

**Real-Time Gross Settlement and hybrid payment systems:
a comparison**

Matthew Willison

Working Paper no.252

Bank of England, Threadneedle Street, London, EC2R 8AH.

Email: matthew.willison@bankofengland.co.uk

The views expressed in this paper are those of the author, and do not necessarily reflect those of the Bank of England. I gratefully acknowledge the help and insightful comments received from Paul Bedford, Stephen Millard, Marco Polenghi and Hyun Song Shin. I would also like to thank seminar participants at the Bank of England and attendees of the ‘Economics of Payments’ conference at the Federal Reserve Bank of Atlanta. I also thank Simon Buckle, with whom I began this work. All remaining errors and omissions are the sole responsibility of the author.

Copies of working papers may be obtained from Publications Group, Bank of England, Threadneedle Street, London, EC2R 8AH; telephone 020 7601 4030, fax 020 7601 3298, email mapublications@bankofengland.co.uk.

Working papers are also available at www.bankofengland.co.uk/wp/index.html.

The Bank of England’s working paper series is externally refereed.

© Bank of England 2004

ISSN 1368-5562

Contents

Abstract	5
Summary	7
1 Introduction	9
2 The model	11
2.1 Banks	11
2.2 Comparing different payment system designs	15
3 Real-Time Gross Settlement	16
4 Hybrid payment systems	20
4.1 Modifications to the model	20
4.2 Equilibria under hybrid payment systems	22
4.2.1 H1: Payment offset in the afternoon	22
4.2.2 H2: Payment offset in the morning	25
4.2.3 H3: Payment offset in the morning and afternoon	25
4.3 Information asymmetry	25
4.3.1 Public information	26
4.3.2 Additional uncertainty	27
4.4 Multilateral offset	28
5 Conclusions	28
References	30

Abstract

This paper contrasts Real-Time Gross Settlement and hybrid payment systems that are based on payment offset, using a two-period, multi-bank model. The comparison is performed according to two criteria: liquidity needs and speed of settlement. We assume that the existence of a payment is common knowledge but that the specific degree of time-criticality of a payment is the private information of the bank sending the payment. Hybrid payment systems are shown to outperform Real-Time Gross Settlement when payments are offset in the first period and when they are offset in both periods. This suggests that in a hybrid system, the offsetting facility should be in operation all day, or, at the very least, for some time after the system opens in the morning. A system in which the offsetting facility was only switched on late in the day would not necessarily be preferred to Real-Time Gross Settlement. These results are shown to be robust to changes in the transparency of the central queue of payments awaiting offset. However, this robustness may not hold with different forms of information asymmetry.

Summary

This paper considers the issue of payment system design. Real-Time Gross Settlement (RTGS) has become the foremost system for the settlement of high-value payments in developed economies. The rationale behind the trend towards RTGS has been the perceived need to reduce the risk potentially found in deferred net settlement (DNS), the predominant system for settling high-value payments previously. RTGS entails payments being settled on a gross basis in real time. As a consequence, credit risk and settlement risk between settlement banks are eliminated. But RTGS does not dominate DNS in all respects. With payments settled on a gross basis, settlement banks' liquidity needs under RTGS are greater than those under DNS. This could induce settlement banks to delay payment activity in order to wait for incoming payments that can then be used as liquidity. Central banks have mitigated these problems to an extent by providing intraday liquidity. Of course, lending liquidity generates credit risk for the central bank. Thus, this lending is collateralised to remove this risk. Yet, the benefits from reducing the risk associated with DNS systems are considered to exceed the costs of greater liquidity needs; hence, the number of RTGS systems has grown.

The debate surrounding the optimal basis for settling payments has shifted of late with the advent of so-called hybrid payment systems. A hybrid, as the name suggests, combines features of both RTGS and DNS. More precisely, a hybrid typically takes one of the designs and augments it with features associated with the other design. Given that RTGS is the prime payment system design, recent debate has mainly concentrated on the benefits of complementing RTGS with a liquidity-saving feature called payment offset. A payment is offset when it is settled simultaneously with a set of other payments rather than being settled individually like in RTGS. When payments are settled simultaneously, the payments are self-collateralising, to the extent that their values are alike. Settlement banks only need liquidity equal to the net value of their payments in the set to settle these payments. An important design feature of any hybrid is that payments can be placed in a central queue. While payments are in this queue, the system operator searches for offsetting payments. Otherwise, payments can still be settled by RTGS without necessarily entering the central queue. The benefit of complementing RTGS with a payment offset facility is that liquidity needs can fall and the incentive to delay placing payments into the system is reduced, relative to RTGS; but offset does not necessarily reintroduce the risk present under DNS.

In this paper, we examine the issue of optimal payment system design. We compare the performance of an RTGS system against six hybrid systems based on payment offset. We assume that, when payments are offset, they are considered legally to be final and irrevocable. So the hybrid systems introduce no credit risk relative to the RTGS benchmark. We compare the system designs based on two criteria: their liquidity demands and the speed with which payments are settled. The second criterion captures the potential impact of operational risk, since any operational event will have a larger effect the more payments still remain to be settled when the event occurs. There is a trade-off between liquidity efficiency and exposure to operational risk.

We assume that some payments are time-critical. Hence, if a settlement bank delays payment settlement it faces a cost. Each settlement bank faces a trade-off between the costs of obtaining liquidity from the central bank and the costs of delaying payments when choosing how many

payments to settle in certain periods of the day. We assume that the existence of payments is common knowledge but the time-criticality of payments is not publicly known. This particular kind of information asymmetry enables us to focus on the problems settlement banks face in co-ordinating their usage of the central queue and how certain features of hybrid systems could potentially affect their ability to overcome these difficulties. We analyse the effects that the frequency at which payments are offset can have. We find that, when payments can only be offset late in the day, a hybrid will not offer improvements on RTGS according to either criterion. However, when offset occurs early or all day the first-best outcome is obtained.

1 Introduction

Real-Time Gross Settlement (RTGS) has become the foremost system for the settlement of high-value payments in developed economies. (See BIS (1997).) The rationale behind the trend towards RTGS has been the perceived need to reduce the risk potentially found in deferred net settlement (DNS), the predominant system for settling high-value payments previously. RTGS entails payments being settled on a gross basis in real time. As a consequence credit risk and settlement risk between settlement banks (hereafter, banks) is eliminated. But RTGS does not dominate DNS in all respects. With payments settled on a gross basis, banks' liquidity needs under RTGS are greater than those under DNS. This implies a burden for the wider economy primarily because it is a drain of resources but also because, in an attempt to save liquidity, banks delay payment activity in order to wait for incoming payments that can then be used as liquidity. Central banks (CBs) have mitigated these problems to an extent by providing intraday liquidity. Of course, lending liquidity generates credit risk for the CB. Thus, this lending is collateralised to remove this risk.⁽¹⁾⁽²⁾ Yet the benefits from reducing the risk one associates with DNS systems are considered to exceed the costs of greater liquidity needs; hence, the number of RTGS systems has grown.

The debate surrounding the optimal basis for settling payments has shifted of late with the advent of so-called hybrid payment systems. A hybrid, as the name suggests, combines features of both RTGS and DNS. (For a survey of this category of payment system design see McAndrews and Trundle (2001)). More precisely, a hybrid typically takes one of the designs and augments it with features associated with the other design. Given that RTGS is the prime payment system design, recent debate has mainly concentrated on the benefits of complementing RTGS with a liquidity-saving feature called payment offset. A payment is offset when it is settled simultaneously with a set of other payments rather than being settled individually like in RTGS. When payments are settled simultaneously, the payments are self-collateralising, to the extent that their values are alike. Banks only need liquidity equal to the net value of their payments in the set to settle these payments. The German large-value payment system, *RTGS^{Plus}* is one RTGS system that incorporates payment offset. However, there do exist hybrid designs with a DNS basis; eg NewCHIPS in the United States. An important design feature of any hybrid is that payments can be placed in a central queue. While payments are in this queue, the system operator searches for offsetting payments. Otherwise, payments can still be settled by RTGS or DNS without necessarily entering the central queue. The benefit of complementing RTGS with a payment offset facility is that liquidity needs can fall and the incentive to delay placing payments into the system is reduced, relative to RTGS, but offset does not necessarily reintroduce the risk present under DNS.

⁽¹⁾ Although lending is collateralised a CB still faces some market risk, the risk that the value of collateral declines dramatically during the day. To counteract this, a CB will typically require a haircut; ie the value of the collateral when it is posted has to exceed the value of the liquidity that is borrowed by some margin.

⁽²⁾ An alternative to collateralised provision of intraday credit is for banks to be allowed to run overdrafts on settlement accounts but be charged for the privilege. This is how the Federal Reserve provides liquidity in the US RTGS payment system, Fedwire. Pricing reduces but does not remove credit risk.

The relative merits of RTGS and hybrid payment systems have been analysed in a number of papers with the focus on various effects. Roberds (1999) compares systems with regard to their impact on banks' portfolio choices. He finds that a system with payment offset can produce results mimicking those found with RTGS and DNS. Other papers in the literature use simulation techniques to assess the advantages of hybrid payment systems. These include Leinonen and Soramäki (1999) and Johnson *et al* (2002). However, a major drawback with these simulations is that bank behaviour is taken to be exogenous. For example, the liquidity savings that hybrids can generate are calculated using RTGS data. Yet we would expect certain actions taken by banks, such as the times at which payments are submitted to the system, to depend on the payment system design in place. In this paper we allow the timing of payment activity to be endogenous and hence, enable it to vary across system designs.

In this paper, we examine the issue of optimal payment system design. We compare the performance of an RTGS system against six hybrid systems based on payment offset. We assume that, when payments are offset, they are considered legally to be final and irrevocable. So, the hybrid systems introduce no credit risk relative to the RTGS benchmark. We compare the system designs based on two criteria: their liquidity demands and the speed with which payments are settled. The reason for using this second criterion is that it captures the potential impact of operational risk, since any operational event will have a larger effect the more payments that still remain to be settled when the event occurs.

The criteria capture an important feature of RTGS. There is a trade-off between liquidity efficiency and exposure to operational risk. This trade-off ensures that the first-best is not attainable under RTGS. For example, if banks increase the amount of payments settled early, they cannot exploit incoming payments as a source of liquidity as heavily and hence, must post more liquidity. Meanwhile, posting less liquidity shifts payment activity towards later in the day. We assume that banks are concerned about exposure to operational risks because their customers are; ie some payments are time-critical. Hence, if a bank delays payment settlement it faces a cost. Each bank faces a trade-off between the costs of obtaining liquidity from the CB and the costs of delaying payments when choosing how many payments to settle in certain periods of the day. The result is that under RTGS banks will delay some payments in order to wait for incoming payments. This agrees with the findings of Angelini (1998), Bech and Garratt (2003), Buckle and Campbell (2003), and Kobayakawa (1997).

In our model we assume that banks face a particular form of information asymmetry. Whereas we assume that the existence of payments is common knowledge, the time-criticality of payments is not publicly known. Only a bank sending a payment knows the time-criticality of this payment. The implication is that although a bank may know how many payments other banks are each settling it does not know how many are directed to it until the payments have arrived. Since incoming payments are a substitute for (costly) borrowing from the CB, a bank is uncertain about the impact of other banks' actions on its pay-off. Of course, discussing this kind of information asymmetry/uncertainty is only sensible if there are more than two banks in the payment system. Even when there are more than two banks, the uncertainty will disappear when banks settle all or none of their payments. We assume this particular kind of information asymmetry as it enables us to focus on the problems banks face co-ordinating their usage of the central queue and how certain features of hybrid systems could potentially affect their ability to overcome these

difficulties. Payments can only be offset if offsetting payments are in the central queue at the same time. Thus, one might suspect that the transparency of the central queue will influence the likelihood of this by improving the information that banks have at their disposal. (Transparency determines whether and to what degree a bank can observe the payments that other banks have submitted to the central queue.)

We analyse the effects that the frequency at which payments are offset can have. When payments can only be offset late in the day, a hybrid will not offer improvements on RTGS according to either criterion. However, when offset occurs early or all day the first-best is obtained. What is surprising is that this result holds whether or not the central queue is visible to all banks. Because the time-criticality of payments is private information, a bank is unsure about what payments other banks will simultaneously submit to the central queue. This means that each of a bank's payments can be offset with a positive probability if each other bank submits at least one payment. Hence, banks will fully exploit the offset facility, by submitting all of their payments at the first opportunity they have. Thus, the largest possible set of payments is offset, early in the day. Introducing further uncertainty will not necessarily jeopardise the result unless we assume that using the central queue is costly for banks. Moving in the other direction, by assuming perfect information, there is the possibility of other equilibria when payments are offset early or all day.

Section 2 introduces the model. Equilibria under RTGS and hybrids are presented in Sections 3 and 4, respectively. Section 5 concludes.

2 The model

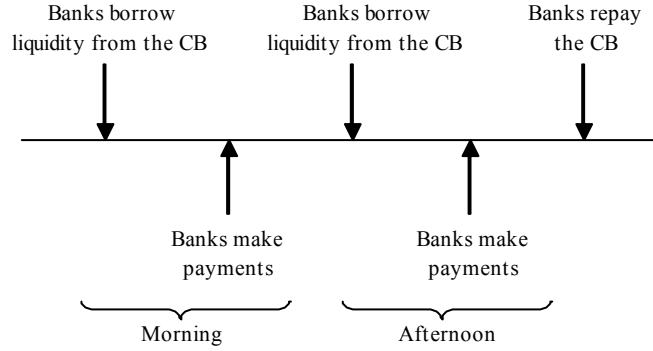
2.1 Banks

We present the model for the benchmark payment system design, Real-Time Gross Settlement (RTGS). There are n , risk-neutral, banks. Each bank has access to the payment system and the fixed cost of access is equal to zero. Assume that n is finite and exceeds two. Each bank has a single payment of one unit to send to each other bank through the payment system.⁽³⁾ In other words, each bank has $n-1$ payments to send. This 'pattern of payments' is common knowledge. Denote by $i \rightarrow j$ a payment from bank i to bank j .

The payment system (in RTGS) is modelled as a two-stage game. At each stage banks can settle payments. The two stages taken together is called a day with the first (second) stage referred to as period 1(2) or the morning (afternoon). The time-line for the game is illustrated in Chart 1.

⁽³⁾ The assumption that all payments have the same value means that the model is more tractable while not hindering the ability of the model to capture the salient features of RTGS. If payments did have different values the only (qualitative) effect would be to complicate the rules for prioritising payments but would not otherwise alter bank behaviour. But conformity in payment values means that when we analyse hybrid payment systems we will not consider 'payment splitting' nor imperfect payment offset. (Imperfect payment offset means that the each bank's payments in the set that are simultaneously settled can have different combined values.)

Chart 1



When a bank fails to settle one of its payments it is said to have cancelled this payment. Of course, in the real world such a payment could be settled on a subsequent day but here we restrict attention to just a single day. Cancellation of a payment results in a cost λ for a bank. This could be direct compensation for the customer requesting the payment or some expected loss of future business. Similarly, if a bank delays a payment it incurs a cost. Because of the two-stage assumption, payment delay occurs when a payment is not settled in the morning. Unlike the cancellation cost, the cost from delaying a payment differs across payments. That is, payments vary in their time-criticality. We denote the delay cost for a payment $i \rightarrow j$ by $\zeta(i \rightarrow j)$ and assume that $\zeta(\cdot)$ can take $n - 1$ possible values. We assume that $\zeta(i \rightarrow j) = \zeta(i \rightarrow k)$ if and only if $j = k$ that is, a bank has no two payments with the same delay cost. Thus, each bank has a single payment associated with each of the $n - 1$ possible delay costs.

A bank receives all requests to send payments at the beginning of the day. It orders its payments in an internal queue Q_i . The ordering of an internal queue is a function of delay costs. If the delay cost associated with a payment $i \rightarrow j$ exceeds that associated with another payment $i \rightarrow k$ it will sit closer to the front of Q_i . Thus, the most time-critical payment is at the head of the internal queue and the least is at the end of the queue.

Degrees of time-criticality are allocated to payments by a move by nature at the start of the day. A bank can perfectly observe the time-criticality of each of its own payments but not those of other banks' payments. Thus, we assume that a bank does not know the ordering of other banks' internal queues. This implies that even when a bank i is certain about how many payments another bank j is going to settle in a particular period i cannot be certain about whether it will receive an incoming payment from j in this period. This is unless j is settling all or none of its payments in its internal queue in this period.

A bank's action set in the morning is the set of payments it chooses to settle at this stage, a_i^1 . It is chosen from the morning action set A_i^1 that is shown in equation (1).

$$A_i^1 = \{\emptyset, \{1\}, \{1,2\}, \dots, Q_i \setminus \{n-1\}, Q_i\} \quad (1)$$

\emptyset means that no payments are settled (ie a bank chooses the empty set). $\{1\}$ is the payment at the head of the internal queue, $\{1,2\}$ is the first two payments in the queue and so on. $Q_i \setminus \{n-1\}$ is all the payments in the internal queue bar the payment at the very back of the internal queue, $\{n-1\}$. A_i^1 does not include all of the feasible combinations of payments a bank could settle in the morning. This follows since it does not make sense for a bank to settle a payment in a certain position in the internal queue if it does not also settle payments ahead of this particular payment. After all, if there is a common cost associated with settling a payment, a bank should show a preference for settling the payments that generate the greatest benefits; ie those payments that generate the largest savings in delay cost.

We can represent delay costs as a function of morning actions. The cost of delaying a payment $i \rightarrow j$, with a position ε in the internal queue, is $\beta(a_i^1) = \zeta(i \rightarrow j)$, where $a_i^1 = \{1, \dots, \varepsilon\}$.

In the afternoon a bank chooses an action a_i^2 from the action set A_i^2 shown in equation (2).

$$A_i^2 = \wp(Q_i \setminus a_i^1) \quad (2)$$

$\wp(Q_i \setminus a_i^1)$ is the power set of what remains of the internal queue after the morning action has been taken, $Q_i \setminus a_i^1$. The power set contains all possible combinations of payments in $Q_i \setminus a_i^1$ as well as \emptyset . That is, we are not restricting A_i^2 in the way that we restrict A_i^1 . The explanation for this is that there is no reason why a payment in the position ε in the internal queue, which is not settled in the morning, should not be settled in the afternoon when the payment in the position $\varepsilon-1$ is cancelled. This is because the cost of including any payment in a_i^2 at the expense of another payment is λ . The more payments there are in a_i^1 the fewer members there are of A_i^2 . The extreme case is when $A_i^2 = \emptyset$ because $a_i^1 = Q_i$.

To be capable of settling a payment in a given period a bank must have sufficient liquidity at the beginning of this period to do so. Let L_t^i be the liquidity available to bank i in period t . Liquidity cannot be withdrawn from the system during the day. All banks begin with zero liquidity. Incoming payments in the morning can only be used as liquidity in the afternoon while intraday credit provided by the CB is available at the beginning of both the morning and the afternoon. In order to borrow a unit of liquidity from the CB, a bank must post one unit of collateral (we assume that the CB does not require a haircut). Thus, we refer to borrowed liquidity as collateral. Banks repay the CB and redeem their collateral only at the end of the day. Each bank has sufficient eligible assets for it not to be credit constrained. Posting a unit of collateral in the morning generates a cost γ^1 . When a unit of collateral is posted in the afternoon a bank incurs a cost γ^2 . Assume that $\gamma^1 \geq \gamma^2 > 0$. This implies that a bank will not have an incentive to post collateral in the morning that it intends not to use until the afternoon.⁽⁴⁾ Three further

⁽⁴⁾ It is not clear which way the inequality between γ^1 and γ^2 should necessarily go. The direction of the inequality that we have assumed could be justified by the idea that posting assets as collateral ties those assets down for the rest of the day and the resulting restriction to a banks freedom implies a cost. Alternatively, one might believe it is harder to obtain eligible assets during the day than it is at the start of the day. In which case, the cost difference should be in the opposite direction. Investigating the case where $\gamma^2 \geq \gamma^1 > 0$ is a subject for future work.

assumptions are made about collateral costs: $\lambda > \gamma^2$; $\beta(\{1\}) > \gamma^1$; and $\beta(Q_i) \leq \gamma^1 - \gamma^2$. These assumptions imply that: (1) the cost of cancelling a payment is greater than the cost of acquiring the necessary liquidity to make the payment in the afternoon; (2) the cost of delaying the payment at the head of a bank's internal queue exceeds the cost of acquiring the necessary unit of liquidity for the payment to be settled in the morning; and (3) the cost of delaying the payment at the end of a bank's internal queue is less than the cost of obtaining the liquidity required for it to be settled in the morning.

We denote the level of collateral chosen by a bank in period t by c_i^t . Collateral choices are functions of the actions taken by all banks in both periods and must be such that a bank has non-negative liquidity at all points in the day. We derive two conditions that collateral choices have to satisfy for liquidity to never drop below zero, one condition for each period. In RTGS, what we call the morning collateral condition is for $i=1, \dots, n$,

$$L_i^1 = c_i^1 \geq |a_i^1| = \sum_{j \neq i} I(i \rightarrow j \in a_i^1) \quad (3)$$

Where $I(i \rightarrow j \in a_i^1) = 1(0)$ when a payment is (not) settled in the morning and $|a_i^1|$ is the value of payments in the set a_i^1 . As incoming payments cannot be used as liquidity until the afternoon, collateral must simply be equal to or greater than the value of outgoing payments in the morning. The afternoon collateral condition is a function of the liquidity available and the payments a bank wants to settle in this period. Once again, a bank cannot use incoming payments that arrive in the afternoon as a source of liquidity to settle payments in this period. Because the ordering of internal queues is private information, a bank does not know with certainty the amount of incoming payments at the beginning of the day when collateral is chosen. The level of afternoon liquidity is shown in equation (4).

$$L_i^2 = c_i^1 + c_i^2 - \sum_{j \neq i} I(i \rightarrow j \in a_i^1) + \sum_{j \neq i} I(j \rightarrow i \in a_j^1) \quad (4)$$

The afternoon collateral condition is

$$c_i^1 + c_i^2 \geq \sum_{j \neq i} I(i \rightarrow j \in a_i^1 \cup a_i^2) - \sum_{j \neq i} I(j \rightarrow i \in a_j^1) \quad (5)$$

Since collateral carries a cost, banks will seek to minimise the amount of collateral that they each hold in either period while satisfying both collateral conditions. It will always be optimal for a bank to use collateral posted in the morning to settle payments in the morning. As a result, equation (3) will always bind. Given this, we can rewrite the afternoon collateral condition as shown in equation (6).

$$c_i^2 = \max \left\{ 0, \sum_{j \neq i} I(i \rightarrow j \in a_i^2) - \sum_{j \neq i} I(j \rightarrow i \in a_j^1) \right\} \quad (6)$$

Equation (5) will hold with equality if $c_i^2 > 0$ but not necessarily if $c_i^2 = 0$. At the end of the day a bank must repay the CB the liquidity that it borrows to cover any shortfall of liquidity it expects

to have during the day. For a bank to have sufficient liquidity to repay the CB at the end of the day it must be that

$$c_i^1 + c_i^2 - \sum_{j \neq i} I(i \rightarrow j \in a_i^1 \cup a_i^2) + \sum_{j \neq i} I(j \rightarrow i \in a_j^1 \cup a_j^2) \geq c_i^1 + c_i^2 \quad (7)$$

Where the left-hand side of this inequality is the bank's end-of-day liquidity. We refer to this inequality as the sustainability condition. For the payment system to be sustainable each bank must be able to borrow the liquidity it needs for settlement purposes while being capable of repaying it at the end of the day. We can see from equation (7) that the payment system is sustainable, when each bank begins with zero liquidity, if and only if for each bank the value of its incoming payments over the day is equal to or greater than the value of its outgoing payments over the day. But equation (7) holds for every bank if and only if for each bank, the value of its incoming payments equals the value of its outgoing payments. Otherwise, interbank lending is required for all banks to have enough liquidity to repay the CB.⁽⁵⁾

2.2 Comparing different payment system designs

Different payment system designs are compared with reference to two criteria. The first is liquidity demand, in particular, the liquidity demand under a given payment system relative to the theoretical minimum amount required. As shown in Roberds (1993), the theoretical minimum is equal to liquidity demand under end-of-day, multilateral DNS. In this paper, where each bank has one payment to send to each other bank and all payments have a common value, the theoretical minimum liquidity demand for each bank equals zero.

The second criterion is the speed of settlement. This captures the exposure of the payment system to operational risk. We assume that the occurrence of an operational incident is exogenous to the system and that incidents can occur in both periods. The speed of settlement affects the impact of an operational incident on the system because the slower it is (ie the fewer payments that are settled in the morning) the more payments could be cancelled if an operational incident occurs in the afternoon. If an incident occurs in the morning, there is the possibility that the payment system could be back up and running come the afternoon, allowing payments to be settled. For simplicity we do not model explicitly the likelihood of such incidents nor their actual effects. Rather, we assume that banks fail to take into account all of the negative effects arising from operational disruption, while the system operator's concern about these effects is reflected in a simple preference for payments being settled in the morning.

The first-best outcome according to these criteria is for all payments to be settled in the morning and for no bank to post any collateral.

We have now set out the basic building blocks of the model we use to analyse RTGS and how we assess different payment system designs. Each bank chooses one action in either period of the day. The actions are the sets of payments it wants to settle in each of the periods. From these decisions we back out the collateral that each bank must post to ensure that it has the necessary liquidity at all times in the day. We now proceed to describe equilibrium actions under RTGS.

⁽⁵⁾ We are grateful to a referee for pointing out the importance of interbank lending in ensuring that all banks can repay the CB at the end of the day.

3 Real-Time Gross Settlement

In this section we characterise equilibria under RTGS. We will show that any equilibrium will involve the posting of collateral and the delaying of some payments. We solve the game backwards and restrict attention to equilibria in pure strategies. The subgame that materialises in the afternoon depends on the history of play up to this point. Denote by h^2 the history of play where $h^2 = \{a_1^1, \dots, a_n^1\}$. Select a subgame and consider an individual bank's optimal afternoon action. Bank i has liquidity L_i^2 available for settlement purposes in the afternoon. As $\lambda > 0$, it is not optimal for any bank to choose its afternoon action a_i^2 to be such that it does not use up all of the liquidity provided by incoming payments that have arrived in the morning when doing so means that some payments are cancelled. It is also not optimal for a bank to choose an afternoon action such that payments are cancelled when the liquidity from incoming payments in the morning has a value less than that of the bank's remaining payments. This is because by assumption the benefit of cancelling one less payment λ exceeds the cost of obtaining the necessary liquidity from the CB, γ^2 . Thus, the optimal action is the set of remaining payments in the afternoon, $Q_i \setminus a_i^1$. This is clearly a bank's strictly dominant action in the afternoon. The result is stated formally in lemma 1.

Lemma 1: $a_i^2 = Q_i \setminus a_i^1$ strictly dominates all other $a_i^2 \in A_i^2$ for all i , given any h^2 .

Given that cancellation costs are sufficiently high we have arrived at the unsurprising result that no bank chooses to cancel any of its payments. This implies that every bank both sends and receives $n-1$ payments during the day. Substituting the values of outgoing and incoming payments into the left-hand side of the sustainability condition in equation (7) we see that the condition holds with equality. Thus, the payment system is sustainable and there is no interbank lending at the end of the day. A further implication of lemma 1 is that the amount of collateral that a bank posts in the afternoon is

$$c_i^2 = \max \left\{ 0, (n-1) - |a_i^1| - \sum_{j \neq i} I(j \rightarrow i \in a_j^1) \right\} \quad (8)$$

A bank's pay-off function in the morning when it chooses an action a_i^1 is shown in equation (9).

It is a bank's expected cost over the entire day from the perspective of the morning. Hence, a bank chooses its morning action in order to minimise expected cost. Let

$a_{-i}^1 = \{a_1^1, \dots, a_{i-1}^1, a_{i+1}^1, \dots, a_n^1\}$; ie the actions taken by all banks other than bank i in the morning.

The first part of expected cost is the delay cost incurred. The second is the morning collateral cost while the third is the expected afternoon collateral cost.⁽⁶⁾

$$c(a_i^1, a_{-i}^1) = \sum_{a_i^1 \supseteq a_i^1} \beta(a_i^1) + \gamma^1 |a_i^1| + \gamma^2 E_i^1 \left[\max \left\{ 0, (n-1) - |a_i^1| - \sum_{j \neq i} I(j \rightarrow i \in a_j^1) \right\} \right] \quad (9)$$

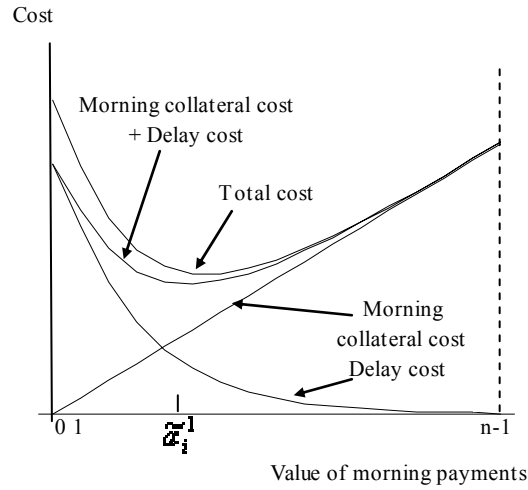
The expected value of afternoon collateral is shown in equation (10).

⁽⁶⁾ Obviously the only source of uncertainty is the value of incoming payments arriving in the morning.

$$E_i^1 \left[(n-1) - |a_i^1| - \sum_{j \neq i} I(j \rightarrow i \in a_j^1) \middle| (n-1) - |a_i^1| \geq \sum_{j \neq i} I(j \rightarrow i \in a_j^1) \right] \cdot \Pr \left((n-1) - |a_i^1| \geq \sum_{j \neq i} I(j \rightarrow i \in a_j^1) \right) \quad (10)$$

When a bank chooses to settle more payments in the morning there are three effects on its expected cost. First, delay cost declines. Second, morning collateral cost rises. Third, it reduces expected afternoon collateral cost. This is because an increase in $|a_i^1|$ reduces the conditional expectation and also reduces the probability that a bank has to post collateral in the afternoon; ie with fewer payments left to settle in the afternoon, it is more likely that incoming payments in the morning provide sufficient liquidity to settle these remaining payments. When deciding to settle fewer payments in the morning the three effects work in the opposite directions.

Chart 2



Delay cost: $\sum_{a_i^1 > a_i^1} \beta(a_i^1)$

Morning collateral cost: $\gamma^1 |a_i^1|$

Total cost: $c(a_i^1, a_{-i}^1)$

The costs faced by a bank are shown in Chart 2. On the horizontal axis is the value of payments settled in the morning, $|a_i^1|$. (One must remember that a given value is associated with a particular member of the morning action set A_i^1 .) The morning collateral cost equals $\gamma^1 |a_i^1|$. The delay cost is a decreasing function of $|a_i^1|$. The way that an internal queue is ordered, with a payment's position in the queue inversely related to its delay cost, generates the shape of the delay cost curve; ie the marginal change in delay cost as a bank settles more payments is declining in the value of payments settled. From these costs we derive a total cost excluding the expected afternoon collateral cost. The assumption that the reduction in the delay cost from settling the payment at the head of the internal queue compared with no payments, $\beta(\{1\})$, is

greater than γ^1 is clearly necessary and sufficient for the action minimising this cost to be non-empty. Hence, we have the ‘U-shaped’ curve illustrated in Chart 2. The cost including the expected afternoon collateral cost lies (weakly) above this curve for all $|a_i^1|$. Since the expected afternoon collateral cost is decreasing in $|a_i^1|$ the gap between the two curves is also decreasing in $|a_i^1|$. (The curves coincide when a bank settles all of its payments in the morning.) Thus, given the morning actions of other banks, the optimal morning action for a bank is that which minimises the total cost curve. We can refer to this action as the best reply of bank i to the actions of others, $BR_i(a_{-i}^1)$.

Before proceeding to discuss banks’ equilibrium action sets in each period, we introduce a mathematical tool that we use in the analysis. This is the covering relation \mapsto . (For more details see Davey and Priestley (2002). The covering relation is used to compare members of an ordered set. Consider an ordered set K , where elements in K are ordered according to an order \leq . Take two elements $x, y \in K$. We can write $x \mapsto y$ when $x > y$ and there exists no element z such that $x > z \geq y$, unless $z = y$.

Denote an equilibrium morning action by a_i^{1*} . (Lemma 1 implies that the equilibrium afternoon action is $Q_i \setminus a_i^{1*}$.) The equilibrium is characterised by

$$a_i^{1*} \in BR_i(a_{-i}^{1*}) \quad \forall i \tag{11}$$

We now describe the equilibrium under RTGS.

Proposition 1: Under RTGS, a positive value of payments will be settled in the morning and in the afternoon by all banks for any $n(> 2)$.

The proof of proposition 1 has three parts. The first demonstrates that no bank will choose to settle zero payments in the morning given the assumptions we have made about delay and collateral costs. Define an action \tilde{a}_i^1 . This is the action such that $\beta(\tilde{a}_i^1) \mapsto \gamma^1$.⁽⁷⁾ Clearly, \tilde{a}_i^1 is uniquely defined. It is depicted in Chart 2. For all actions to the left of it in A_i^1 (ie $|a_i^1| < |\tilde{a}_i^1|$), deviating from this action, by settling more payments, will always reduce a bank’s total cost regardless of the other banks’ actions. In other words, ignoring its impact on expected afternoon collateral cost, the act of settling additional payments is optimal for a bank. (Of course, given that expected afternoon collateral cost is declining in the number of payments settled, if a deviation is worthwhile without taking into account this cost reduction it will also be if we did consider it.) This is so for any actions where the reduction in delay cost exceeds the rise in morning collateral cost. That is, any action a_i^1 to the left of \tilde{a}_i^1 in A_i^1 . Hence, no bank will ever choose such an action. The assumption that $\beta(\{1\}) > \gamma^1$ implies that choosing \tilde{a}_i^1 entails the settlement of at least one payment by each bank in the morning.⁽⁸⁾

⁽⁷⁾ We define this relationship by applying the covering relation to the ordered set $\langle \beta(Q_i), \dots, \gamma^1, \dots, \beta(\{1\}); \leq \rangle$.

⁽⁸⁾ As all banks have the same cost structure, \tilde{a}_i^1 has the same value for all banks.

The second step is to show that it is not an equilibrium for all banks to settle all of their, respective, payments in the morning. Suppose banks choose the actions $\{Q_1, \dots, Q_n\}$ and consider the incentive of an individual bank i to deviate from its action Q_i . If any deviation strictly reduces i 's costs $\{Q_1, \dots, Q_n\}$ cannot be an equilibrium. As each other bank is settling its entire internal queue of payments in the morning, i is certain about the value of its incoming payments in the morning. It knows for sure that it will receive $n-1$ payments in the morning. This means that for any morning action it chooses, i will have sufficient liquidity in the afternoon to meet its needs without having to borrow from the CB for a second time. Settling one less payment in the morning generates a benefit γ^1 , the reduction in morning collateral cost. It also results in a cost $\beta(Q_i)$, which is the increase in delay cost. However, the expected afternoon collateral cost is unchanged. The assumption that $\beta(Q_i) \leq \gamma^1 - \gamma^2$ means that this deviation strictly reduces the costs of bank i . Therefore, $\{Q_1, \dots, Q_n\}$ cannot be an equilibrium. The intuition behind this is that when all other banks settle all of their payments in the morning, a bank optimally shifts settlement activity towards the afternoon so as to exploit the high (and certain) value of incoming payments in the morning given that the cost of delaying some payments is not too high.

The final step shows that no bank will choose to settle all of its payments in the morning. Suppose banks choose the actions $\{Q_1, \dots, Q_{j-1}, a_j^1, Q_{j+1}, \dots, Q_n\}$, where $a_j^1 \subset Q_j$. Consider the bank i 's incentive to deviate from its action Q_i . If it settles one less payment it incurs an increase in its delay cost $\beta(Q_i)$ but enjoys the benefit of reduced morning collateral cost, γ^1 . Its afternoon collateral cost remains equal to zero because it receives at least one payment in the morning with certainty because other banks settle all of their payments in the morning and each bank has one payment to send to i . The deviation strictly reduces bank i 's costs because $\gamma^1 > \beta(Q_i)$. Now suppose that two banks do not settle all of their payments in the morning while all other banks do. Again bank i has an incentive to deviate from its action Q_i because by doing so it incurs an increase in delay cost of $\beta(Q_i)$, a reduction in morning collateral cost γ^1 , and no change in its expected afternoon collateral cost since it receives one payment in the morning with certainty because $n-3$ banks are settling all of their payments in the morning. The same argument ensures that it can be used to show that there is no equilibrium such that more than one bank settles all of its payments in the morning.

Suppose that banks choose the actions $\{a_1^1, \dots, a_{i-1}^1, Q_i, a_{i+1}^1, \dots, a_n^1\}$ where $a_j^1 \subset Q_j$ for all $j \neq i$. If bank i settles one less payment it incurs an increase in cost $\beta(Q_i)$ and a decrease, γ^1 . It also experiences an increase in its expected afternoon collateral cost because, with each other bank settling fewer than all of its payments, bank i cannot be certain that it will receive a payment in the morning. The increase in its expected afternoon collateral cost can be represented by $p\gamma^2$ where $p(<1)$ is the probability that i has to post one unit of collateral in the afternoon. But the deviation strictly reduces bank i 's costs because $\gamma^1 - \gamma^2 \geq \beta(Q_i)$ implies that $\gamma^1 - p\gamma^2 > \beta(Q_i)$.

In equilibrium, under RTGS, payments are settled in both periods. Thus, the system will be exposed to operational risk in the afternoon. The fact that payments are settled in the morning implies that collateral posted in the morning has a strictly positive value. In the afternoon, collateral posted can also be strictly positive or equal to zero. The basic structure of RTGS

precludes the first-best being attainable. To speed up settlement requires banks to post more collateral in the morning (with a less than fully compensating reduction in expected afternoon collateral cost). To reduce the amount of collateral posted in the morning, the system's exposure to operational risk has to rise. It is this trade-off that implies that the first-best is simply not possible.

4 Hybrid payment systems

As we said in the introduction, the debate about the relative merits of different payment systems used to be centred on two opposing classes of system designs: RTGS and DNS. A preference for one of the two depends on the perceived optimal mix of credit risk and liquidity efficiency. DNS requires low levels of liquidity but generates credit risk. RTGS needs more liquidity but there is no risk. However, there now exists a set of alternatives to RTGS and DNS known as hybrid payment systems. As the name suggests, a hybrid incorporates elements of both RTGS and DNS. But to be more precise, a hybrid envelops one of RTGS and DNS but augments this design with rules and procedures one would normally associate with the other. The argument in favour of hybrid designs is that they will implement levels of social welfare that are not attainable with either RTGS or DNS.

As we have taken RTGS to be our benchmark payment system design, we consider hybrid designs that are based on RTGS rather than DNS. In particular, we augment the standard features with a payment offset facility.

A payment is deemed to be offset when it is part of a set of payments that are simultaneously settled. We say that there is bilateral offset when a set of offset payments includes payments from only two banks. There is multilateral offset when a set of offset payments can include payments from more than two banks. Settlement is on a gross basis in a legal sense. But in terms of the value of liquidity banks need to settle these payments, offset payments are in effect netted. This is because payments are self-collateralising (McAndrews and Trundle (2001)). Thus, overall liquidity demands are lower than in RTGS. Of course, offsetting need not be perfect, with incoming payments in the set being sufficient to allow each bank to settle its outgoing payments in the set without further liquidity. That is, offset can generate residual payments unless further liquidity is provided. The system operator will typically put a ceiling on the value of such residual payments. This settlement procedure has obvious parallels with net settlement. However, payment offset need not generate the risk we would associate with DNS. This is because of the absence of settlement lags. The lack of lags arises from the fact that settlement is simultaneous and that no offset payment is settled until all of the residual payments have been settled (either by RTGS or subsequent offsetting opportunities).

We model bilateral rather than multilateral offset since in this model, with each bank having a single payment to make to each other bank, it is capable of resulting in the theoretical minimum of liquidity demands, zero. (Although we will return to the issue of multilateral offset below.)

4.1 *Modifications to the model*

To be able to analyse hybrid payment systems, and changes to behaviour that they could bring about, we must modify the model in several key areas.

Payments can now be settled through two streams. There is a RTGS stream that operates in exactly the same way as before. This is consistent with the basic idea that a hybrid system design should always fully encompass either RTGS or DNS. The second stream is the central queue in which payments are placed to await offsetting payments. We assume that there is no direct cost of submitting a payment into the central queue. When two offsetting payments are both in the central queue at the same time they can be offset and settled instantly. There is no lag between the identification of offsetting payments and settlement because all payments have the same value; ie there are no residual payments to be settled. No payment can reside in both streams at the same time. So if a payment is placed in the central queue and the offsetting payment is not there, to achieve quick settlement a bank has to withdraw the payment and put it in the RTGS stream in the next period.

The second major change is that we expand the number of times a bank acts. We now have a four-stage game. Each period is split into two subperiods. Denote the first subperiod in period $1(2)$ with $1a(2a)$ and the second with $1b(2b)$.⁽⁹⁾ We assume that there is no delay cost associated with a payment settling in subperiod tb rather than subperiod ta . (The subject of whether there is a cost to using the central queue is returned to below.) We also assume that there is no risk of an operational event happening between subperiods.

The hybrid designs considered in this paper are defined along two dimensions. The first is the timing of payment offset. If offset occurs in either period it will take place in both subperiods in this period. We consider three cases:

H1: offset in the afternoon only

H2: offset in the morning only

H3: offset in the morning and the afternoon.

The other dimension is the transparency of the central queue. We consider two different degrees of transparency:

Opaque: a bank can only see its own payments in the central queue.

Visible: a bank can see its own payments and payments intended for it in the central queue.

In any period that the offset facility is not operating a bank's action set is just the set of payments that are settled in the RTGS stream. In H1, the morning action set is that in equation (1). In H2, the afternoon action set is the power set of all remaining payments, given morning actions. The

⁽⁹⁾ The results for RTGS that we found above also apply in a four-stage version of the game. In the two-stage game, the only factor that is causing banks to settle payments in the morning rather than the afternoon is the delay cost. By the same argument, in the four-stage game, given that payments will be settled in each period, no bank chooses to settle payments in subperiods $1a$ and $2a$. Delaying settlement to the second subperiod generates no cost for a bank but does open up an opportunity to exploit incoming payments that arrive in the first subperiod. Of course, if all banks behave in this way, there will be no incoming payments in the first subperiods of the morning and the afternoon. Although, when one bank settles in ta while all others settle in tb could be an equilibrium. But this can be removed if one assumes that if a bank is indifferent between settling in each subperiod (as it would be in this alternative equilibrium) it chooses to settle in the second subperiod.

set of remaining payments is the internal queue minus the payments settled by RTGS in the morning and those payments that are offset in the morning.

When the offset facility is working, banks take a number of actions. First, there are the sets of payments placed in the central queue in each subperiod tk , $q_{i,c}^{tk}$, $t=1,2$, $k=a,b$. Second, banks choose the sets of payments to settle by RTGS in each subperiod, a_i^{tk} . Assume that $q_{i,c}^{tk} \cap a_i^{tk} = \emptyset$; ie no payment can be submitted to the central queue and be settled by RTGS in the same subperiod. In subperiod $1a$ we assume that a bank shows a preference for submitting more time-critical payments to the central queue: $q_{i,c}^{1a} \in A_i^1$, where A_i^1 is the same action set as that in equation (1). The set of payments settled by RTGS in subperiod $1a$ is a member of $A_i^1 \setminus q_{i,c}^{1a}$. In subperiod $2a$, $q_{i,c}^{2a}$ is taken from the set of payments that can be offset. If a bank submits ϕ payments to the central queue it is equally likely to choose any of the payments that can be offset.

4.2 Equilibria under hybrid payment systems

4.2.1 H1: Payment offset in the afternoon

In this section we consider and discuss important results for each of the hybrid payment system designs. We begin with H1 when the central queue is opaque. The morning action is the same as under RTGS, $a_i^1 \in A_i^1$. The afternoon actions are $\{a_i^{2a}, q_{i,c}^{2a}, a_i^{2b}, q_{i,s}^{2b}\}$. If a payment remains unsettled in subperiod $2b$ and the offsetting payment has been received through the RTGS stream, it is optimal for a bank to settle this payment by RTGS in $2b$. This is shown in lemma 2. It follows from lemma 1.

Lemma 2: If $j \rightarrow i \in a_j^1$ or $j \rightarrow i \in a_j^{2a}$ and $i \rightarrow j \notin a_i^1, a_i^{2a}$, then $i \rightarrow j \in a_i^{2b}$.

If a payment remains unsettled in subperiod $2a$ and the offsetting payment has not yet been received, there are a number of possible outcomes. A bank might submit such payments to the central queue if it believed that other banks were going to do the same. If it believed other banks were not going to, it would choose to settle such payments by RTGS in subperiod $2b$. From the perspective of subperiod $2b$, there is no unique outcome.

If a bank has not settled a payment in the morning, but the offsetting payment arrived in the morning, the bank will choose to delay settling the payment in the RTGS stream until subperiod $2b$. The delay between subperiods entails no cost but allows the bank to use any RTGS payments it receives in subperiod $2a$ to make payments in $2b$. This result is shown in lemma 3.

Lemma 3: $a_i^{2a} = \emptyset$ for all i .

If a bank has not settled a payment in the morning and the offsetting payment was not made in the morning, it is optimal for a bank to submit this payment to the central queue in subperiod $2a$ if each other bank submits at least one payment. The bank enjoys a liquidity saving if the payment is offset but no cost if the payment is not offset in $2a$ and is settled in $2b$. If each other bank submits at least one payment, each of a bank's payments will be offset in $2a$ with a positive

probability. But if all payments that can be offset are submitted to the central queue in 2a, all of these payments will be offset. Thus, in this equilibrium the central queue will be empty in subperiod 2b.

Lemma 4: $q_{i,c}^{2b} = \emptyset$ for all i .

The other possible outcome in period 2a is that banks submit no payments to the central queue. Proceeding to subperiod 2b, there is no unique outcome (for reasons we discussed above). We will ignore equilibria in which banks choose not to use the central queue at the earliest opportunity in the analysis below.

A bank will settle a payment by RTGS in subperiod 2b if and only if it is not settled in the morning. If a bank does not settle a payment in the morning, it will be settled by RTGS in the afternoon if and only if the offsetting payment was settled in the morning. If the offsetting payment was not settled in the morning, then both payments are submitted to the central queue in subperiod 2a and are offset. Therefore,

$$I(i \rightarrow j \in a_i^{2b}) = 1 - I(i \rightarrow j \in a_i^1) - [1 - I(i \rightarrow j \in a_i^1)][1 - I(j \rightarrow i \in a_j^1)]$$

Or

$$I(i \rightarrow j \in a_i^{2b}) = I(j \rightarrow i \in a_j^1)[1 - I(i \rightarrow j \in a_i^1)] \quad (12)$$

The value of the payments a bank settles in the RTGS stream in the afternoon is

$$\sum_{j \neq i} I(j \rightarrow i \in a_j^1)[1 - I(i \rightarrow j \in a_i^1)] \quad (13)$$

That is, the more payments that are settled in the morning, the fewer payments are settled by RTGS in the afternoon. The afternoon collateral condition is

$$c_i^2 + \sum_{j \neq i} I(j \rightarrow i \in a_j^1) \geq \sum_{j \neq i} I(j \rightarrow i \in a_j^1)[1 - I(i \rightarrow j \in a_i^1)]$$

The level of collateral that satisfies the afternoon collateral condition is

$$c_i^2 = \max \left\{ 0, \sum_{j \neq i} I(i \rightarrow j \in a_i^1)(j \rightarrow i \in a_j^1) - \sum_{j \neq i} I(i \rightarrow j \in a_i^1) - \sum_{j \neq i} I(j \rightarrow i \in a_j^1) \right\} \quad (14)$$

Or

$$c_i^2 = 0$$

A bank will know with certainty that, for any morning action and value of incoming payments, it does not post collateral in the afternoon. This is because if a payment is offset in the afternoon, because it and its offsetting payment were not settled in the morning, it requires no collateral while if a payment is settled in the RTGS stream in the afternoon it must be that the offsetting

payment was settled in the morning. Hence, the necessary liquidity is provided by this incoming payment.

The implication of this is that banks optimisation problems are disentangled from each others actions. Its cost function is just the first two elements of equation (9). In Chart 2 total costs under H1 correspond to total costs under RTGS minus expected afternoon collateral cost. Since the expected afternoon collateral cost is declining in the value of payments settled in the morning under RTGS, the optimal morning action under H1 is either equal to that under RTGS or entails fewer payments being settled. Because all banks have the same morning collateral cost and delay cost functions, each bank chooses to settle the same value of payments in the morning, $|\tilde{a}_i^1|$.

The intuition behind this result is that the opportunity for payments to be offset in the afternoon tempts banks to delay a greater number of payments than they do under RTGS. The implication for welfare is not clear. If the same number of payments are settled in the morning under RTGS and H1, the total value of collateral posted is (weakly) greater under RTGS. Exposure to operational risk is the same under both designs. Therefore, H1 is (weakly) preferred. When fewer payments are settled in the morning under H1 than under RTGS, which design is preferred depends on the relative weights placed on each criterion. RTGS has a greater speed of settlement but H1 involves less collateral being posted.

As under RTGS, there is an inherent trade-off between liquidity demands and speed of settlement under H1. The more payments are settled in the morning, the more collateral has to be posted in the morning. The fewer payments settled in the morning, the greater the exposure of the system to operational risk. As before, the implication of the trade-off is that the first-best is not attainable when there is payment offset in the afternoon only and the central queue is opaque.

Let the central queue be visible. In subperiod $2b$, a bank faces the same situation as it does when the central queue is opaque. Although a bank can observe whether offsetting payments were in the central queue in subperiod $2a$, it knows that posting a payment to the central queue in $2a$ is not a binding commitment on the part of a bank to keep the payment in the central queue in $2b$ if it is not offset immediately. Whether a bank keeps a payment in the central queue in $2b$ depends only on what it expects the other bank to do. As before, there is no unique outcome. In subperiod $2a$, a bank will find it optimal to submit any payment that could be offset to the central queue for the same reasons as outlined above. Lemma 3 and 4 also hold when there central queue is visible. Thus, the same equilibrium arises as that when the central queue is opaque.

H1 is sustainable in equilibrium. A bank posts collateral only in the morning. It is such that the morning collateral condition binds; ie

$$c_i^1 = \sum_{j \neq i} I(i \rightarrow j \in a_i^1) \quad (15)$$

For a bank to repay the CB and redeem its collateral at the end of the day, its total incoming payments, through the RTGS stream, minus the payments it makes by RTGS in subperiod $2b$, must be at least equal to c_i^1 . The condition is shown in equation (16).

$$\sum_{j \neq i} I(j \rightarrow i \in a_j^1) + \sum_{j \neq i} I(j \rightarrow i \in a_j^{2b}) - \sum_{j \neq i} I(i \rightarrow j \in a_i^{2b}) \geq \sum_{j \neq i} I(i \rightarrow j \in a_i^1) \quad (16)$$

If we substitute equation (12) into equation (16) we find that the payment system is sustainable if

$$\sum_{j \neq i} I(j \rightarrow i \in a_j^1) I(i \rightarrow j \in a_i^1) \geq \sum_{j \neq i} I(j \rightarrow i \in a_j^1) I(i \rightarrow j \in a_i^1) \quad (16b)$$

H1 is sustainable because equation (16b) holds with equality for all banks.

Proposition 2: The same equilibrium arises under H1 when the central queue is opaque and when it is visible. H1 is sustainable.

4.2.2 H2: Payment offset in the morning

We now turn to H2. Lemma 1 confirms that a bank will settle all of its remaining payments by RTGS in the afternoon. In subperiod 1b, banks face a similar situation as they do in 2b under H1. But in subperiod 1a, a bank will choose to submit all of its payments to the central queue if each other bank submits at least one payment, when the central queue is opaque or visible. (There are equilibria where banks choose to not submit any payments to the central queue in subperiod 1a. We will ignore these as we did under H1.) This is because by doing so a bank maximises the likelihood that its payments are offset, while placing a payment in the central queue involves no cost either in the form of an explicit charge or a restriction on what a bank can do with the payment in the future. As each bank has a single payment to send to each other bank, and that all payments share a common value, all payments are offset in subperiod 1a. This implies that no bank posts collateral in either period. The equilibrium corresponds to the first-best; ie liquidity demands and exposure to operational risk are both minimised.

Proposition 3: The first-best is attained under H2 when the central queue is opaque and when it is visible.

Because in equilibrium no bank posts any collateral, the sustainability condition is satisfied.

4.2.3 H3: Payment offset in the morning and afternoon

H3 produces the same equilibrium as under H2. In subperiod 1a, a bank will choose to submit its entire internal queue of payments to the central queue, irrespective of whether the central queue is opaque or visible. (Ignoring possible equilibria where no payments are submitted to the central queue in subperiod 1a.) All payments are offset. The system is sustainable.

Proposition 4: The first-best is attained under H3 when the central queue is opaque and when it is visible.

4.3 Information asymmetry

A key assumption in the model is that the ordering of internal queues is private information. Changing this assumption could have an impact on the results and hence, the optimal payment system design.

4.3.1 Public information

In Section 4 we argued that in a subperiod in which payments could be offset, banks either submit no payments (the case we ignored) or all payments that could be offset. No intermediate cases can arise in equilibrium because a bank knows that each of its payments that can be offset will be offset with positive probability if each other bank submits at least one payment. A bank is uncertain about the identity of payments submitted by other banks in subperiod $2a$ because each other bank is equally likely to submit any of its payments that can be offset. A bank is uncertain in the morning because the orderings of internal queues is private information.

In this section we assume that internal queue orderings are public information and see how the results for H2 and H3 could be affected with the aid of an example. We show that intermediate cases are possible in equilibrium. Let $n=3$. The internal queues of the three banks are respectively,

$$Q_1 = \{1 \rightarrow 2, 1 \rightarrow 3\}$$

$$Q_2 = \{2 \rightarrow 3, 2 \rightarrow 1\}$$

$$Q_3 = \{3 \rightarrow 2, 3 \rightarrow 1\}$$

As before, $\beta(\{1\}) > \gamma^1$ and $\beta(Q_i) \leq \gamma^1 - \gamma^2$. Suppose that the sets of payments that banks forward to the central queue in $1a$ are

$$q_{1,c}^{1a} = \{1 \rightarrow 2\}$$

$$q_{2,c}^{1a} = \{2 \rightarrow 3, 2 \rightarrow 1\}$$

$$q_{3,c}^{1a} = \{3 \rightarrow 2\}$$

In the morning, all of the payments placed in the central queue are offset and no payments are settled in the RTGS stream in subperiod $1b$. Under H2, the levels of expected afternoon collateral are $c_1^2 = 1$, $c_2^2 = 0$, and $c_3^2 = 1$. With H3, all remaining payments are offset in the afternoon and no collateral has to be posted by any bank.

The above does not represent an equilibrium if internal queues are private information. When internal queues are private information, a bank submits all of its payments to the central queue in $1a$ since each could be offset with a positive probability.

When the ordering of internal queues is public information bank 2 has no incentive to deviate from the actions described above since doing so will (strictly) increase its costs. Banks 1 and 3 also do not have an incentive to deviate from the above set of actions under H2 and H3. Consider bank 1. (As 1 and 3 have the same cost structure analysing the incentives of just one of them will suffice.) Under H2, if 1 forwards the payment $1 \rightarrow 3$ into the central queue in $1a$ but is does not settle it by RTGS in $1b$, the benefit and cost of deviating are both zero. There is no benefit and no cost, since the payment is not settled in the morning. If 1 does settle $1 \rightarrow 3$ by RTGS in the morning, the benefit is $\beta(Q_i) + \gamma^2$; ie the reduction in total delay costs plus the expected fall in

afternoon collateral costs. The cost is the increase in morning collateral costs, γ^1 . As it is assumed that $\beta(Q_i) \leq \gamma^1 - \gamma^2$, 1 does not strictly gain from this deviation.

Under H3, if 1 deviates by placing 1→3 in the central queue but does not settle it by RTGS in 1*b*, there is no benefit and no cost. The payment is not offset or settled by RTGS in the morning but is offset in the afternoon, just as before the deviation. When 1→3 is settled by RTGS, there is a benefit $\beta(Q_i)$ but a cost γ^1 . The benefit arises from the fact that the payment is settled in the morning rather than the afternoon. The cost is due to the payment being settled by RTGS rather than being offset in the afternoon. The deviation is not optimal as $\gamma^1 > \beta(Q_i)$.

$q_{i,c}^{1a} = Q_i$ for all i is also an equilibrium when internal queues are publicly known. As with bank 2 in the above case, none of the banks have an incentive to deviate. Therefore, we have multiple equilibria, ignoring those equilibria where banks fail to submit any payments to the central queue in 1*a*. The intermediate cases we discussed above can arise in equilibrium. The equilibria can be Pareto-ranked with the one equilibrium corresponding to the first-best. So, if internal queues are public information the welfare gains associated with H2 and H3 could be overstated. But there could be policy remedies that ensure that ‘intermediate’ equilibria cannot exist. One policy is to prohibit the withdrawal of payments once they have been placed in a visible central queue. This enables banks to commit to not withdrawing a payment from the central queue even if it is not immediately offset. A bank which observes an incoming payment in the central queue in 1*a* will naturally respond by forwarding the offsetting payment to the central queue in 1*b* since it is sure that the payments will be offset. Working backwards, each bank has an incentive to place all of its payments in the central queue in 1*a*.⁽¹⁰⁾

4.3.2 Additional uncertainty

The result contained in propositions 3 and 4 that the same equilibrium arise for both opaque and visible central queues is perhaps unexpected and leads one to ask whether there is something about the form of information asymmetry in the model that generates the result? The answer is that the result appears to be robust as long as one maintains a certain other assumption made above. The key assumption is that submitting a payment to the central queue costs a bank nothing.

One can think of a bank potentially facing a number of layers of uncertainty. The first, which we consider above, is where the overall pattern of payments and banks’ costs are universally known but the time-criticality of payments is the private information of the banks sending them. In such a climate, a bank does not know with certainty how other banks’ actions map into its own pay-off. There could be additional uncertainty. A bank could be uncertain about the existence of incoming payments. The situation would be complicated further if banks were to receive payment requests at several times in the day; ie does a payment $i \rightarrow j$ not exist or is it that bank i is yet to have received the order to send it? On top of this layer of uncertainty, a bank could be uncertain about its own as well as other banks’ payments.

⁽¹⁰⁾ In a repeated game, banks could achieve more co-ordinated use of the central queue without the need for a policy remedy such as prohibited payment withdrawal.

The argument is that with any type of uncertainty, a bank will still forward all of its payments to the central queue at the earliest opportunity. By doing so it ensures that any payment for which an offsetting payment exists is offset.

However, if banks incur a cost when using the central queue, different kinds of uncertainty may lead to different equilibria.

Suppose that submitting a payment to the central queue is costly for a bank.⁽¹¹⁾ Assume that this charge is less than the collateral saving if the payment is subsequently offset in the same period. This ensures that if a bank is sufficiently confident that a payment it places in the central queue will be offset in the same period it will use the central queue. Introducing further uncertainty reduces the belief that offset will happen. If the perceived probability that it is possible is low enough a bank could be persuaded not to bother using the central queue but instead settling payments by RTGS immediately. However, it is likely that collateral costs are such that a bank will not settle all of its payments by RTGS in subperiod $1a$. In which case, the way that beliefs are updated after subperiod $1a$ becomes relevant to bank actions in subsequent subperiods. We can then discern how opaqueness and visibility could generate different equilibria.

4.4 Multilateral offset

Under bilateral offset, it is possible for all payments to be offset. Indeed, we find that this is the case under H2 and H3 in equilibrium. If there was multilateral offset, the results should be no different. Under H1, multilateral offset could allow for more payments to be offset in the afternoon.

Suppose that under H1 multilateral offset is available in the afternoon. Whenever payments can be multilaterally offset they can also be bilaterally offset. The set of payments that is offset is naturally greater when there is multilateral offset than when there is bilateral offset. Yet, the equilibrium under H1 will be just the same as before. No bank needs to post collateral in the afternoon and each knows this with certainty. So each bank's actions are determined by morning collateral cost and delay cost, which are independent of the actions of other banks.

5 Conclusions

In this paper we have considered the properties of RTGS and hybrid payment systems. We have compared different designs with reference to two criteria: liquidity demands and the speed of settlement, which captures the exposure of a system to exogenous operational risk. In the model, the time-criticality of a payment is the private information of the bank sending it. Under RTGS, the first-best is unattainable. This is because there is an inescapable trade-off between liquidity needs in the morning and speed of settlement. As in previous work considering bank behaviour under RTGS, we find that some payments will be delayed and then settled using liquidity

⁽¹¹⁾ One such cost arises from the fact that payments in the central queue will not necessarily be settled instantly on arrival in the central queue. Hence, leaving a payment in the central queue will, in practice, impose a delay cost. There may also be operational risk associated with a large number of payments residing in the central queue at any one time.

provided by incoming payments. The key to obtaining welfare improvements is to escape this trade-off. With this in mind we consider a series of hybrid payment system designs, each augmenting RTGS with payment offset. Varying the timing of payment offset we find that improvements in terms of exposure to operational risk are not possible if payments can be offset only in the afternoon. This implies that the first-best is not attainable under the hybrid payment system design H1. However, the first-best is attained when payments can be offset in the morning only or all day. What is more, none of the results applying to hybrids depend on the transparency of the central queue. Relaxing the assumption that the orderings of internal queues are private information can generate multiple equilibria. Assuming that there is additional uncertainty uncovers the important role that costs to using the central queue could play.

References

Angelini, P (1998), ‘An analysis of competitive externalities in gross settlement systems’, *Journal of Banking and Finance*, Vol. 22, pages 1-18.

Bank for International Settlements (1997), ‘Real-time gross settlement systems’, *CPSS Publications*, No. 22.

Bech, M L and Garratt, R (2003), ‘The intraday liquidity management game’, *Journal of Economic Theory*, Vol. 109, pages 198-219.

Buckle, S and Campbell, E (2003), ‘Settlement bank behaviour in a real-time gross settlement payment system’, *Bank of England Working Paper no. 209*.

Davey, B A and Priestley, H A (2002), *Introduction to lattices and order*, Cambridge University Press, Second edition.

Johnson, K, McAndrews, J and Soramäki, K (2002), ‘Complements to real-time gross settlement systems: economizing liquidity with queuing and netting’, *mimeo*.

Kobayakawa, S (1997), ‘The comparative analysis of settlement systems’, *Centre for Economic Policy Research Discussion Paper no. 1667*.

Leinonen, H and Soramäki, K (1999), ‘Optimizing liquidity usage and settlement speed in payment systems’, *Bank of Finland Discussion Paper no. 16*.

McAndrews, J and Trundle, J (2001), ‘New payment system design: causes and consequences’, *Bank of England Financial Stability Review*, December, pages 127-36.

Roberds, W (1993), ‘The rise of electronic payments networks and the future of the Fed with regard to payment finality’, *Federal Reserve Bank of Atlanta Economic Review*, March/April, pages 1-22.

Roberds, W (1999), ‘The incentives effects of settlement systems: a comparison of gross settlement, net settlement, and gross settlement with queuing’, *Institute For Monetary And Economic Studies – Bank of Japan Discussion Paper*, No. 99-E-25.