Settlement trends in real-time payment systems

What is happening today and where is the industry going?

Contents

Executive summary

"Various local factors appear to drive the choice of strategy: the role the central bank wishes to play in the real-time payment system, and customs and practice in existing payment systems."

As economies become increasingly digital, the expectation is that financial interactions will keep pace. This is reflected in the deployment in many countries around the world of real-time payments services, with their 24/7 availability. The volumes have grown steadily and are widely expected to continue to do so as real-time payments become ever more important in our daily lives.

Real-time systems, like other payment systems, involve two main processes: clearing, which is the exchange of payment messages between banks, and settlement, which is the transfer of funds between banks arising from these payments. This paper looks specifically at the way that inter-bank settlement processes are operating today and the trends we observe in the market.

Central banks play a key role in the design and operation of payment systems, particularly the settlement procedures. Their design needs to ensure that the risk that a bank cannot settle its payments is eliminated or minimized, and must ensure smooth, reliable operation to avoid impacting citizens, businesses and the wider economy. Various strategies have evolved in bulk systems to address these requirements. However, real-time payments present particular challenges that affect the design, in particular, high volumes, 24/7 operation, and the need for settlement to be completed in real time. This has led to two broad approaches, each implemented in several ways.
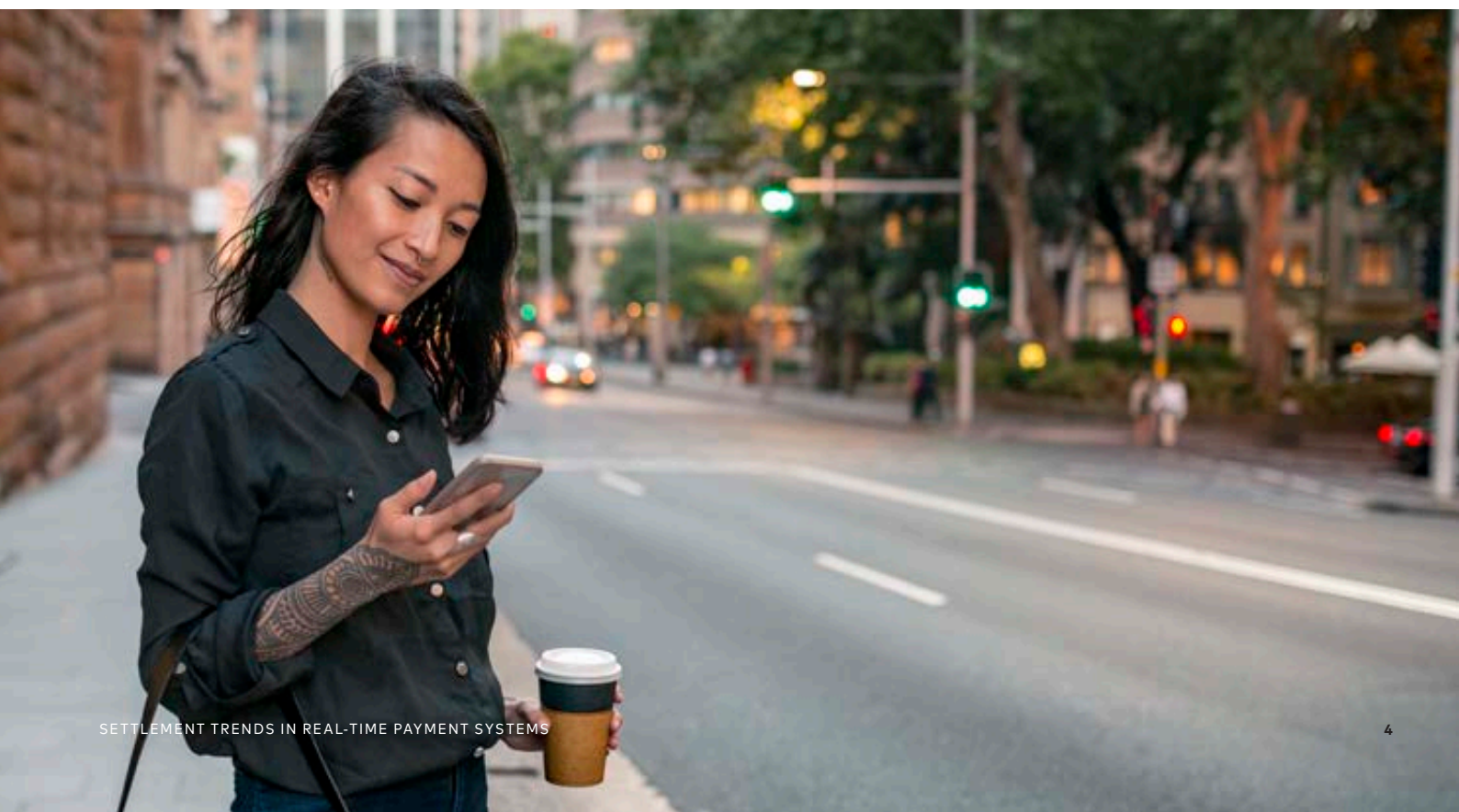
- **Deferred multilateral net basis:** In such systems, payments are settled net at the central bank several times per day after messages have been exchanged. The successful completion of settlement is ensured in various ways, including prefunding, loss sharing and hybrid solutions.
- **Real-time gross basis:** This means that each payment is settled individually as part of the clearing process. This may take place within commercially operated real-time payment systems, or the central bank may provide 24/7 real-time payment or settlement capabilities.

Various local factors appear to drive the choice of strategy. In our experience, two (probably related) factors stand out: firstly, the role the central bank wishes to play in the real-time payment system and its broader policies for payment system settlement; and secondly, customs and practice in existing payment systems in the market which often appear to lead to similar solutions being adopted for real-time payments. These factors tend to lead to market-specific solutions, which may be similar to those of other markets but are seldom identical. It also probably means that there is no single best strategy, only the right solution for each market.

"Tools that exist in a few systems to dynamically predict and manage liquidity may become more common."

The increased use of real-time payments is likely to lead to increased focus on their efficiency, including their settlement procedures. While fundamental change is unlikely, several themes for the ways that settlement may evolve are emerging. Although a few central banks are providing 24/7 settlement services, most provide services only during "bankers' hours."

This is a focus for banks and the authorities in a broader context. For example, in 2023, the CPMI (Committee on Payments and Market Infrastructures) published its report "Operational and technical considerations for extending and aligning payment system operating hours for cross-border payment: An analytical framework." This indicates that there will be a move toward more central banks providing real-time gross settlement for individual transactions in the same way that the Reserve Bank of Australia and the European Central Bank (ECB) already do. Alternatively, other central banks may choose to provide extended settlement windows to support deferred net systems as well as other settlement requirements. Another potential area of focus may be liquidity efficiency. As payment values rise and during times when liquidity costs are high, banks and other parties may look at the amount of liquidity dedicated to real-time payment settlement, particularly in systems that use prefunding. Tools that exist in a few systems to dynamically predict and manage liquidity may become more common.

Introduction

"The real-time payments market continues to evolve, and there is significant interest in the way that real-time payment settlement is also evolving."

Today we live in a world where there are multiple ways to pay for goods and services. In recent years, real-time payment systems have been developed in many countries, allowing consumers and businesses to make payments instantly between their bank accounts 24/7. The speed at which real-time payment volumes are growing varies according to local factors, such as use cases supported. But there is no doubt that volumes will continue to grow, and there is already increased focus on the efficiency of such systems.

All payment systems, including real-time systems, involve two processes: clearing and settlement. Clearing is the exchange of payment messages between banks, usually involving a processor or automated clearing house (ACH). Settlement is the process by which the funds relating to these payment messages are transferred between banks. In most systems, settlement involves the transfer of funds between the central bank accounts of the banks involved, often facilitated by the processor.

The real-time payments market continues to evolve, and there is significant interest in the way that real-time payment settlement is also evolving. This paper explores settlement procedures around the world, based on Mastercard's learnings as a major supplier of real-time payment services. Although many real-time payment systems, particularly those developed in the past 10 years, have similar features (e.g., use of ISO 20022 messaging), the way payments are settled often varies considerably. This paper explores this in terms of:

- The requirements and challenges for real-time payment settlement
- The settlement strategies that Mastercard has encountered in the market and the different ways these strategies have been implemented
- Influencing factors that determine the use of particular strategies
- Examples (not exhaustive) of where the different strategies are used and observations about trends
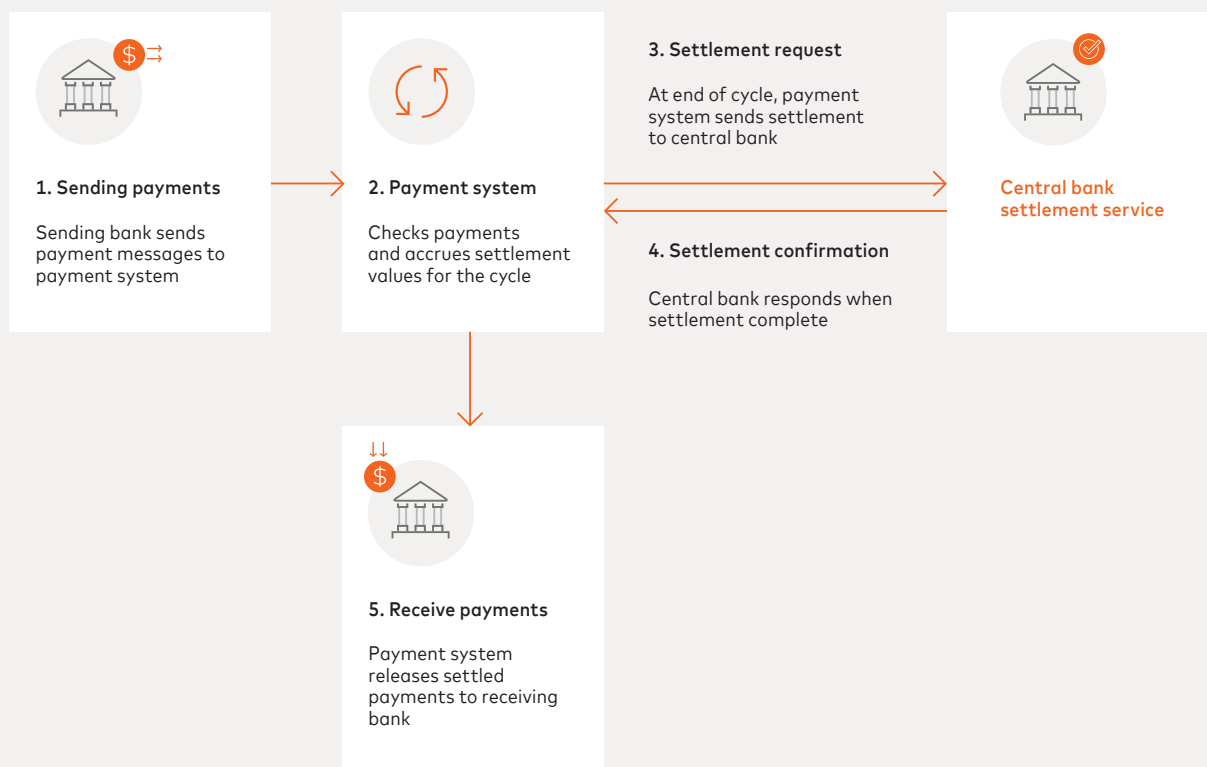- Some advantages and disadvantages of different settlement models

Requirements and challenges for real-time payment system settlement

"If a bank is unable to settle, its payments will stop flowing. This is typically called settlement before output or settlement before exchange."

When any new payment system is developed, the regulators, banks and other market players will consider a range of requirements that apply to settlement. First, it is important that the system eliminates or minimizes the risk that a bank is unable to settle its transactions. Second, even if a bank does fail, payment systems should continue to provide reliable continuous operation between the remaining banks, minimizing the impacts on businesses and citizens.

Non-real-time payment systems (otherwise known as bulk) use several strategies to do this. The most common strategy is to require settlement to be completed successfully at the central bank *before* messages are exchanged between banks (cleared). If a bank is unable to settle, its payments will stop flowing. This is typically called settlement before output or settlement before exchange and is illustrated in Figure 1 below.

Figure 1: Settlement before exchange

1. Sending payments

Sending bank sends payment messages to payment system

2. Payment system

Checks payments and accrues settlement values for the cycle

3. Settlement request

At end of cycle, payment system sends settlement to central bank

4. Settlement confirmation

Central bank responds when settlement complete

Central bank settlement service

5. Receive payments

Payment system releases settled payments to receiving bank

"Real-time payment systems, by their very nature, pose specific challenges for settlement."

Real-time payment systems, by their very nature, pose specific challenges for settlement that make it harder to meet these requirements using non-real-time payment settlement strategies.

- **24/7 processing:** Settlement needs to take place continuously, even if the central bank settlement procedures are not available.

- **Real-time settlement and finality:** In real time, the processing must ensure each payment is (a) legally complete (payment finality) and (b) settled or guaranteed for later settlement (settlement finality).

- **High volumes:** Real-time payments are processed singly, not in batches. Even if the central bank settlement system is available 24/7, it may need to be able to handle thousands of settlement transactions per second. This is not something that bank real-time gross settlement (RTGS) systems are usually designed to do.

Settlement strategies

"Irrespective of the settlement model, all real-time systems... involve some form of bank prefunding."
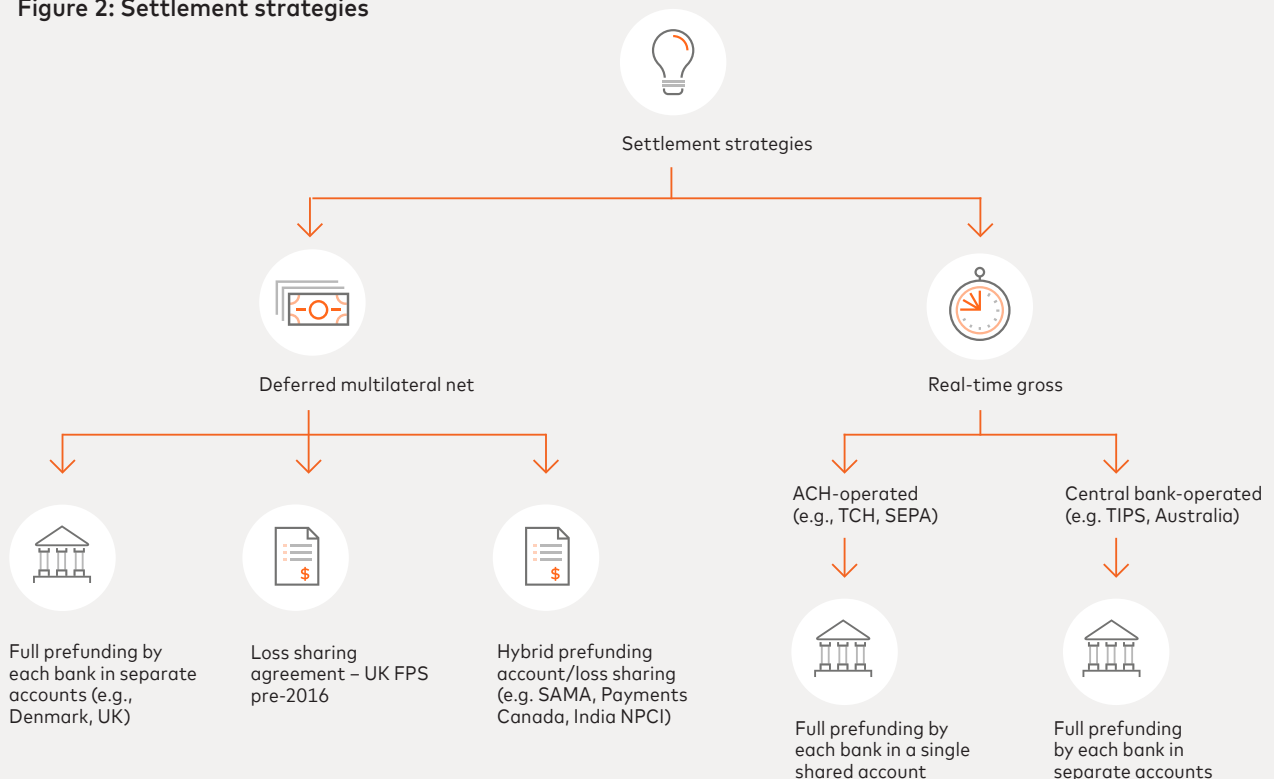
Broadly speaking, there are two strategies in use:

- **Deferred multilateral net**, where settlement happens after the payment message is cleared. Settlement is based on multilateral net positions that are settled several times per day.
- **Real-time gross** (also known as "line by line"), where clearing and settlement happen at the same time during the processing of each payment, and where each payment is settled individually on a gross basis.

Irrespective of the settlement model, all real-time systems, apart from those that solely use loss-sharing agreements, involve some form of bank prefunding. In other words, participant banks are required to deposit liquidity (cash or other defined collateral) at the central bank before making any payments. They can make payments only up to the value of their deposits. This is necessary to ensure that settlement can always be completed.

The different implementations of each of these strategies are illustrated in Figure 2 below.

Figure 2: Settlement strategies

Settlement strategies

Deferred multilateral net

- Full prefunding by each bank in separate accounts (e.g., Denmark, UK)
- Loss sharing agreement – UK FPS pre-2016
- Hybrid prefunding account/loss sharing (e.g. SAMA, Payments Canada, India NPCI)

Real-time gross

- ACH-operated (e.g., TCH, SEPA)
  - Full prefunding by each bank in a single shared account
- Central bank-operated (e.g. TIPS, Australia)
  - Full prefunding by each bank in separate accounts

Deferred multilateral net

The settlement procedure is usually the same, involving several settlement cycles per day. However, the way that settlement risk is managed varies.

- **Full prefunding** by each bank in separate accounts: Banks deposit funds or pledge collateral at the central bank "covering" 100% of the value of payments they will make. These funds are ring-fenced and cannot be used for other purposes to guard against bank insolvency. In some systems, these funds are held in a separate account from the account used to fund settlement (e.g., the U.K.'s FPS). In other systems, the banks have a single account that is used for prefunding and settlement (e.g., Denmark's Express Clearing). Bank caps/limits for clearing payments in these systems are based on the value of prefunding deposits.

- **Loss sharing:** Banks agree that if one or more banks fail, the surviving banks will settle the obligations of the failed bank or banks — referred to as a "survivors pay" scheme. Bank caps/limits in these systems are set by the scheme/central bank based on risk and the loss-sharing agreement. FPS used this strategy before 2016; no other loss-sharing systems are known.

- **Hybrid loss sharing/prefunding:** Several systems fall into this category (e.g., Saudi Arabia's SAMA, Payments Canada). In such systems, each bank is subject to a cap on outgoing payments in the real-time payment system. However, unlike in fully prefunded systems, the cap does not correspond to an equivalent amount of dedicated liquidity held at the central bank specifically for real-time payment system settlement. The cap is usually set by the scheme/central bank based on a local strategy. In some systems, banks contribute to a central fund whose size will be based on a scheme/central bank formula, e.g., the sum of the largest debit positions of the two biggest banks in the preceding six months. This fund will be used if a bank fails to settle. But if it is not big enough or the fund is used, the loss-sharing element kicks in to top up the fund. This means banks do not individually fully prefund their payments, so it is more liquidity-efficient than full prefunding, but with some residual risk. In other systems (e.g., the UAE), the central bank takes a holistic view of banks' assets/positions and controls banks' real-time payment system limits (caps) as part of a broader oversight role. Where there is a settlement problem, it is also possible that the central bank, not the remaining banks, would make good on any shortfall.

> "In Denmark, every 15 minutes the central system checks whether any banks need to add liquidity because they are below the low funds threshold."

Real-time gross

- **ACH/processor-operated:** Full prefunding by each bank based on a single shared prefunding account for the system into which all participants deposit liquidity. The processor is the system of record and tracks the balance of each participant as payments are made. TCH in the U.S. and the current Swedish BIR system are examples of this strategy. This is sometimes considered as settlement in commercial bank money because the accounting is outside the central bank.

- **Central bank-operated:** Technically very similar to the above, but in this case the central bank will operate the settlement system, which may or may not be 24/7. Each bank fully prefunds its payments via its own central bank settlement account. The accounts of each bank are tracked and updated by the system.

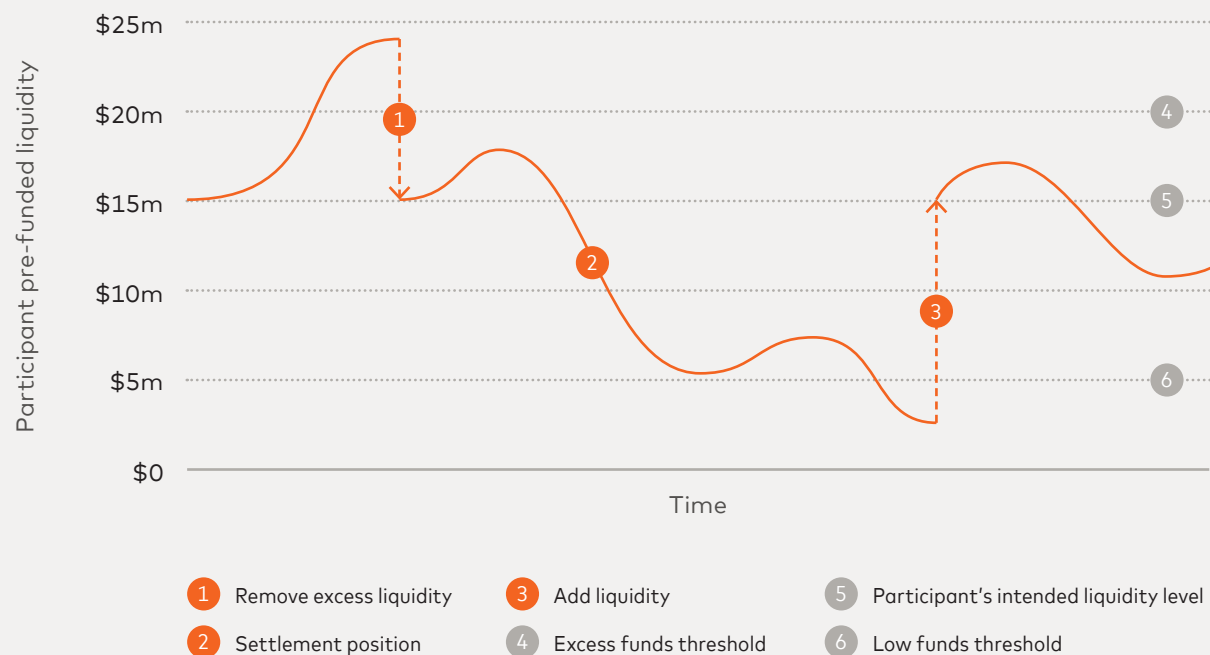There is more detail on how these models operate in the Appendix on Page 17.

Other controls

There are two other areas of real-time payment settlement that may vary independently of the model.

Liquidity monitoring and management: The first area relates to the way that banks can add and remove liquidity committed to settlement prefunding. These processes are usually called funding (also known as top-ups) and defunding (also known as drawdowns). Typically, banks send requests to the central bank to effect these changes. Most systems also have high and low funds thresholds that trigger alerts to help banks manage their liquidity and ensure they stay within their desired liquidity range. See Figure 3 on next page.

Inter-bank settlements usually takes place in central banks. They play a key role in Denmark, for example, every 15 minutes the central system checks whether any banks need to add liquidity because they are below the low funds threshold. If so, an automatic top-up request is sent from the central system to the central bank. Additionally, at the end of each settlement cycle, excess liquidity is automatically returned to the bank's RTGS account by the central system.

Figure 3: Liquidity monitoring

Legend:
1. Remove excess liquidity
2. Settlement position
3. Add liquidity
4. Excess funds threshold
5. Participant's intended liquidity level
6. Low funds threshold

Axis labels: Participant pre-funded liquidity ($0, $5m, $10m, $15m, $20m, $25m) vs Time

"Because the settlement banks are ultimately responsible for all entities they settle, they need to be able to manage the resulting exposures."

Management of non-settling participants: Along with large established banks, a range of other actors may be given direct access to the real-time payment system (e.g., smaller banks, fintechs, etc.). This means they are technically connected and can submit payments, as distinct from indirect participants that submit payments via another bank. Such entities are known as funded participants and are "sponsored" by a settlement bank (funding participant) that will settle payments on their behalf, or are required to use another bank, system (e.g., smaller banks, fintechs, etc.) who do not settle their own transactions. This may be their choice or it maybe because they are restricted from from obtaining a central bank settlement account. Because the settlement banks are ultimately responsible for all entities they settle, they need to be able to manage the resulting exposures to ensure that any financial problems in their settled entities do not have any systemic effect. Most systems have controls to accomplish this, managed by the settlement participants. The control may be a simple daily credit limit, or the payment system may track the funds the entity has on deposit at the settlement bank. For example, in the Philippines' real-time system, the caps of funded participants are updated as they add or remove funds from their account with the funding participant.

Factors that may determine the choice of settlement strategy

A number of different strategies and designs have emerged. The choice seems to be determined by various market-specific factors, such as:

- **Central bank policies:** These might determine the bank's appetite to invest in new settlement procedures. The ECB and the Reserve Bank of Australia have both developed specific services to support real-time payment settlement. On the other hand, the TCH service in the U.S. and the FPS service in the U.K. were both developed to operate within the then-existing capabilities of national central banks.

- **Commercial money versus central bank settlement:** Some central banks insist that settlement must involve the transfer of funds between the accounts of participant banks at the central bank (i.e., central bank money settlement). Other central banks do not insist on this approach and accept settlement strategies that are managed by the commercial banks (i.e., commercial bank money settlement).

- **Local custom and practice:** While markets are keen to embrace real-time payments, existing customs and practices for payment system settlement often determine how any new real-time payment system is settled. This is why the payment flows in real-time systems are often similar but the settlement procedures may be quite different.

- **Bank liquidity management:** The design of some systems includes tools to help optimize bank liquidity. While values remain low, this may be less of an issue, but as values grow it is likely there will be more focus on this. There are also issues around whether the funds put aside for real-time payment system settlement count toward bank reserve targets and whether interest is paid on these deposits.

Examples of where different strategies are in use and the trend

A non-exhaustive list of how settlement takes place in different real-time payment systems around the world can be found in Table 1 below.

Table 1: Settlement strategy usage

Deferred net, full bank prefunding	Deferred net, hybrid prefunding/loss sharing	Real-time gross, ACH-operated	Real-time gross, central bank-operated
InstaPay, Philippines	Real-Time Rail, Canada	TCH, USA	TIPS, Europe
FPS, U.K.	NPCI, India	BIR, Sweden	NPP, Australia
Express Clearing, Denmark	PromptPay, Thailand	Bankart, Slovenia	SIC, Switzerland
VIPPS, Norway	QPS, Qatar	STET, France	SPEI, Mexico
FAST, Singapore	Aani, UAE	SIBS, Portugal	FedNow, USA
IRT, Peru	SAMA, Saudi Arabia	Nexi, Italy	BI-FAST, Indonesia
DuitNow, Malaysia		EquensWorldline, Netherlands	
		Iberpay, Spain	
		EKS, Latvia	
		EBA Clearing RT1, EU	
		DIAS, Greece	
		CENTRO, Lithuania	
		CEC, Belgium	

Advantages and disadvantages of different models

There is no best strategy for settlement. Each market is different, so it is a question of the right strategy for that market. Nevertheless, there are advantages and disadvantages for different models. Table 2 below has some illustrative points for comparison.

Table 2: Comparison of settlement strategies

Criteria	Deferred net, full bank prefunding account per bank	Deferred net, loss sharing	Deferred net, hybrid prefunding /loss sharing	Real-time gross, processor-operated	Real-time gross, central bank-operated
Settlement risk	None	Settlement risk managed, but risks are uncapped, balance sheet issue	Largely eliminated, but still some residual risk	None	None
Competition issues[1]	None	Concern that loss sharing creates barriers to entry for small banks because they cannot take on the risk of a big bank failing	May still be some concern about barriers to entry for small banks; depends on the formula used for managing settlement risk	None	None
Central bank (CB) concerns re. central bank vs. commercial bank money settlement	None	None	None	Settlement in commercial bank money means central bank does not "see" settlement flows; concerns that processor may fail	None
Central bank impact	Limited/none	Limited/none	Some impact; depends on design of the system	Limited/none (CB needs to support trust account construct)	High because CB needs to build real-time payment settlement platform
Liquidity requirement for banks	Requires banks to fully prefund payments	Depends on CB requirements	Lower than full prefunding	Requires banks to fully prefund payments but may mean that they are not paid interest on deposits	Requires banks to fully prefund payments

1.   It is assumed that any bank is permitted to hold a central bank settlement account, otherwise they would not be able to join the system.

Where is the industry going?

As outlined above, there are several factors that can determine the settlement model used within a particular jurisdiction, and different countries will adopt different models. However, we are seeing several trends emerging as the real-time payments market matures, mainly associated with the role of central banks.

Most central banks do not provide payment and payment system settlement facilities outside of "bankers' hours." Consequently, banks' real-time payment liquidity management strategies need to consider that they cannot easily adjust the funds set aside for settlement when the central banks' systems are unavailable. In some countries, the real-time payment system supports inter-bank payments to facilitate arrangements between participating banks, whereby they can provide funding to each other while the central bank is unavailable. In any event, limited central bank operating hours create some risks and potentially lead to liquidity inefficiencies. As the value of real-time payment traffic increases, these issues are likely to become more important.

The issue of central bank operating hours has broader implications. The technical report issued by the CPMI in February 2023 laid out an analytical framework to assist central banks and operators planning to extend real-time gross settlement (RTGS) system operating hours.[2] The framework is designed to help determine the most appropriate approach to extending operating hours. It evaluates related technical and operational issues and designing and an implementation plan. For the short term, it argues that increasing operating hours on current operating days would be the most achievable option and that other end states could be considered in the medium to long term. However, so far, only a few central banks operate 24/7, and a few more have announced that they are actively considering extending their operating hours, potentially culminating in full 24/7 operation (e.g., Bank of England).

2. Bank for International Settlements, "Operational and technical considerations for extending and aligning payment system operating hours for cross-border payments: An analytical framework," Committee on Payments and Market Infrastructures technical report, February 2023.

Another area in which central banks have been active is the provision of their own real-time payments system settlement services, notably the ECB and the Reserve Bank of Australia (i.e., central bank operated real time gross systems as described above). There are still relatively few such services, possibly because of the investment required. However, this may become more common in the future.

In a number of other countries, where the central bank operates the real-time payment clearing system, including settlement, a different approach to managing settlement risk is being adopted. In these markets, the central banks have complete visibility of banks' activities, exposures and reserves and they are not requiring the commercial banks to set aside dedicated real-time payments liquidity. The central banks are managing risk by taking a holistic view of banks' assets / positions and setting real-time payment system limits (caps) as part of a broader oversight role. This may also become more common.

As already mentioned, higher values of real-time payments being processed and higher liquidity costs for commercial banks may also lead to greater focus on liquidity optimization. There are systems today that already have well-developed tools for ensuring that bank liquidity is managed efficiently. The Danish Express Clearing system, for example, automatically checks and adjusts bank liquidity throughout the day. There could be wider adoption of such tools. The search for increased efficiency could also lead to greater use of the hybrid strategies already described, which exchange lower liquidity requirements for increased levels of risk. It could even lead to different strategies being provided for large and small banks based on their risk appetite.
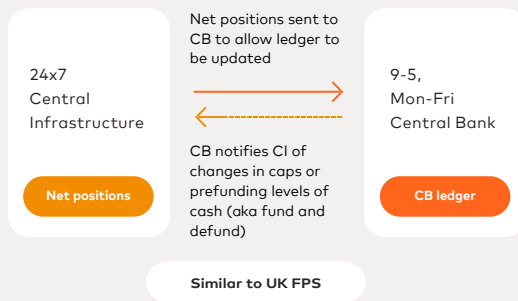
Appendix

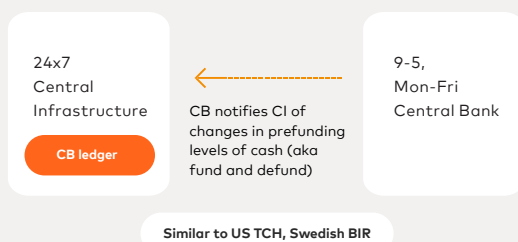## Settlement models in operation

The operation of these different models and the location of the central bank settlement ledger are illustrated in the diagram below.
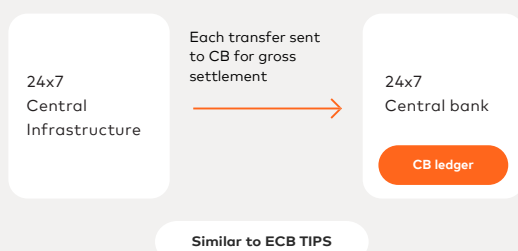
Deferred net settlement

24x7 Central Infrastructure

Net positions

Net positions sent to CB to allow ledger to be updated

→

←

CB notifies CI of changes in caps or prefunding levels of cash (aka fund and defund)

9-5, Mon-Fri Central Bank

CB ledger

Similar to UK FPS

- Based on central bank ledger service being open only during "banking hours."
- To guarantee settlement, banks set aside a dedicated amount of cash to prefund their transfers in an account in their name at the CB, OR banks sign a loss-sharing agreement, OR a hybrid.
- Central infrastructure clears messages in real time and verifies each transfer is covered by the amount of cash set aside by the payer's settlement bank OR within the cap set by the scheme where loss sharing is used.
- Periodically (settlement cycles) the central infrastructure sends positions of each bank to the CB to allow it to adjust its ledger.

ACH/processor-operated real-time gross settlement

24x7 Central Infrastructure

CB ledger

←

CB notifies CI of changes in prefunding levels of cash (aka fund and defund)

9-5, Mon-Fri Central Bank

Similar to US TCH, Swedish BIR

- Based on central bank outsourcing the CB ledger record-keeping to the central infrastructure where it would operate 24/7.
- Banks would set aside dedicated amounts at the CB to prefund their transfers in a single account for the system that holds the balances of all banks.
- Central infrastructure clears messages and settles them in real time in the CB's outsourced ledger. Before settling each transfer, it verifies the payer's settling bank has a sufficient balance in the ledger.

Central bank-operated real-time gross settlement

24x7 Central Infrastructure

Each transfer sent to CB for gross settlement

→

24x7 Central bank

CB ledger

Similar to ECB TIPS

- Based on central bank ledger service being open 24/7.
- Banks would set aside a dedicated amount of cash to prefund their transfers in an account in their name at the CB. This amount may be all or part of their cash holding.
- Central infrastructure clears messages but sends each transfer to CB for settlement in the CB ledger (aka line by line). CB performs all liquidity checking, CI role limited to message exchange.

For more information, contact us:

Richard Jones,
Director
Industry relations
richard.jones@mastercard.com

Jim Mortimer
Director
Business development
jim.mortimer@mastercard.com

Susan Hall
Vice president
Industry standards
susan.hall@mastercard.com

Designed by Mastercard Creative Studio