back to Shamir (1979) and the theory was founded in the 1980s, see e.g. Goldreich et al. (1987); Ben-Or et al. (1988); Chaum et al. (1988), but only recently has the ideas been refined and made applicable in practice, see e.g. Bogetoft et al. (2005), Bogetoft et al. (2008) and Malkhi et al. (2004)[1].

To provide an idea of how secure multiparty computation works, consider the following simplified problem of adding the two privately held numbers a and b. Let c be a secret key and the encrypted information submitted be c^a and c^b. Now multiplying the two numbers yield c^{a+b} and if you know c you know the result, though, unfortunately you also known the private inputs. To solve this problem, let c be the following solution to a linear function $f(0) = c$ and the $f(x_1) = y_1$ and $f(x_2) = y_2$ two random numbers on the function. Now, one

[1] The three papers represent the following two research projects: SIMAP (www.sikkerhed.alexandra.dk/uk/projects/simap/index.htm) and FAIRPLAY (www.cs.huji.ac.il/project/Fairplay/home.html).

of the points provides no information while both points provide all information about c. Although other operations like that of comparing two numbers require many more operations, the basic intuition is the same[2], it is feasible without revealing any information other than what was agreed upon a priori.

This paper compares two distinct trust institutions based on respectively, a single TTP (traditional trust) and an organized network of N TTPs using secure multiparty computation (threshold trust). The two institutions are compared in a game theoretic model, where corruption is a tempting strategy. Unlike traditional trust, no single TTP as a member of threshold trust can make corruption by misusing the private information. As in the example above, revealing the private information requires coordination. In general threshold trust is designed such that corruption is avoided as long as no coalition of $T + 1$ colludes[3]. Although, the computational complexity depends on the choice of N and T, it is reasonable to consider N and T as fundamental design variables. This paper aims at providing economic reasoning for this fundamental design issue. I will refer to the two different trust institutions as T-TTP (traditional TTP) and NT-TTP (threshold trust).

In the game theoretic model, a TTP plays repeatedly against a market that demands a mediation job from the TTP. The TTP gets a high payoff from playing "corrupt" as oppose to playing "honest", but the market can punish the TTP by selecting another competing TTP. In an infinitely repeated game, a sufficient valuation of future punishment (weighted by a discount factor) makes it economically optimal to play honest as oppose to playing corrupt and getting a high payoff in the short run. The question is whether it requires a higher or a lower weight on future punishment to make corruption unattractive to NT-TTP as opposed to T-TTP. With NT-TTP, corruption can only happen if a predefined number of $T + 1$ TTPs co-operate. By intuition such a system may seem superior to a T-TTP. However, if $T + 1$ is less than N, payoff from playing honest is divided by more than payoff from playing corrupt. The analytical model illustrates these two counteracting effects.

The modeling has many similarities to models of cartels. Although the basic game typically has some of the same structures, the players in cartel games are competing firms and the demand side is represented by the underlying elastic demand function, see e.g. Motta (2004). Here we assume perfect competition where the market pays the competitive price or selects another TTP. In a cartel game, payoff from co-operation is the illegal collusion, and the tempting deviation is a unilateral deviation from the coordinated monopoly profit. In this model the co-operation strategy is the honest play, and the deviation the illegal corruption. Furthermore, the NT-TTP structure forces the deviation to be

[2] Comparing two numbers depends on the size of the numbers and requires a lot of communication between the involved TTPs. Comparing 2 32 bit integers in a NT-TTP where $N = 3$ and $T = 1$ takes approx. 1 second, see e.g. Bogetoft et al. (2006).

[3] Here the threshold T is the maximum number of TTPs that can not reveal the information, other parts of the literature operate with a threshold for the minimum number of TTPs that can reveal the information.

coordinated. How this coordination is taking place is not modeled, it is simply assumed that the most profitable number of $T+1$ TTPs form a collusive coalition[4]. From a welfare perspective, the deviation is a positive thing in a cartel game (since it may start a price war) and a negative thing in our model. On the other hand, with NT-TTP, the task of the authorities is to make collusion difficult in both cases. In cartel games, the participating firms are looking for deviations from cooperation and the punishment from deviation from the cartel's point of view, is to play the competitive equilibrium. In our setting the punishment for detected corruption comes directly from the exclusion from a large part of the market. In our model the information about corruption is modeled simply by an exogenous detection rate as opposed to the more advanced cartel models, where the market in one way or the other provides indications of deviation. E.g. Rotemberg and Saloner (1986) base the deviation on expected profit given demand fluctuations, while Porter (1983) introduces a certain trigger price that triggers the punishment period.

A related line of literature, models the so-called leniency programs, where members of the cartel get reduced penalty for helping the authorities in cartel cases, see e.g. Motta and Polo (2003). These models involve exogenous probability about things like the chance of being reviewed by the authorities and the chance of being proved guilty in case of no co-operation with authorities. This model is more simple and operates with a single detection rate and that detection triggers punishment forever. In a leniency program, it may, e.g., be optimal to remain in the cartel although the firm is under review.

Another line of research consider the game theoretical rational of sharing a secret using threshold trust. The primary setup is where the involved TTPs each have a higher value of the secret if no one else sees it. In a one-shot game (where all are suppose to submit their shares simultaneously) non of the parties have an incentive to distribute their share, see Halpern and Teague (2004). Several papers suggest mechanisms and setups that counter act this finding and makes it rational to share the secret, see e.g. Halpern and Teague (2004) and Abraham et al. (2006) a.o. Recently Maleka et al. (2008) extends these ideas by modeling it as a repeated game, where lack of co-operation (not sending the share) is punished in future repetition of the game. In this paper I differ from this line of research by considering a different setup, where the involved TTPs are paid to supply a service and hereby to participate with their individual shares[5].

The outline of the paper is as follows. Section 2 describes the characteristics of threshold trust. Section 3 provides the game theoretic modeling and an immediate comparison of the two trust institutions. More comparative results and

[4] It is assumed that the remaining $N-T-1$ participating TTPs have no more insight in the corruption than any other outside authority. This is supported by the technology.

[5] It is assumed that defecting within the corrupt coalition may be sufficiently avoided or punished by the remaining members of the coalition. This means that the situation where a single member of the coalition tries to gain the others' shares without supplying his own is not considered.

policy recommendations are provided and discussed in Section 4, and Section 5 concludes.

2 Threshold Trust

The purpose of a TTP is to confidentially handle private information according to a prescribed protocol. In this paper failure to do so is considered corruption. Corruption that does not involves the TTP, e.g. bidding rings, is not considered in this paper.

Corruption may either be performed internally by the TTP or in coordination with an outside party that gains from the corruption. A simple example is a second price sealed bid auction where the price may be manipulated by an extra bid just below the highest bid. This is clearly valuable to the seller. Also, if the TTP's salary depends on the selling price, corruption may be directly beneficial for the TTP as well.

With NT-TTP, the choice of N and T are fundamental design variables that in different ways counteract corruption. To illustrate the differences between the two trust institutions, consider the following three general security concerns (Pfleeger and Pfleeger (2003)):

Integrity: Prevent manipulation of the protocol.

Confidentiality: Prevent revelation of information outside the protocol.

Availability: Prevent the protocol from being blocked.

Corruption can be categorized as a violation of one or more of these three concerns. Clearly, all three concerns may be directly violated by a T-TTP. This is opposed to NT-TTP, where violation of each of the three concerns requires a different number of the N TTPs to cooperate[6]. Table 1 illustrates the required coordination in order to violate the three concerns in case of $N = 5$ and varying threshold (T).

Table 1: *Required coordination to violate the three security concerns with NT-TTP.*

NT setup	Integrity	Confidentiality	Availability
(5,1)	5	2	4
(5,2)	5	3	3
(5,3)	5	4	2
(5,4)	5	5	1

[6] Corruption by software engineers is not considered in this paper.

Manipulating the protocol will involve all N TTPs. Therefore, integrity is independent of the chosen threshold, unlike confidentiality and availability that is inversely dependent on the threshold. Compromising confidentiality, may be done independently by $T + 1$ TTP without any traceable signals outside the coalition. On the other hand, $N - T$ TTPs can prevent the protocol from being executed. This creates a fundamental trade-off between confidentiality and availability.

In this paper it is assumed that the gain from a successful corrupt act is the same for a T-TTP and a coalition of $T+1$ out of N in case of NT-TTP. Hereby, we basically assume that breaking the confidentiality is both necessary and sufficient to get the high payoff from playing corrupt.

The neglected higher integrity with NT-TTP may be supported by the following statements: 1) that knowledge about the private inputs is sufficient to perform a corruption, like in the case of the second price auction and 2) that breaking the confidentiality is less detectable than manipulating the protocol, since the protocol is public and the public result has to correspond with each participant's submitted information.

Availability seems of less importance in terms of corruption, although it may be of value to prevent the protocol from being performed. As illustrated in Table 1, the higher threshold the more coalitions may prevent availability. Especially with the maximum threshold of $N - 1$ where each individual TTP may veto the protocol. Apart from intended blocking, unintended dropout may be a significant problem if N is large or if timely precision is important, e.g. in most online services. Also, if just one of the keys are lost, the collected information is useless. On the other hand, setting $T = N - 1$ and let each of the participants constitute a TTP makes up a perfect trust institution in terms of confidentiality.

Apart from the three security concerns, the complexity of secure multiparty computation is significant and depends on N and T. In general the computation time increases as N, T and the relation $\frac{T}{N}$ increases. Altogether, there is no a priori dominating choice of N and T.

In the analysis we will assume that the N members are identical and independent and discuss the numeric choice of N and T. Though, in reality one might have prior expectations about likely coalitions among the N members, and use this to select the threshold. In general one may consider the likely gain from corruption by any $T + 1$ coalition and select T according to this, when defining stable coalitions in a cooperative game. As an example, consider a sealed bid double auction between one seller and one buyer, where the mediator's job is to compute the trading price, e.g. the average of the two submitted bids. Consider a NT-TTP with $N = 3$ and $T = 1$ where the TTPs are the seller, the buyer and a consultancy house. Since likely corruption may happen in a coalition between either the buyer or the seller and the consultancy house, the required coordination with T-TTP is the same. On the other hand, if the NT-TTP consisted of three independent consultancy houses, any corruption would require fundamentally more coordination.

3 The Game Theoretic Modeling

This section presents the applied game theoretic models and some immediate comparative results. As mentioned above, for a successful coalition to maximize the gain from corruption, it is assumed that it consists of exactly $T+1$ TTPs. It turns out that the preferred trust institution is determined by two counteracting effects:

The division effect: Unlike T-TTP, the gain from corruption is divided among less than the gain from playing honest with NT-TTP[7]

The coordination effect: Unlike T-TTP, corruption requires co-operation among more independent TTPs with NT-TTP

The modeling is first presented in a simple basic model, that only involves the division effect, and then extended to a model that captures both effects.

3.1 The Basic Model

It is assumed that the two trust institutions face the same competitive prices for a given mediation job. One may think of the job of handling the private bids in a second price sealed bid auction. **The TTP**'s opponent is customers (represented by **The Market**), who perceives **The TTP** as being reliable or unreliable. If **The TTP** is perceived unreliable, a large part of the customers drop the TTP, and the TTP gets a low payoff (V^l). If **The TTP** is perceived reliable, playing "honest" generates a medium payoff (V^m) while playing "corrupt" generates a high payoff (V^h). **The Market** has an advantage of using the same TTP but a disadvantage of facing corruption. The payoffs have the same properties as in the well known game of the prisoners' dilemma. Below, the game is presented as a 2×2 matrix game between **The TTP** and **The Market**.

$$\begin{array}{llcc} & & \multicolumn{2}{c}{\textbf{The market}} \\ & & \textit{Reliable} & \textit{Unreliable} \\ & \textit{Honest} & (V^m, V^h) & (V^l, V^m) \\ \textbf{The TTP} & & & \\ & \textit{Corrupt} & (V^h, V^l) & (V^l, V^m) \end{array}$$

The game has a weakly dominating Nash equilibrium in pure strategies, where **The TTP** plays corrupt, and *The market* considers **The TTP** to be unreliable and chooses another TTP. Although both players would have been better off by playing respectively "honest" and "reliable", it is not a best response. However,

[7] Unless $T = N - 1$ where both gains from corruption as well as honest play are divided among N TTPs.

if the two players repeatedly meet and play the same game, supporting the cooperative strategy (honest, reliable) may be possible.

We assume that the players play this game in every period and that there is always a positive probability for another period - meaning that we consider an infinite number of periods. Also, we will consider the so-called *Grim trigger strategy*, which in this setup means that **The Market** plays "reliable" as long as **The TTP** plays "honest". If **The TTP** plays "corrupt" **The Market** will play "unreliable" in the next period and forever hereafter. It is well known, that with a sufficiently high discount factor, future punishment may ensure that the cooperative strategy (honest, reliable) is a Subgame Perfect Nash Equilibrium, see e.g. the seminal paper Friedman (1971) for an introduction to this so-called supergame. The intuition is simple; a higher discount factor puts higher weights on future punishments which at some point make it economically optimal to play "honest" to avoid punishment.

Since the focus is on comparing two trust institutions facing the same supergame, the simplified assumption of repeating the game as well as the punishment period in infinity is of less importance. Nevertheless, one may e.g. implement a return to a co-operative equilibrium after a given number of punishment periods, see Abreu (1986). This may reflect a situation where trust is reestablished after a period of corruption.

The expression below provides the smallest discount factor that makes "honest" the T-TTP's best response.

$$\begin{array}{l} V^{\text{honest}} > V^{\text{corrupt}} \Leftrightarrow \\ V^m \cdot \frac{1}{1-\delta} > V^h + V^l \cdot \frac{\delta}{1-\delta} \Leftrightarrow \\ \frac{V^h - V^m}{V^h - V^l} < \delta \end{array} \tag{1}$$

As mentioned, the gain from playing "honest" is divided among N and the gain from playing "corrupt" is divided only among $T+1$ with NT-TTP. Therefore, the player "NT-TTP" represents both the whole group of N TTPs as well as the successful coalition of $T+1$ TTPs. If NT-TTP plays "corrupt", the remaining $N-T-1$ TTPs are unaware of any corruption before the subsequent period. When corruption is detected, the assumption is that the NT-TTP institution continues with a smaller part of the market which is collectively shared among all N TTPs. Setting $V^l = 0$ is a simple way to model the case where the NT-TTP institution breaks down when corruption is detected - this is discussed further in section 4. Like before, the expression below provides the smallest discount factor that makes "honest" the NT-TTP's best response[8].

$$\begin{array}{l} V^{\text{honest}} > V^{\text{corrupt}} \Leftrightarrow \\ \frac{V^m}{N} \cdot \frac{1}{1-\delta} > \frac{V^h}{T+1} + \frac{V^l}{N} \cdot \frac{\delta}{1-\delta} \Leftrightarrow \\ \frac{\frac{N}{T+1} V^h - V^m}{\frac{N}{T+1} V^h - V^l} < \delta \end{array} \tag{2}$$

[8] The result is independent of a proportional increase in the payoff i.e. independent of the market share.

Note that the difference between the two situations is $\frac{N}{T+1}$, which is larger than or equal to 1. Therefore, all successful coalitions among at least $T+1$ and no more than $N-1$ require a higher discount factor to support "honesty" as oppose to T-TTP. This means that it is easier to support honesty with T-TTP in the sense that the required valuation of future punishments is less for T-TTP than for NT-TTP. This is illustrated in Figure 1 where payoffs are fixed at $V^l = 3, V^h = 10$ and $V^m = 3 + w, w \in [0; 7]$ and 4 different choices of N and T are pictured ((5,1), means $N = 5$ and $T = 1$).

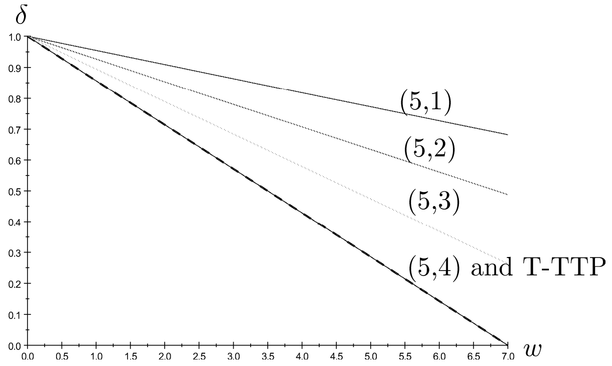

Figure 1: *T-TTP is a relatively more trustworthy institution if all corruption is detected.*

3.2 The Extended Model

In the analysis above, all corruption is detected with certainty and punished in the following period. In the following we will assume that less than all incidences of corruption is detected. Though, if corruption is detected the TTP is punished in the following period and forever hereafter as before.

It is assumed that the payoffs are expected payoffs, supported by overlapping intervals, such that the realized payoffs do not leave **The Market** player with any certain signals about corruption. Corruption may be detected by the market participants or some third party supervising the market. The detection of corruption is assumed to be the same for all TTPs in every period, no matter if they operate individually as T-TTP or as a member of NT-TTP. Also, the detection rates for the individual TTPs are assumed to be independent.

Now, let β be the probability that corruption by a given trust institution (T-TTP or NT-TTP) in a given period is *not* detected. If $\beta = 0$ corruption is always detected as in the model above. For $\beta > 0$, the TTP can either be detected and receive V^l forever hereafter or move undetected to the next period. If no corruption is detected, the game is repeated, and the TTP plays "corrupt" again and receive V^h, which may or may not be detected and punished from the subsequent period a.o. Figure 2 illustrates the different paths a TTP can take.

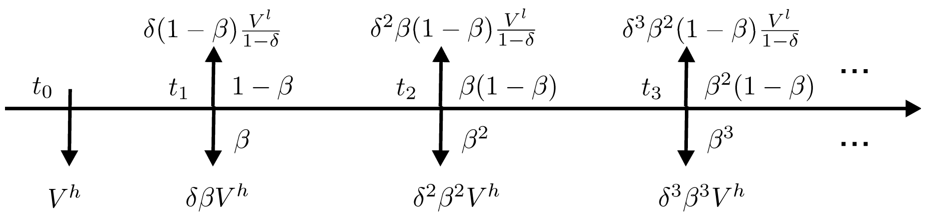

Figure 2: *The discounted expected payoffs on a timeline.*

In terms of comparing the two institutions, let α be the probability that corruption by a given TTP in a given period is *not* detected. If the TTPs involved in NT-TTP are independent, the highest β is α^{T+1}. For a T-TTP $\beta = \alpha$. Clearly, the detection rate is an increasing function of the threshold T. Though, the relative higher detection rate between T-TTP and different threshold values for NT-TTP depends on the actual detection rate. Figure 3 pictures $\alpha - \alpha^{T+1}$ for different values of α and T. Hereby, the relative gain from the coordination effect is illustrated. For high and low values of α, the relative coordination effect is small. Also, the maximum relative coordination value increases with T. In the following, we study how this coordination effect counteracts the division effect, illustrated in the previous subsection.

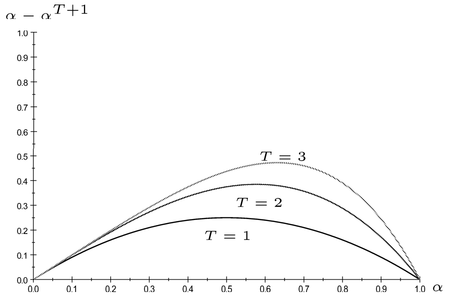

Figure 3: *The relative coordination effect.*

Now, weighting the different paths a T-TTP can take (see Figure 2), the inequality that makes "honest" the economically best response is given as:

$$\underbrace{\frac{V^m}{1-\delta}}_{\text{honest}} > \underbrace{V^h + \frac{\delta\alpha V^h}{1-\delta\alpha}}_{\text{not detected}} + \underbrace{\frac{\delta}{1-\delta\alpha}\frac{(1-\alpha)V^l}{1-\delta}}_{\text{detected}} \tag{3}$$

As above, solving for δ provides a lower bound on the discount factor in order to support "honesty":

$$\begin{aligned}
&\tfrac{V^m}{1-\delta} > V^h + \tfrac{\delta\alpha V^h}{1-\delta\alpha} + \tfrac{\delta}{1-\delta\alpha}\tfrac{(1-\alpha)V^l}{1-\delta} \Leftrightarrow \\
&(1-\delta\alpha)V^m > (1-\delta)V^h + \delta(1-\alpha)V^l \Leftrightarrow \\
&0 > -\delta(V^h - V^l - \alpha V^m + \alpha V^l) + V^h - V^m \Leftrightarrow \\
&\delta > \tfrac{V^h - V^m}{V^h - V^l - \alpha V^m + \alpha V^l}
\end{aligned} \tag{4}$$

Likewise, the discount factors that makes "honest" the NT-TTPs best response is given as:

$$\begin{aligned}
&\tfrac{1}{N}\tfrac{V^m}{1-\delta} > \tfrac{1}{T+1}V^h + \tfrac{1}{T+1}\tfrac{\delta\alpha^{T+1}V^h}{1-\delta\alpha^{T+1}} + \tfrac{1}{N}\tfrac{\delta}{1-\delta\alpha^{T+1}}\tfrac{(1-\alpha^{T+1})V^l}{1-\delta} \Leftrightarrow \\
&(1-\delta\alpha^{T+1})V^m > (1-\delta)\tfrac{N}{T+1}V^h + \delta(1-\alpha^{T+1})V^l \Leftrightarrow \\
&0 > -\delta(\tfrac{N}{T+1}V^h - V^l - \alpha^{T+1}V^m + \alpha^{T+1}V^l) + \tfrac{N}{T+1}V^h - V^m \Leftrightarrow \\
&\delta > \tfrac{\frac{N}{T+1}V^h - V^m}{\frac{N}{T+1}V^h - V^l - \alpha^{T+1}V^m + \alpha^{T+1}V^l}
\end{aligned} \tag{5}$$

A comparison of the two institutions while playing the same supergame is given in Figure 4. The result provides the lower bound on the discount factor δ in order to support the TTP to play "honest" as a function of α for respectively T-TTP and NT-TTP. The other parameters are set to: $N = 5, T = 2, V^l = 3, V^m = 4$ and $V^h = 10$.

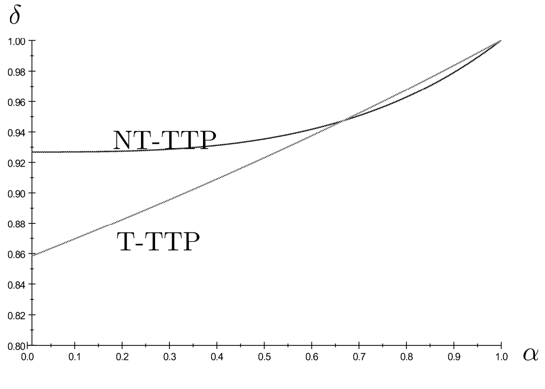

Figure 4: *Comparing T-TTP and threshold trust with* $N = 5$ *and* $T = 2$.

The figure illustrates the counteraction between the division and the coordination effect. If all corruption is detected, the division effect makes it easier to support honesty with T-TTP. On the other hand, if no corruption is detected, there is no economic reasoning for any TTP to play "honest". The interesting point is where the curves cross - where the division effect is suppressed by the coordination effect and makes NT-TTP a relatively more trustworthy institution.

In the following we present some comparative results that follow the same logic as in Figure 4. We explore where the two curves cross with respect to the different application specific parameters. First we consider the choice of N and T and then the payoff matrix.

3.3 The Choice of N and T

To explore the effect of increasing the threshold, N is fixed at 7 and the threshold T is varied. In Figure 5, T-TTP is the thick dashed curve and the intersecting curves represent the 6 different NT-TTP setups. With (7,0) each of the individual TTPs may play "corrupt" and get the high payoff, while payoff from playing "honest" should be divided among all 7 TTPs. Hereby the division effect dominates and T-TTP will always be a relatively more trustworthy institution. As T increases, α for which NT-TTP is preferred increases rapidly. For $T = 5$ NT-TTP is a dominating choice with the chosen payoffs: $V^l = 3, V^m = 4$ and $V^h = 10$. For $T = 6$ NT-TTP will always be a relatively more trustworthy institution. To see this, note that when $T = 6$ the division effect disappears and the coordination effect makes NT-TTP more trustworthy.

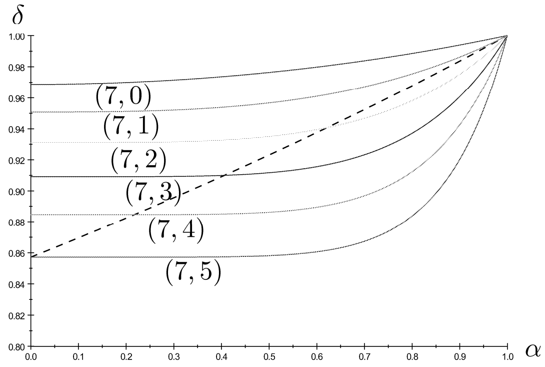

Figure 5: *The effect of T.*

Since majority trust has a computational advantage it may be relevant to consider the effect of majority trust with an increasing N. To explore this, majority trust based on (3,1), (5,2), (7,3), (9,4) and (11,5) are compared to T-TTP. Figure 6 provides the required discount factor as a function of α for each of the different setups. T-TTP is the thick dashed curve, and the intersecting curves represent the 5 different NT-TTP setups. Figure 6 shows that N has a small positive but diminishing effects in favor of NT-TTP. Also the Figure show that the order of the curves representing the different NT-TTP setups changes for smaller α. However, applying the same payoff matrix in the two institutional setups, an increasing N will always make NT-TTP a relatively more trustworthy institution. The intuition is that the division effect is approximately the same as N increases, while the coordination effect increases as T increases. Like before, the chosen payoffs are: $V^l = 3, V^m = 4$ and $V^h = 10$.

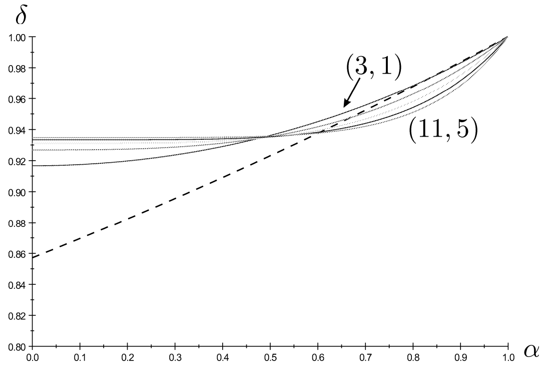

Figure 6: *The effect of N.*

3.4 The Payoff Matrix

Here we explore the relative trustworthiness between the two trust institutions when facing the same but varying payoffs. NT-TTP is represented as majority trust based on $N = 5$ and $T = 2$.

As mentioned before, Figure 7 provides the required discount factor as a function of α for T-TTP and NT-TTP. The two institutions (T-TTP and NT-TTP) are compared in three different situations (A,B and C), corresponding to different payoffs from corruption and punishment. In situation A, the gain from corruption is high and the punishment low (9:1). In situation C, V^m is raised such that the gain from corruption is small and the punishment higher (1:9). Situation B is in between. As the Figure illustrates, the different corruption/punishment scenarios have a significant effect on the relative trustworthiness of the two institutions. As the gain from corruption decreases and the punishment increases, the T-TTP becomes relatively more trustworthy. The intuition is that with NT-TTP the high gain from cooperation is collectively shared among all N TTPs, while the small extra gain from corruption is only shared among the $T+1$ TTPs. Therefore, NT-TTP requires a higher weight on the future since punishment is relatively less.

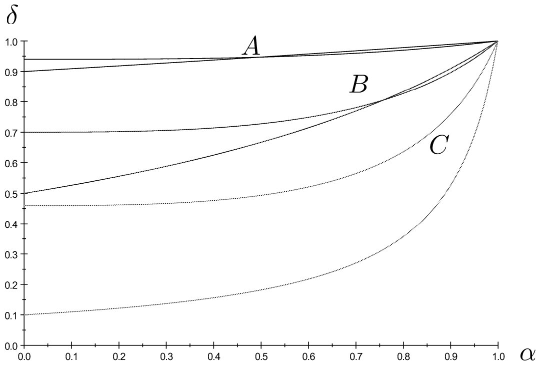

Figure 7: *Corruption and punishment.*

To conclude, even though no single TTP holds any information in case of NT-TTP, traditional trust based on a single TTP (T-TTP) can be more trustworthy. On the other hand, the choice of N and T can make NT-TTP a more trustworthy institution. This can happen by diminishing the division effect (by diminishing the relative difference between N and T) and increasing the coordination effect (by increasing the size of T).

4 Discussion and Policy Recommendation

From the previous section we have that neither of the two trust institutions is a dominating choice per se. In this section we discuss differences between the two trust institutions as well as initiatives that may counteract corruption and make NT-TTP a relatively more trustworthy institution.

4.1 NT-TTP Has a Different Cost Structure

It is sometimes suggested that a NT-TTP is a less costly way to establish trust. The basic argument is that no sensitive information is available to the individual TTP. This is opposed to a traditional TTP, were strict procedures prevent any leakage of information. On the other hand, the NT-TTP involves more TTPs. In terms of the game theoretic modeling above, it is clear that the TTP institution that makes the most profit is the most trustworthy. The intuition is simply that the more profitable TTP has more to lose from playing “corrupt”.

4.2 Breakdown of The NT-TTP

In the analytical model it is assumed that NT-TTP continues in a smaller market after corruption is detected - a cost that is collectively covered. However, corruption may cause the NT-TTP to breakdown with one or more TTPs leaving. Nevertheless, this may cause the NT-TTP institution to be more trustworthy for several reasons.

In case that the remaining group of $N - T - 1$ TTPs leave the NT-TTP after detected corruption, they may experience a temporary loss of reputation or business opportunities. This risk of being associated with a corrupt NT-TTP may affect the behavior of the TTPs in two opposite directions. In the initial phase of establishing the NT-TTP, the risk of sullying a good name may bias the selecting in a positive direction. On the other hand, if the TTPs expect the others to form a corrupt coalition, they might as well try to join it to get a part of the high payoff. However, if the later effect causes the corrupt coalition to include more than $T+1$ TTPs, the division effect makes corruption less attractive. Therefore, in both cases the NT-TTP becomes relatively more trustworthy.

A more direct effect comes from the fact that if a NT-TTP breaks down it can not continue in a smaller market as opposed to a T-TTP. This makes the punishment more harsh to the NT-TTP and therefore corruption less tempting. Figure 8 illustrates the situation where $V^l = 0$ for the NT-TTP[9] and respectively 0,1,2 and 3 for the T-TTP. The other parameters are chosen to be: $N = 3$, $T = 1$, $V^m = 4$ and $V^h = 10$. As the Figure illustrates, a relatively more harsh punishment makes NT-TTP a relatively more trustworthy institution.

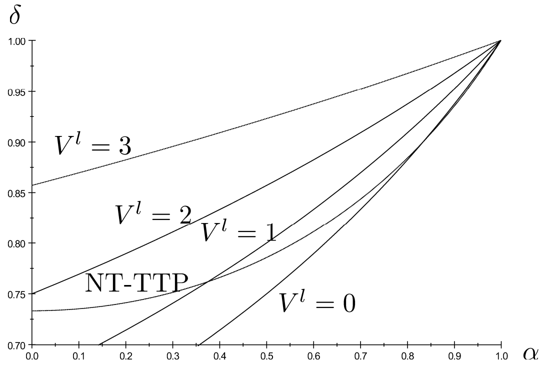

Figure 8: *If NT-TTP breaks down it becomes relatively more trustworthy.*

[9] If the punishment V^l is positive, a comparison between T-TTP and NT-TTP may provide the NT-TTP with a weird advantage, since the "positive" punishment is divided among a smaller number.

4.3 Counteract Stable Coalitions

If the information about corruption is disseminated outside the successful coalition, the risk of being detected increases, or the coalition may be forced to expand the coalition or bribe outside parties. Therefore, it seems reasonable to assume that, if possible, the coalition will consist of the same $T + 1$ TTPs in every period in order to avoid disseminating information outside the coalition. Though, if one or more of the N TTPs are replaced in each round, playing corrupt in every period may involve new coalitions. Hereby a successful group of $T+1$ colluding TTPs in a given period should either choose to 1) bribe outside TTPs that hold superior information about likely corruption, 2) accept a higher detection rate or 3) play only corrupt when the same TTPs meet. Either way, the NT-TTP institution becomes relatively more trustworthy since the expected gain from corruption is lower one way or the other.

To give an example, consider the following simple extension where 4 TTPs are initially assigned and where the NT-TTP consists of $N = 3$ and $T = 1$. Now, in every new period, one of the 3 TTPs is replaced with the 4'th TTP. Hereby, any given successful coalition of 2 TTPs will only meet every second period. Assuming that a successful coalition of 2 TTPs decide to play corrupt only when they meet, the situation may be modeled simply by lowering the payoff from playing "corrupt" with 50 % in the present model. However, the cost is that 4 instead of 3 TTPs have to share the same gain from playing "honest" as well as the payoff in the punishment period. Figure 9 illustrates 4 different situations. T-TTP and NT-TTP^A represent the benchmark with the usual payoffs: $V^l = 3, V^m = 4$ and $V^h = 10$. 50 % less gain from corruption reduces V^h to 7, and dividing the honest gain and punishment with 4 instead of 3 reduces V^l and V^h to respectively $2\frac{1}{4}$ and 3, this is represented by NT-TTP^B. Although the Figure shows significant improvement from introducing a fourth TTP, it is not unambiguously because V^l and V^m are relatively lower. However, if the fourth TTP is subsidized or represents a public authority, the effect is unambiguously in favor of NT-TTP. This is illustrated in Figure 9 by NT-TTP^C, where the payoffs are $V^l = 3, V^m = 4$ and $V^h = 7$.

4.4 NT-TTP and Leniency Programs

By assumption, if corruption is detected, the punishment is V^l in every future period. In reality there might be an additional penalty if the corruption can be proven in court. To the extent that corruption can be proven in court, an additional leniency program may counteract corruption with NT-TTP. With a leniency program a member of a corrupt coalition gets a reduced penalty for helping the authorities in court, see e.g. Motta and Polo (2003). Therefore, with a positive probability of being convicted in court (and detected in the

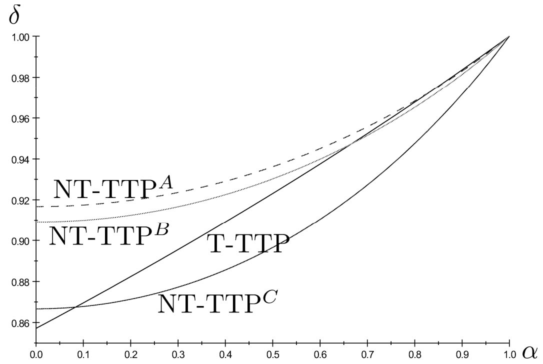

Figure 9: *Replacing one TTP in each period,* $N = 3$ *and* $T = 1$.

first place), each of the colluding TTPs may be tempted to cooperate with the authorities for economic reasons. Although, the real effect of a leniency program may be limited, it will cause the detection rate to be higher for each member of a NT-TTP.

5 Conclusion

Traditional trust is compared to threshold trust in a repeated game where corruption is a tempting deviation. If corruption is detected, a part of the market chooses another TTP, and the TTP is punished by a low payoff forever.

NT-TTP has a fundamental division problem where the gain from corruption is divided among less than are honest gains. Therefore, if all corruption is detected, corruption is a relatively more tempting deviation with NT-TTP, which makes T-TTP a more trustworthy institution.

On the other hand, if not all incidents of corruption are detected, the trustworthiness depends on the actual configuration of the NT-TTP. Since corruption with NT-TTP requires cooperation among at least two independent TTPs, the chance of being detected is higher with NT-TTP. This coordination effect counteracts the uneven division of payoffs from corruption and honest play.

One computational efficient configuration is majority trust, where any majority can use and misuse the NT-TTP. Majority trust based on a large number of TTPs is a more trustworthy institution. Increasing the threshold makes the NT-TTP an unambiguously better choice. Setting $T = N - 1$ completely removes the problem of uneven division. However, higher confidentiality is at the cost of availability, since anyone of the N TTPs can prevent the protocol from being performed.

The TTP institution with the lowest costs is likely to be the most profitable and therefore, also the most trustworthy TTP. Since no single member of NT-TTP holds any information, the variable costs for the individual TTPs is likely to be low. This is unlike a T-TTP, where strict and (probably) costly procedures are required for being reliable. This likely difference in cost structures may be in favor of the NT-TTP.

As a public authority, the structure of NT-TTP allows for efficient intervention. Introducing a fourth (subsidized) TTP that systematically replaces the TTPs in a simple majority trust based on 3 TTPs, makes the NT-TTP institution relatively more trustworthy. On the other hand, the classical leniency programs may have limited effect due to the problem of proving corruption.

The modeling neglects the cooperative game within the NT-TTP. Division of the gains may cause instability among the N TTPs while playing the cooperative strategy and during punishment period, like among the $T + 1$ TTPs, while playing the deviation strategy. As an example, if the NT-TTP breaks down due to instabilities among the N TTPs after corruption is detected, the punishment period may be relatively more harsh to NT-TTP, which makes it relatively more trustworthy. Though, incorporating the cooperative requirements in the present non-cooperative game is one of the more challenging extensions.

Another future challenge is to conduct laboratory or field experiments to uncover how the two trust institutions are perceived. The trust institution that is perceived to be the most trustworthy may attract a larger part of the market and hereby become even more trustworthy as a result of corruption being less tempting.

References

Abraham, I., Dolev, D., Gonen, R. and Halpern, J.: 2006, Distributed computing meets game theory: robust mechanisms for rational secret sharing and multiparty computation, *PODC '06: Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing*, ACM Press, New York, NY, USA, pp. 53–62.

Abreu, D.: 1986, Extremal equilibria of oligopolistic supergames, *Journal of Economic Theory* **39**, 191–225.

Ben-Or, M., Goldwasser, S. and Wigderson, A.: 1988, Completeness theorems for non-cryptographic fault-tolerant distributed computation, *Proc. of ACM STOC 1988*, pp. 1–10.

Bogetoft, P., Christensen, D. L., Damgaard, I., Geisler, M., Jakobsen, T., Krøigaard, M., Nielsen, J. D., Nielsen, J. B., Nielsen, K., Pagter, J., Schwartzbach, M. and Toft, T.: 2008, Multiparty computation goes live, Cryptology ePrint Archive, Report 2008/068. `http://eprint.iacr.org/`.

Bogetoft, P., Damgaard, I., Jacobsen, T., Nielsen, K., Pagter, J. and Toft, T.: 2005, Secure computing, economy, and trust - a generic solution for secure auctions with real-world applications, *Report RS-05-18*, Basic Research in Computer Science.

Bogetoft, P., Damgaard, I., Jacobsen, T., Nielsen, K., Pagter, J. and Toft, T.: 2006, A practical implementation of secure auctions based on multiparty integer computation, *Proceedings of Financial Cryptography 2006*, Lecture Notes in Computer Science, vol. 4107, Springer Verlag, pp. 142–147.

Chaum, D., Crépeau, C. and Damgaard, I. B.: 1988, Multi-party unconditionally secure protocols, *Proc. of ACM STOC 1988*, pp. 11–19.

Friedman, J. W.: 1971, A non-cooperative equilibrium for supergames, *Review of Economic Studies* **38**, 1–12.

Gibbard, A.: 1973, Manipulation of voting schemes: A general result, *Econometrica* **41**, 587–601.

Goldreich, O., Micali, S. and Wigderson, A.: 1987, How to play any mental game or a completeness theorem for protocols with honest majority, *Proc. of ACM STOC 1987*, pp. 218–229.

Halpern, J. and Teague, V.: 2004, Rational secret sharing and multiparty computation: extended abstract, *STOC '04: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, ACM, New York, NY, USA, pp. 623–632.

James Jr., H. S.: 2002, The trust paradox: A survey of economic inquiries into the nature of trust and trustworthiness, *Journal of Economic Behavior & Organization* **47**, 291–307.

Maleka, S., Shareef, A. and Rangan, C. P.: 2008, Rational secret sharing with repeated games. forthcoming as proceeding from ISPEC 2008.

Malkhi, D., Nisan, N., Pinkas, B. and Sella, Y.: 2004, Fairplay - a secure two-party computation system, Presented at Usenux Security 2004.

Motta, M.: 2004, *Competition Policy*, Cambridge University Press.

Motta, M. and Polo, M.: 2003, Leniency programs and cartel prosecution, *International Journal of Industrial Organization* **21**, 347–379.

Myerson, R. B.: 1979, Incentives compatibility and the bargaining problem, *Econometrica* **47**, 61–73.

Pfleeger, C. P. and Pfleeger, S. L.: 2003, *Security in Computing*, 3rd edn, Prentice Hall.

Porter, R. H.: 1983, Optimal cartel trigger price strategies, *Journal of Economic Theory* **29**, 313–338.

Rotemberg, J. J. and Saloner, G.: 1986, A supergame theoretic model of business cycles and price wars during booms, *American Economic Review* **76**, 390–407.

Shamir, A.: 1979, How to share a secret, *Commun. ACM* **22**(11), 612–613.