Vlastimir Vuković*

# RTGS: The way of backward**

CBM *Research* Paper 11, May 2025


## *Abstract*

*The net settlement model dominated operations in clearing houses for over two centuries, up until the early 1990s. At that period, central bankers began favoring* ***explicit*** *transactions with explicit money at the account, i.e., gross settlement (Leigh-Pemberton). The same central bankers proposed sound banking principles: '[f]unds must be in an account at a central bank before payment transfers will be honored' (Angell). This became the core idea behind the construction of RTGS systems. As a result, all participants were effectively pushed into a waiting room.*

*Despite the dominance of RTGS systems in wholesale payments, the practice of gross settlement could not eliminate the interdependence between incoming and outgoing transactions, nor the restrained use of reserves. Therefore, RTGS systems have since been upgraded by introducing netting schemes to bridge illiquidity. All available evidence points to the conclusion that RTGS systems are not superior, but rather inherently inferior to net systems.*

*The gross settlement, in a transaction-by-transaction model, has intensified competitive behavior and blurred the conditionality of own interest by common interest. Therefore, the spirit of cooperation and mutual trust among participants is necessarily stronger in the CHIPS network than in Fedwire, or in CHAPS than in the BoE's RTGS system.*

*A high level of reserves in the banking sector does not imply timely payments. On the contrary, in stress scenarios, bankers instinctively engage in hoarding of liquidity. A potential solution could lie in Real-Time Net Settlement (RTNS), based on accelerated sequential netting. The proposed RTNS model can be succinctly defined as a real-time net settlement payment system with accelerated sequential multilateral netting/clearing in predefined cycles on a final offsetting basis, along with the use of prefunded balances and intraday credits.*


***Key words:*** *RTGS, wholesale payments, clearing, settlement, banks, central banks, reserves, system liquidity, payment intermediaries (PSPs), CHIPS, Fedwire, CHAPS, accelerated sequential netting, RTNS solution, prefunded balances, intraday credits, mediated money.****

---

* Central Bank Money *Research*.

** This paper is the third part of the Discussion paper No. 5 - '*Immediate money and mediated money: Two primary forms of money and their concealed differences*', Central Bank Money *Research*, November 2024.

*** **Mediated money** exclusively circulates account-to-account in books of payment institutions. It implies that money stays in-network, i.e. in banking system (central bank and commercial banks).

## 1 Introduction

The net settlement model dominated operations in clearing houses for over two centuries, up until the early 1990s. At that period, central bankers began favoring ***explicit*** *transactions* with explicit money at the account, i.e., gross settlement (Leigh-Pemberton). Since then, the stigma of implicitness has followed the net settlement model, although it has managed to survive in the shadow of RTGS as the most efficient method to *reduce settlement risk.*

The same central bankers proposed sound banking principles: '[f]unds must be in an account at a central bank before payment transfers will be honored' (Angell, p. 7). This became the core idea behind the construction of RTGS systems. As a result, all participants were effectively pushed into a waiting room. In order to initiate the chain of payments, banks must wait for incoming payments from other banks, or borrow from the central bank or on the money market. This is a fundamental design flaw of RTGS payment systems

Despite the dominance of RTGS systems in wholesale payments, the practice of gross settlement could not eliminate the interdependence between incoming and outgoing transactions, nor the restrained use of reserves. Therefore, RTGS systems have since been upgraded by introducing netting schemes to bridge illiquidity.

All available evidence points to the conclusion that RTGS systems are not superior, but rather inherently inferior to net systems. Nevertheless, gross settlement systems have become dominant; their expansion during the 1990s was driven by the need for safer foreign exchange transactions (PvP mechanism) and securities trading (DvP mechanism). Yet, this trinity of gross settlement mechanisms neither prevented nor mitigated the GFC 2007–9.

The gross settlement, in a transaction-by-transaction model, has intensified competitive behavior and blurred the conditionality of own interest by common interest. Since the early 1990s, the payment game has become increasingly noncooperative and causes de-networking in interbank payments. Therefore, the spirit of cooperation and mutual trust among participants is necessarily stronger in the CHIPS network than in Fedwire, or in CHAPS than in the BoE's RTGS system.

A high level of reserves in the banking sector does not imply timely payments. On the contrary, in stress scenarios, bankers instinctively engage in hoarding of liquidity. The pattern of a liquidity crisis: Distress => Uncertainty => Untrust => FEAR => Hoarding => Panic => Sale => Freezing. Once hoarding of reserves begins, the connections between incoming and outgoing payments. In the panic phase, comprehensive de-networking threatens the liquidity of all banks, even though they attempt to isolate themselves. However, bank illiquidity is probably the only contagion that is worsened by isolation.

A potential solution could lie in Real-Time Net Settlement (RTNS), based on accelerated sequential netting. The proposed RTNS model can be succinctly defined as a real-time net settlement payment system with accelerated sequential multilateral netting/clearing in predefined cycles (from a minimum of two minutes to a maximum of fifteen minutes) on a final offsetting basis, along with the use of prefunded balances and intraday credits.

The fundamental rule is that each **payment message** = **payment order**. At the end of the cycle, the system operator calculates interbank obligations from payment messages through multilateral netting. *Offset payments* are immediately entered into system records and bank books as final, irrevocable, and unconditional, without any deferment. Remaining non-settled payment messages are placed in queuing according to the system's rules and instructions from the payer bank.

The paper after this Introduction (1) is organised as follows: RTGS: the way backward (2), RTNS solution: accelerated sequential netting (3), and Conclusion (4).

## 2 RTGS: the way backward

The net settlement model dominated operations in clearing houses for over two centuries, up until the early 1990s. At that period, central bankers began favoring ***explicit*** *transactions* with explicit money at the account, i.e., gross settlement (Leigh-Pemberton, p. 454). 'I am quite clear, therefore, that real-time gross settlement is the way forward, [...]' (ibid, p. 452). Since then, the stigma of implicitness has followed the net settlement model, although it has managed to survive in the shadow of RTGS as the most efficient method to *reduce settlement risk*.[1]

Among central bankers, the belief took hold that they had discovered a *quicker* and *more efficient* system. 'In this case, large-value payment systems would be able to run more quickly and efficiently on the basis of sound banking principles: funds must be in an account at a central bank before payment transfers will be honored' (Angell, p. 7). This became the core idea behind the construction of RTGS systems. Therefore, in the RTGS model, '[…] offsetting payment obligations cannot be netted and must instead be pre-funded. […] Banks do not have to fund all of their payments directly, however. Liquidity is recycled throughout the day as banks use incoming payments to fund outgoing ones' (Benos et al, p. 144–145). The cited sound banking principles represent a rejection of the clearing house concept and its practice. Of course, netting could not disappear entirely—it was merely suppressed (via incoming payments) and made more difficult (by requiring funds at account).

As a result, all participants were effectively pushed into a waiting room. A simple example: two banks hold $10 million in mutual payments in the morning, both on their own behalf and on behalf of their clients. However, the smallest individual payment is $800,000, and neither bank has that amount in its account at the central bank at the beginning of the day. The banks are formally illiquid, despite holding sufficient settlement assets to meet their total obligations, and they face congestion.[2] Each bank is required to send the full amount of money to the other bank at account. From this metallist's view, does it follow that a clearing instruction is not a payment order, but rather a worthless message? In order to initiate the chain of payments, banks must wait for incoming payments from other banks, or borrow from the central bank or on the money market.

This simple example highlights a fundamental design flaw of RTGS payment systems. 'The fundamental problem of a payment system is whether promised payments will actually occur' (Kahn et al, p. 3). Hence, the gridlock of payment flows is the chronic illness of RTGS systems.

This illness was not chronic in the clearing houses of previous centuries: '[a] key advantage of net settlement over RTGS, which is that net settlement can eliminate gridlock equilibria, [...]. Net settlement is effective because it *de facto* gives absolute priority to offsetting claims, since such claims are automatically discharged under net settlement rules' (ibid, p. 17).

Notions such as gridlock, congestion, delays, cascades, shortfalls, deadlocks, queuing, hoarding, and many others are frequently found in titles of research papers on RTGS systems, describing various functional problems. The primary cause lies in the *metallist concept of explicit transactions*, or so-called gross settlement, which operates on a transaction-by-transaction basis. 'In a gross settlement system, on the other hand, the settlement of funds occurs on a transaction-by-transaction basis, that is, without netting debits against credits' (BIS, 1997, p. 5). In short, it functions without clearing, on a purse-to-purse model.

Despite the dominance of RTGS systems in wholesale payments, the practice of gross settlement could not eliminate the interdependence between incoming and outgoing transactions, nor the restrained use of reserves. '[W]e find evidence that banks in the U.S. still economize in the use of intraday liquidity and rely on incoming payments to make outgoing payments, generating a high degree of strategic complementarity in payment decisions' (Afonso et al., 2022, p. 15). The persistence of clearing or netting technology, as already emphasized, is rooted in a fundamental identity: ***sent payments*** = ***received payments*** in network. From this identity, one can derive a function that describes the position of each individual payment intermediary within the network of intermediaries: total payments = f (cumulative inflows). At the same time, this provides further evidence that in any settlement system, whether gross or net, *payment decisions* are governed by absolute complementarity—the liquidity of one participant depends on the payments of all others and on the liquidity of the entire system.

Net settlement has been accompanied by two pejorative attributes – implicit transactions and deferred payments. The first has already been clarified as a product of monetary blindness, which fails to recognize clearing as a legitimate method of payment. The second is tied to the early days of modern clearing houses, in the late 18th and early 19th centuries, when bankers would meet at the end of the business day to perform clearing of mutual claims and, finally, settlement. This practice persisted for the next two centuries, and such systems became known as Deferred Net Settlement (DNS). Even today, there are similar private payment systems operating as hybrids between the RTGS and DNS models.[3] By far the most well-known and largest such system, both in the United States and globally, is the CHIPS network. However, deferred settlement does not mean delays of payment—it simply refers to the system rules. Technically, net settlement can be performed every minute throughout the business day. Therefore, RTGS systems have since been upgraded by introducing netting schemes to bridge illiquidity.[4]

All available evidence points to the conclusion that RTGS systems are not superior, but rather inherently inferior to net systems.[5] Nevertheless, gross settlement systems have become dominant; their expansion during the 1990s was driven by the need for safer foreign exchange transactions and securities trading. The imposition of RTGS as the standard model for large-value payment systems had two key motivations. First, to minimize the risk of default in the largest financial market—the global foreign exchange market—through the use of the *Payment versus Payment*

(PvP) mechanism. Second, to mitigate problems of securities delivery and funds transfer through the *Delivery versus Payment* (DvP) mechanism.[6] Technically, both PvP and DvP would not be able to function without RTGS.[7] Yet, this trinity of gross settlement mechanisms neither prevented nor mitigated the GFC 2007–9.

Thus, these anachronistic gross settlement models (RTGS with PvP and DvP) have proven to be the way backward. A potential solution could lie in Real-Time Net Settlement (RTNS), based on accelerated sequential netting. A proposal for an RTNS payment system is outlined in the next section.

## 3 RTNS solution: accelerated sequential netting

The basic prudence post-crisis measures focus on short-term resilience of a bank's liquidity and longer-term sustainability of its solvency. These measures involve two minimum standards: Liquidity Coverage Ratio (LCR) and Net Stable Funding Ratio (NSFR), which form the backbone of the Basel III framework (BIS, 2010). For the analysis of the RTNS solution (Real-Time Net Settlement), the LCR standard is primarily relevant.

**LCR = Stock of high-quality liquid assets / Total net cash outflows over the next 30 calendar days ≥ 100%**

The numerator of this ratio is commonly referred to by the acronym HQLA (*High-Quality Liquid Assets*). 'At a minimum, the stock of liquid assets should enable the bank to survive until Day 30 of the stress scenario' (ibid, p. 3). The denominator of this *ratio = outflows – Min {inflows; 75% of outflows}. Cash outflows include retail deposit run-off, outflows of other deposits, as well as unsecured wholesale funding run-off* (ibid, p. 12–16).

The definition of the LCR standard reveals the theoretical basis of a partial approach, analyzed through *the BIS–BCBS practical example* (Vuković, '*Two primary forms of money'* pp. 20-24). Deposits, by definition, can only flow into other banks, not outside the banking sector, just like most wholesale funding. An exception is the portion that is redirected toward the purchase of government bonds and central bank securities.

However, even under a more comprehensive approach that evaluates the level of systemic liquidity, a high level of reserves in the banking sector does not imply timely payments. On the contrary, in a stress scenario, bankers instinctively engage in hoarding of liquidity, driven by precautionary motives and speculative purposes (Acharya and Merrouche, p. 152). This leads to an uneven distribution of liquidity within the banking sector and a slowed movement of reserves, nearly to the point of freezing.[8] Nevertheless, the most important generator of crisis is fear, although it is largely overlooked in analyses of systemic illiquidity. Therefore, it is specifically emphasized here in the development of a liquidity crisis.

**The pattern of liquidity crisis:**
**Distress ⇒ Uncertainty ⇒ Untrust ⇒ FEAR ⇒ Hoarding ⇒ Panic ⇒ Sale ⇒ Freezing**

With the onset of hoarding of reserves, the links between incoming and outgoing payments—which are the core of mediated money, or payments in a network—rapidly break down. In the panic phase, comprehensive de-networking threatens the liquidity of all banks, even though they attempt to isolate themselves. However, the illiquidity of banks is likely the only 'contagion' which is actually worsened by the isolation.

The description of the banking panic of 1933 best illustrates this process of breakdown: 'The System was demoralized. Each Bank was operating on its own. All participated in the general atmosphere of panic that was spreading in the financial community and the community at large' (Friedman–Schwartz, p. 391). It is evident that in interbank payment systems, all banks are mutual counterparties, regardless of whether they have direct payment relationships or not.

Therefore, the spirit of cooperation and mutual trust among participants is necessarily stronger in the CHIPS network than in Fedwire, or in CHAPS than in the BoE's RTGS system.[9] After all, the first clearing house was created precisely due to the spirit of cooperation among participants and their mutual trust (Thornton, p. 101). Respect for clearing house rules has always been unquestioned, even when those rules limited *excessive balances* by defining a "maximum available balance" (PRC).[10]

A more detailed comparison of clearing house rules and the LCR minimum standard shows that internal rules have been more important than the official standards in achieving a more balanced distribution of systemic liquidity and in preventing precautionary hoarding of reserves. Nevertheless, neither rules nor standards would have had significant impact on the prevention of systemic liquidity risk without the traditional measure – expansionary monetary policy, but on an unprecedented scale. 'The massive expansion of reserve balances since fall 2008 and the payment of interest on reserve balances have altered the intraday liquidity management practices of financial institutions' (Bech et al, p. 2). This was a true Big Bang of central bank mediated money: 'Consequently, the level of reserve balances ballooned from $10 billion on average during August 2008 to $850 billion by year-end' (ibid, p. 12).

Despite such excessive reserves in the payment network, which remained largely abundant during the 2010s and early 2020s, there are no guarantees that systemic illiquidity contagion has been preempted. This was clearly demonstrated by the eruption of the international liquidity crisis in March 2023 (BIS, 2023). The explanation lies in the pattern of liquidity crisis, which reflects the behavior of payment intermediaries, as well as game theory, which assists in analyzing their strategies in cooperative and noncooperative games characterized by strong interdependence.

Up to the phase of untrust in the crisis, banks behave relatively cooperatively, and then conflict actions prevail along with the domination of own interest. Understandably, fear triggers the survival instinct, which ignores a common interest and leads to deeper isolation from other participants. A collapse ensues if the central bank and government do not intervene quickly and decisively.

Hence, the eternal search continues for the 'alchemical liquidity stone': '[A]n adequate regulatory mechanism should be in place to prevent banks from being "too interconnected to fail" and thus reduce the likelihood of a system-wide liquidity strain occurring' (Silva, p. 2). Unfortunately, such a mechanism cannot exist due to the total interdependence in the payment system: all actions of all payment agents are interdependent.[11] The gross settlement system of transaction-by-transaction is precisely a product of such an alchemical attempt, which survives only on the support of expansionary monetary policy.

On the other hand, the gross settlement system of transaction-by-transaction has reinforced a competitive spirit and blurred the conditionality of own interest by common interest. Since the early 1990s, the payment game has become increasingly noncooperative.[12]

The insoluble problem is the conflict-dominated own interest in a system of strongly interdependent participants. At first glance, the dominating strategy of each participant in the payment system is to release outgoing payments at the end of the working day. In essence, this is a self-destructive strategy. In the payment system, a common interest maximizes the own. The most important components of common interest are system liquidity and payment efficiency. These goals are inevitably threatened by competitors' confrontation in a noncooperative game. Even in the case of an oligopoly game, the oligopolist succumbs to the pressures of uncertainty and untrust. In the end, a cooperative game can to some extent be externally imposed on participants—through law, central bank regulation, and payment system rules.

The gross-settlement system of transaction-by-transaction causes a form of dictatorship of incoming payments, i.e., de-networking of interbank payments. As a result, only reserves at the bank account are perceived as one's own liquidity, while incoming payments are viewed as free funds from other participants. In doing so, they separate their outgoing payment from our incoming payment, even though it is in fact a single unified transaction between the payer bank and the payee bank. This form of de-networking in gross settlement can be described in a simple schematic way.

The elucidation of gross settlement system:

***Account A => outgoing payment / incoming payment => Account B => outgoing payment / incoming payment => Account A.***

Opposite to him is the elucidation of net settlement system:

***Account A => outgoing payment => SETTLEMENT <= outgoing payment <= Account B.***

**The proposition** of the **RTNS** (Real-Time Net Settlement) payment model is based on the presented net settlement system. Briefly – banks send their own outgoing payments to the system operator by payment messages for sequential multilateral netting and settlement at the end of predefined cycles – minimum every two minutes, maximum fifteen minutes. The two-minute sequence corresponds to the maximum cognitive capabilities of liquidity managers for monitoring a payment algorithm and dynamic decision-making, although the technological capacity for processing payments is nearly unlimited in the AI era. It is evident that such accelerated sequential netting is incomparably less exposed to systemic liquidity risk than deferred net settlement. For example, the liquidity risk of a two-minute cycle is approximately 300 times lower than end-of-day settlement for ten working hours. Likewise, credit risk is proportionally reduced, as well as overall settlement risk.

The fundamental rule is that each **payment message** = **payment order**. At the end of the cycle, the system operator calculates interbank obligations from payment messages through multilateral netting. *Offset payments* are immediately entered into system records and bank books as final, irrevocable, and unconditional, without any deferment.[13] Remaining non-settled payment messages are placed in queuing according to the system's rules and instructions from the payer bank.

The main motive of banks for a deferred cycle is to collect as many payment messages as possible for facilitated netting and settlement. However, for high-value payment systems, the problem is not the accumulation of payments, but rather their restriction. For instance, the value of a single payment in Fedwire is limited to one cent less than $10 billion. The bigger problem is the insufficiency of small transactions to complete the multilateral *mosaic of clearing*. Prefunded balances also can help to complete clearing, as the maximum possible net debit position. These balances in net settlement systems are smaller than in RTGS, by a definition (Dent and Dison, p. 236).[14] In addition, intraday credits of the central bank can be used.

In the proposed RTNS payment system, participants will be strongly motivated to send as many payment messages earlier, instead of transferring scarce and costly reserves from their transaction account. Friedman's monetary deceiving – the interpretation of netted payments as implicit transactions – would have a strong effect here. Of course, the truth is the opposite: netted payments are explicit transactions, just like all other easily visible moneyness transactions. From this arises the general rule of payment: **payment flows = money flows**.

For a comparison of the RTNS model's performance with prevailing RTGS payment systems, *the BIS – BCBS practical example*, (BIS, 2013b, p. 13-14), can again be used. In the presented projection of interbank payments (Table 1), the banking sector, i.e., the payment network, consists of Bank A and Bank B. The entire analysis was conducted according to RTGS rules. In the upcoming Table 1, alongside RTGS payments, we will simulate net transactions in the payment network under RTNS assumptions – columns *Clearing and Queuing*. It should be noted that the RTGS simulation used bank reserves and intraday credits, while the RTNS projection excluded these, relying in this case on pure clearing and a queuing mechanism.

Table 1 - RTGS vs RTNS

| *Time* | *Sent* | *Received* | *Net* | *Sent* | *Received* | *Net* | ***Clearing*** | *Queuing* |
|---|---|---|---|---|---|---|---|---|
| 07:05 | 450 | | -450 | | 450 | +450 | | 450 |
| 07:10 | | 200 | -250 | 200 | | +250 | 400 | 250 |
| 07:15 | 100 | | -350 | | 100 | +350 | | 350 |
| 07:20 | 200 | | -550 | | 200 | +550 | | 550 |
| 07:25 | | 400 | -150 | 400 | | +150 | 800 | 150 |
| 07:30 | | 300 | +150 | 300 | | -150 | 300 | 150 |
| 07:35 | 300 | | -150 | | 300 | +150 | 300 | 150 |
| 07:40 | | 350 | +200 | 350 | | -200 | 300 | 200 |
| 07:45 | 250 | | -50 | | 250 | +50 | 400 | 50 |
| 07:50 | 100 | | -150 | | 100 | +150 | | 150 |
| 07:55 | | 150 | 0 | 150 | | 0 | 300 | 0 |
| Total | 1,400 | 1,400 | | 1,400 | 1,400 | | 2,800 | |

Sources: BIS (2013b), p. 13-14, and author's calculation.

The simulation shows that RTGS payments were processed faster, which is expected given the use of bank reserves and intraday credits. An additional influence on this result was the unusual distribution of payments in *the BIS – BCBS practical example* – the accumulation of payments (50%) in the first quarter of the working day. Even under such adverse conditions, the RTNS model proves superior, as payments are executed in short cycles, which in this case were projected at five minutes.

The following Table 1a vividly illustrates the dynamics of RTGS and RTNS payments per each five-minute settlement cycle and in total. All payments in the cycles represent the sum of payments by Bank A and Bank B (1,400 + 1,400 = 2,800). In the RTNS model, all netted payments are summed within the predefined cycle according to the rule:

Account A => outgoing payment => SETTLEMENT <= outgoing payment <= Account B.

Table 1a – Payment dynamics: **RTGS** (with reserves and intraday credits) vs **RTNS** (without reserves and intraday credits)

| *Time* | *Payments* | *Total payments* | *Payments* | *Total payments* |
|---|---|---|---|---|
| 07:05 | 450 | **450** | | **0** |
| 07:10 | 200 | **650** | 400 | **400** |
| 07:15 | 100 | **750** | | **400** |
| 07:20 | 200 | **950** | | **400** |
| 07:25 | 400 | **1,350** | 800 | **1,200** |
| 07:30 | 300 | **1,650** | 300 | **1,500** |
| 07:35 | 300 | **1,950** | 300 | **1,800** |
| 07:40 | 350 | **2,300** | 300 | **2,100** |
| 07:45 | 250 | **2,550** | 400 | **2,500** |
| 07:50 | 100 | **2,650** | | **2,500** |
| 07:55 | 150 | **2,800** | 300 | **2,800** |
| Total | 2,800 | | 2,800 | |

Sources: BIS (2013b), p. 13-14, and author's calculation.

The comparison of payment effectiveness between the RTGS system, with reserves and intraday credits, and the RTNS model, without reserves and intraday credits, testifies to the superiority of the proposed model. It is necessary to emphasize that the author's motive for proposing the RTNS model is not 'saving' liquidity or faster liquidity 'recycling', but the creation of more reliable, liquid, and efficient wholesale and retail payment systems.

The superiority of net settlement systems is proven in practice by the record-setting efficiency of the largest system of its kind in the world – the CHIPS network. 'The liquidity efficiency of CHIPS averaged 26:1 in 2023, meaning that $1 contributed to the network in funding supported $26 in settled value, with most payments settling in mere seconds. This contrasts with the average of 6.6:1 liquidity efficiency in other major large-value payment systems around the world' (CHIPS News, April 02, 2024). The cited CHIPS network payment efficiency can be explained by its fundamental characteristics. 'CHIPS is a real-time final settlement system that continuously matches, nets and settles payment messages. […] On a daily basis, the system provides real-time finality for all payment orders released from the CHIPS queue' (BIS-CPSS). [15] Additional support for the functioning of this system is provided by the CHIPS prefunded balance account (ibidem). [16]

The proposed RTNS model differs from the CHIPS network by its own accelerated sequentiality and finality at the end of a predefined cycle. However, the most important thing is that *funds on account* and *funds in a payment order* are treated equally in settlement processes. Thus, money flow is not hidden behind money stock.

According to an official source from the Federal Reserve, RTNS systems already exist. 'Some countries use real-time net settlement, where transactions net every few seconds, essentially in real time' (FRB services, p. 5). However, a detailed internet search reveals no such payment system.

Finally, the proposed RTNS model can be succinctly defined as a real-time net settlement payment system with accelerated sequential multilateral netting/clearing in predefined cycles (from minimum two minutes to maximum fifteen minutes) on a final offsetting basis, along with the use of prefunded balances and intraday credits.

## 4 Conclusion

The net settlement model dominated in clearing houses for more than two centuries, until the early 1990s. At that time, central bankers concluded that explicit transactions with explicit money at the account, i.e., gross settlement, were more desirable. Since then, the stigma of implicitness has followed the net settlement model, although it has continued to survive—even in the shadow of RTGS—as the most efficient way to reduce settlement risk.

Despite the dominance of RTGS systems in wholesale payments, the practice of gross settlement has not been able to eliminate the linking of incoming and outgoing transactions, nor the cautious use of reserves. All available evidence suggests that RTGS systems are not superior, but rather inherently inferior to net systems.

On the other hand, gross settlement, in a transaction-by-transaction model, has intensified competitive behavior and blurred the conditionality of own interest by common interest. Since the early 1990s, the payment game has become increasingly noncooperative. RTGS causes a kind of dictatorship of incoming payments, i.e., de-networking of interbank payments. Thus, only reserves at the bank account are seen as own liquidity, while incoming payments are treated as free funds provided by other participants. At the same time, there is a separation between our incoming payment and their outgoing payment, even though this is a single transaction between the payer bank and payee bank.

Elucidation of the **gross** settlement system: Account A => outgoing payment / incoming payment => Account B => outgoing payment / incoming payment => Account A. Opposite to this is the **net** settlement system: Account A => outgoing payment => SETTLEMENT <= outgoing payment <= Account B.

A high level of reserves in the banking sector does not imply timely payments. On the contrary, in stress scenarios, bankers instinctively engage in hoarding of liquidity, driven by precautionary and speculative motives. Yet, the most important generator of crisis is fear.
The pattern of a liquidity crisis: Distress => Uncertainty => Untrust => FEAR => Hoarding => Panic => Sale => Freezing. Once hoarding of reserves begins, the connections between incoming and outgoing payments, which constitute the essence of mediated money, i.e., payments in network—rapidly break down. In the panic phase, comprehensive de-networking threatens the liquidity of all banks, even though they attempt to isolate themselves. However, bank illiquidity is probably the only contagion that is worsened by isolation.

A potential solution could be Real-Time Net Settlement (RTNS). The proposed RTNS model can be succinctly defined as a real-time net settlement payment system with accelerated sequential multilateral netting/clearing in predefined cycles (from a minimum of two minutes to a maximum of fifteen minutes) on a final offsetting basis, along with the use of prefunded balances and intraday credits.

The fundamental rule is that each **payment message** = **payment order**. At the end of the cycle, the system operator calculates interbank obligations from payment messages through multilateral netting. *Offset payments* are immediately entered into system records and bank books as final, irrevocable, and unconditional, without any deferment. Remaining non-settled payment messages are placed in queuing according to the system's rules and instructions from the payer bank.

*First published on* https://centralbankmoneyresearch.com/

***References***

Acharya, Viral and Quarda Merrouche (2012), Precautionary hoarding of liquidity and interbank markets: Evidence from the Subprime crisis, *Review of Finance*, No 17, pp. 107-160.

Afonso, Gara, Darrell Duffie, Lorenzo Rigon, and Hyun Song Shin (2022), How Abundant Are Reserves? Evidence from the Wholesale Payment System, *Staff Reports no. 1040,* FRB of New York, November 2022.

Angell, Wayne (1992), Large-Value Payment Systems: What have we learned? *Remarks*, Board of Governor of the Federal Reserve System.

Bech, Morten, Antoine Martin, and James McAndrews (2012), Settlement liquidity and monetary policy: Lessons from the Financial Crisis, *FRBNY Economic Policy Review*, FRB of New York, March 2012, pp. 1-25.

Benos, Evangelos, Rodney Garratt, and Peter Zimmerman (2014), The role of counterparty risk in CHAPS following the collapse of Lehman Brothers, *International Journal of Central Banking*, Vol. 10, No. 4, December 2014.

Beyeler, Wolter, Robert Glass, Morten Bech, and Kimmo Soramaki (2006), Congestion and cascades in payment systems, *Staff Report*, no. 259, FRB New York.

BIS (1997), Real-time gross settlement systems, *Report*, Committee on Payment and Settlement Systems, Bank for International Settlements, March 1997.

BIS (2010), Basel III: International framework for liquidity risk measurement, standards and monitoring, Basel Committee on Banking Supervision, Bank for International Settlements, December 2010.

BIS (2013a), Basel III: The Liquidity Coverage Ration and liquidity risk monitoring tools, Basel Committee on Banking Supervision, Bank for International Settlements, January 2013.

BIS (2013b), Monitoring tools for intraday liquidity management, Basel Committee on Banking Supervision, Bank for International Settlements, April 2013.

BIS (2023), Report on the 2023 banking turmoil, Basel Committee on Banking Supervision, Bank for International Settlements, October 2023.

BoE (2024), A brief introduction to the Real-Time Gross Settlement system and Chaps, Bank of England.

Dent, Andrew and Will Dison (2012), The Bank of England's Real-Time Gross Settlement infrastructure, *Quarterly Bulletin*, Volume 52, Number 3 (September), Bank of England, pp. 234-243.

FRB services (2024), Unpacking clearing and settlement, The Federal Reserve – FRB services, 3/13/2024.

Friedman, Milton and Anna Schwartz (1993), *A Monetary History of the United States, 1867-1960*, Ninth Ed., NBER & Princeton University Press.

Garreth, Rule (2015), Understanding the central bank balance sheet, Centre for Central bank Studies, Bank of England, 2015.

Johnson, Kurt, James McAndrews and Kimmo Soramaki (2004), Economizing on liquidity with Deferred Settlement Mechanisms, *FRBNY Economic Policy Review*, FRB of New York, December 2004, pp. 51-72.

Kahn, Charles, James McAnrews, and William Roberds (1999), Settlement risk under Gross and Net settlement, *Working Paper*, 99-10a, Federal Reserve Bank of Atlanta, November 1999.

Leigh-Pemberton, Robert (1992), Developments in wholesale payments systems, *Quarterly Bulletin*, Volume 32, Number 4 (November), Bank of England, pp. 449-457.

PRC (2000), Intraday liquidity management in the evolving payment system: A study of the impact of the euro, CLS Bank, and CHIPS finality, The Payment Risk Committee, New York, April 2000.

Silva, Andre (2016), Strategic complementarity in banks' funding liquidity choices and financial stability, *Working Paper Series*, No 19, European System Risk Board, July 2016.

Thornton, Henry (1802), *An inquiry into the nature and effects of the paper credit of Great Britain*, Reprint, New York, 1965.

Turocy, Theodore and Bernhard von Stengel (2001), Game Theory, *CDAM Research Report*, 2001-09, LSE, October 2001.

[1] 'We believe that the capacity of netting to reduce settlement risk is an important factor behind the historical predominance of net settlement arrangements for large-value transactions' (Kahn et al, 17).
[2] 'When the system becomes congested payment settlement loses correlation with payment arrival. Payment settlement takes place in cascades and is governed by the internal dynamics of the coupled payment queues' (Beyeler et al, p. 13).
[3] 'In addition, in the United States and the euro area, a key feature in the landscape of large value payment systems (LVPS) is "cohabitation" between RTGS system operated by the central bank and a deffered net settlement (DNS) system operated by a private entity, as is the case with CHIPS in the United States and the EURO1 system in Europe' (Banque de France (2017), Payments and market infrastructures in the digital era, p. 120).
[4] 'However, because payment transfers between participants are made immediately in the full amount, and because of the asynchronous timing of payments by participants, maintaining the liquidity needs of RTGS systems can be costly. Indeed, some system operators have altered their RTGS systems in recent years to economize on the funds needed to complete settlements. One way to reduce a system's liquidity needs is by using deferred settlement mechanisms such as netting' (Johnson et al, p. 51).
[5] '[O]ur analysis leads to the conclusion that one should not interpret this trend as proof that RTGS systems are inherently superior or less risky than netting systems' (Kahn et al, p. 28).
[6] **Payment versus payment** – A settlement mechanism that ensures that the final transfer of a payment in one currency occurs if and only if the final transfer of a payment in another currency or currencies takes place' (BIS, 2021, Principles for financial market infrastructures, p. 177). '**Delivery versus payment** - A securities settlement mechanism that links a securities transfer and a funs transfer in such a way as to ensure that delivery occurs if and only if the corresponding payment occurs' (ibid, p. 176).
[7] 'In addition, RTGS can contribute to the reduction of settlement risk in securities and foreign exchange transactions by providing a basis for delivery-versus-payment (DVP) or payment-versus-payment (PVP) mechanism' (BIS, 1997, p. 1).
[8] 'Many commercial banks started to hoard reserves instead of lending in interbank markets, partly to self-insure against potential adverse shocks to themselves and partly out of heightened credit risk fears with regards to potential counterparties' (Garreth, p. 7).
[9] 'Since November 2017, the Bank has also been responsible for operating the **CHAPS system**: setting the rules and technical standards for the CHAPS system and acting as an end-to-end risk manager. […] The Bank operates RTGS and CHAPS with combined operations, risk and analytical teams, all sitting under an integrated and strengthened set of governance and risk management arrangements' (Bank of England, 2024, p. 4).
[10] 'The initial prefunded balance becomes a participant's opening balance on CHIPS. […] However, in no case will a participant's available balance be permitted to fall below zero or to rise to more than two times the initial prefunded balance (the "maximum available balance"). This latter limitation prevents a participant that might be having system problems releasing payment messages from building an excessive balance and absorbing excessive system liquidity' (PRC, p. 45).
[11] 'We find that, even in the era of "ample" reserves, the amount of payments that a bank makes in a given minute depends significantly on the amount of payments it has received over preceding minutes, indicating a high degree of strategic complementarity' (Afonso et al, 2022, p. 1).
[12] 'The term "noncooperative" means this branch of game theory explicitly models the process of players making choices out of their own interest. Cooperation can, and often does, arise in noncooperative models of games, when players find it in their own best interests' (Turocy and von Stengel, p. 6-7).
[13] 'The netting and the posting of the debits and credits to the available balance constitute final settlement of all payment messages in each batch. It therefore satisfies, finally and irrevocably, the obligation of each paying participant to pay the amount of the payment message to the receiving participant. CHIPS payment messages will no longer be subject to final settlement at the end of the day. They will be finally settled upon receipt' (PRC, p. 45).

[14] 'The 'liquidity efficiency' of a model refers to the value of payments that can be settled for a given amount of liquidity. The DNS model is more liquidity efficient than the RTGS model as only the net obligations incurred between banks during a clearing cycle are settled, and these will always be less than (or equal to) the gross values' (Dent and Dison, p. 236).
[15] BIS-CPSS (2012), Red book - Payment, clearing and settlement systems in the United States, Bank for International Settlement - The Committee on Payment and Settlement Systems, p. 489.
[16] 'To facilitate settlement, the CHIPS prefunded balance account (CHIPS account) was established at the federal Reserve Bank of New York. Under the real-time finality arrangement, each CHIPS participant has a pre-established opening position requirement, which, once funded via a Fedwire Funds transfer to the CHIPS account, is used to settle payment orders throughout the day' (ibidem).