Balancing Liquidity and Risk in Modern Payment Systems

Use of AI-controlled dynamic periodic net settlement mechanisms in real-time payment market infrastructures

By Ainsley Ward & Neel Mehta

Overview

Optimizing the period in Deferred Net Settlement

Following the 2008 global banking crisis, it has become significantly more important to ensure that prominent Retail Payment Market Infrastructures such as national Automated Clearing Houses or Real-time Payment platforms or systems are fully collateralized and no longer dependent on the taxpayer to cover default risk. This has led to an increased variety of models and a move away from traditional batch Deferred Net Settlement systems (DNS) toward line-by-line Direct Settlement (DS) models which ensure transactions only pass through when liquidity is available to cover them

However, DS models are traditionally liquidity intensive, seldom achieving liquidity efficiency ratios above single digits. This means that participation is expensive as the liquid assets or cash on account needed to fully collateralize payments are encumbered and unavailable for investment purposes. The counterpoint to DS, DNS typically relies on a fixed period between netting cycles, usually optimized to minimize trade-off between settlement risk inherent in the market infrastructure and lower collateral requirements for participants. This optimized single period for netting leads to compromises that can work for some participants, but not all.

In this paper, we are proposing a new model where the netting period used in a DNS model is dynamic to manage liquidity efficiently based on market variables. Essentially the system will alter, monitor and optimize itself as required based on the prevailing conditions.

In this way the system can retain a higher liquidity efficiency while ensuring that risk levels remain constant regardless of transaction volume or velocity.

The Drivers for Change

As markets begin to modernize their payment or financial market infrastructure in response to the changing political environment, decreased appetite for payment risk, and the need to introduce faster or more complex payment types, there is an increasing focus on settlement models as these may introduce systemic risk. For core settlement systems and high value payments systems (HVPS), these are typically migrating to Real Time Gross Settlement (RTGS), where each transaction is settled immediately by leveraging an asset-backed liquidity pool belonging to the sender. In core settlement systems, a small degree of transactional delay is deemed reasonable to facilitate the use of Liquidity Savings Mechanisms (LSMs) which ensure that the systems are highly liquidity efficient. This fully-collateralized RTGS model is exceptionally low risk as it usually operates on a cover-all basis and ensures that any risk taken is born by the benefactor of that risk, typically the initiator of the transaction.

In the past it has been accepted that lower value bulk payment systems, such as the Automated Clearing House (ACH) model, which under PFMI rules now need to be collateralized, generally have accepted higher risk levels due to the lower importance of this traffic to economic health. Typically ACH systems deploy Deferred Net Settlement (DNS) models, where transactions are batched together and a multi-lateral netting process determines the amount to be settled between participants. This has led to a prevailing settlement model based on collateralizing the window, and a highly efficient liquidity system. A potential risk in this model is any situation where there is a net imbalance in the system during any settlement window, which leads to a participant's pledged collateral being out-of-sync with their collateral requirements; this results in inefficiencies and requires a mechanism to rapidly pledge new collateral. The traditional method of countering this is by use of ever-decreasing periods between settlement windows.

However, in modern Real-Time Payment (RTP) systems where volumes are closer to ACH, but speed of processing needs to replicate that of HVPS, there is a growing movement towards an RTGS-like model, more commonly known as Direct Settlement (DS). DS models use the clearing message as the initiator for settlement and transactions are dealt with line-by-line on a FIFO basis, unlike traditional DNS models which use separate clearing and settlement messages.

In RTP systems there is no tolerance for settlement delay due to the large volumes involved and the 24/7/365 operation of the systems. This normally precludes the use of LSMs and results in systems that are very low risk but highly inefficient in terms of liquidity use or management. This inefficiency manifests itself as direct costs for the participants as cash or pledged assets are locked up to create risk-free liquidity in the system. Where DNS models are deployed for Real-time Payments, the netting period is usually short in order to reduce the settlement risk. Netting periods typically vary from 6 seconds to 20 minutes depending on the peak transaction levels and the risk appetite of the central regulator.

"Direct Settlement uses the clearing message as the initiator for settlement and transactions are dealt with line-by-line on a FIFO basis."

Regulators continue to look at prevailing models

Despite ongoing modernization efforts in many markets and significant spend on experimental technologies such as blockchain, there is little indication that regulators and market orchestrators are looking beyond the traditional methods and mechanisms to support core settlement systems. Nearly all current, and planned, settlement systems can be categorized into a small number of models, all based on legacy thinking' to the sentence prior to the:

- **Direct Settlement (DS)** – participants prefund the system with liquidity (cash or liquid assets) against which each transaction is settled separately at the same time as clearing occurs. This model is being used in Australia's NPP and is similar to that used in a RTGS. In a DNS Liquidity needs are high and LSMs cannot be introduced as these would cause unacceptable delay.
- **Deferred Net Settlement (DNS)** – transactions are cleared immediately, but settlement occurs at the end of a predefined netting period. Settlement can be against prefunded assets or demanded at the time of settlement. Liquidity needs are relatively lower than DS but tend to be less efficient during peaks and troughs.

There are also two emerging models that are deployed in small number of markets:

- **Rule-based Deferred Net Settlement (rDNS)** – transactions are cleared immediately, but the netting process is triggered by breech of a rule-based metric. For example "total system transactional volume exceeds $XXm". Again, settlement can be against prefunded assets or demanded at the time of settlement. Liquidity needs can be lower than periodic DNS dependent on the rules specified as this is a more flexible implementation of traditional periodic DNS.
- **Hybrid Model** – transactions are cleared immediately, but netting and settlement may differ depending on the class of participant or the availability of infrastructure. The model may deploy both rule-based and periodic DNS, optimised for a particular metric such as lowest liquidity use or lowest settlement risk.

However, all of these models have limitations in that they are generally designed around static parameters and optimized for either minimizing credit and liquidity risk, or liquidity efficiency. Given that a settlement system does not function in a static environment, but an environment with strong peaks and troughs, there is growing evidence that dynamic flexibility should be the focus rather than boxing settlement into a single model.

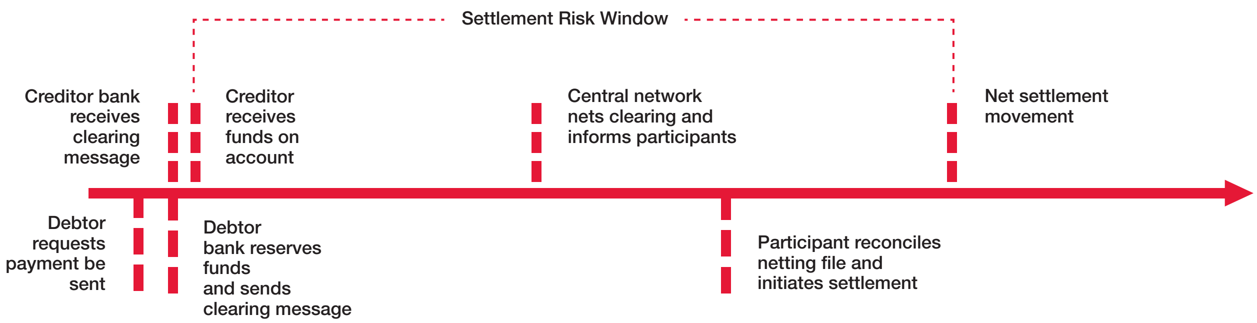

Figure 1: The collateralized risk in Net Settlement systems is the window between clearing and settlement

Creating an Intelligent Settlement Model

In a system where transaction volumes are not uniformly distributed, there are large parts of the payment processing cycle where a single optimized netting period is ineffective. Having multiple shorter periods designed to lower settlement risk during peak volumes may be too conservative for low velocity times as the netting frequency would be ineffective in delivering efficient nets. Similarly, a longer net period will suit lower volume times, but may introduce greater risk at peak volumes. At peak optimization a netted system will naturally balance, tending to near-zero values for each participant and meaning that collateral needs are at their lowest. When volumes passing through the system are not adequate to support this, or the balance leans towards a particular participant within a netting cycle, there is an opportunity to revisit the optimal netting period and dynamically adjust it to rebalance the system.

In an intelligent settlement model, it should be possible to vary the DNS period depending on key values such as transaction velocity, inbound vs outbound, or overall value flow imbalance for participants. This would allow longer or shorter netting periods depending on the prevailing market conditions – a situational application of the balance between risk and liquidity efficiency.

To achieve this, there needs to be three major components:

- **Pre-determined day plans** with forecasted volumes based on historic data. These plans could be published in advance to allow Cash Management teams to manage Bank-side liquidity effectively
- **Re-active Artificial Intelligence** to ensure that if the forecasted day plan is not matching actual trends, the netting frequencies can be automatically amended to effectively balance the trade-off between settlement risk and liquidity efficiency
- **Notification system** to ensure that regulators and participants are kept appraised of changes to the netting period, particularly during traditional lower volume times

This model will allow the period between netting cycles (P) to flow from a pre-determined maximum value (Pmax) and zero, at which point the settlement system will operate identically to the traditional DS model. The flow between different values of P can be based on pre-defined increments but could also leverage a completely sliding scale driven by the AI.

The determinant factors in deciding the value of P are the ratio of settlement risk (S) and liquidity efficiency (E) with an upper limit of Pmax.

Therefore $Pcurrent = \frac{Scurrent}{Ecurrent}$ where $\lim_{P \to \infty} P = Pmax$

This allows a settlement system that becomes optimized to the current traffic while helping to ensure low settlement risk and maximized liquidity efficiency.

Visualising this model creates an idea of how it might work in practice. In this example, a real-time retail system is available 24/7/365. It is used heavily during lunch breaks and evenings, the latter of which coincides with the time that the Central Bank’s collateral provisioning system is closed for end-of-cycle processing.

> **“In an intelligent settlement model, it should therefore be possible to vary the DNS period depending on transaction velocity or overall value flow. This would allow short netting periods during high volumes and longer netting periods when volumes are lower – a situational application of the balance between risk and liquidity efficiency.”**

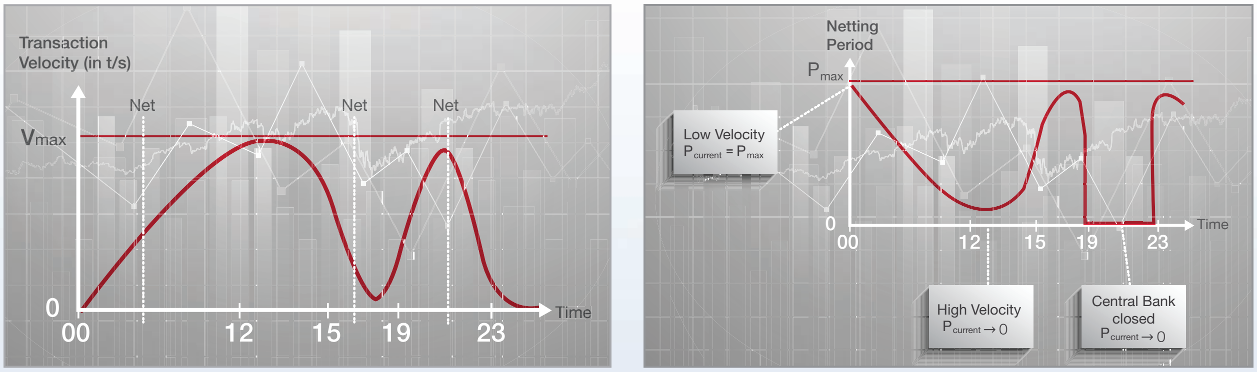

Figure 2: Transaction flow in a traditional system

In the illustrative model above, this system is using three netting periods during the day (06:30; 16:00; 21:00). This means that the settlement risk building up at peak hours is significant and liquidity costs to the market for a ‘cover all’ collateralization model are close to those of a DS model. In this situation the market regulator would typically push for DS to be implemented, increasing liquidity costs, or move to shorter netting periods, which typically decreases liquidity efficiency.

Moving to a model with forecasted volumes based on historical data where the netting period can be varied dynamically would allow the highest efficiency net to be applied when needed. It will also allow risk to be mitigated when the Central Bank is closed by moving to a zero period, or DS-equivalent, which prevents any participant overstepping its liquidity threshold.

The following represents a system where velocity is the driving factor for optimal netting.

Further supplementing this model with additional rules that prevent the market from being impacted by significant volume changes by a single or limited number of participants will help with balancing settlement in the most efficient manner. In short, using AI technology would make net settlement systems more responsive to dynamic market conditions and better minimize systemic risk.

“What this allows is a settlement system that becomes optimized to the current traffic while understanding the need for low settlement risk and maximized liquidity efficiency.”

Supporting Participant-side Collateral Management

One of the inherent complexities of a reactive settlement model could be that it makes it difficult for Participant Bank cash and collateral management teams to work within such a dynamic model.

It may be simpler, although less efficient, for participants to work within a static DS or low-P DNS model. However, if the Participants are supplied with multi-day forecasting based on historical data and current trend analytics then their ability to manage within the model should become easier.

Intraday liquidity management has proven to be a primary key to the successful implementation of retail and wholesale real-time payment systems (RTGS and RTR) around the world, yet the tools that have been developed to support participants have remain relatively poor. Many banks are investing heavily in cash management and enterprise liquidity tools in lieu of central market services, with the downside that forecasting is based purely on their own historical data rather than that available across the market. It is clear that with a move to more dynamic retail settlement systems, there is a need to supply whole market forecasts which can be tailored to each participant and compared and contrasted with the participant's own data.

While forecast data would be made available from the central settlement system through API's, it is envisaged that the central system would also need to provide app or dashboard-based tools similar to those used for accessing weather forecasts on today's smartphones. (See figure 4 at right)

The tool could give a current status "Netting Period is currently 5 minutes"; current trend "The Netting Period is likely to decrease"; daily forecast in detail; and medium-term forecasts built using data analytics. This trend data would become an important tool in ensuring participants can maintain the necessary liquidity to reduce the chances of settlement system gridlock.

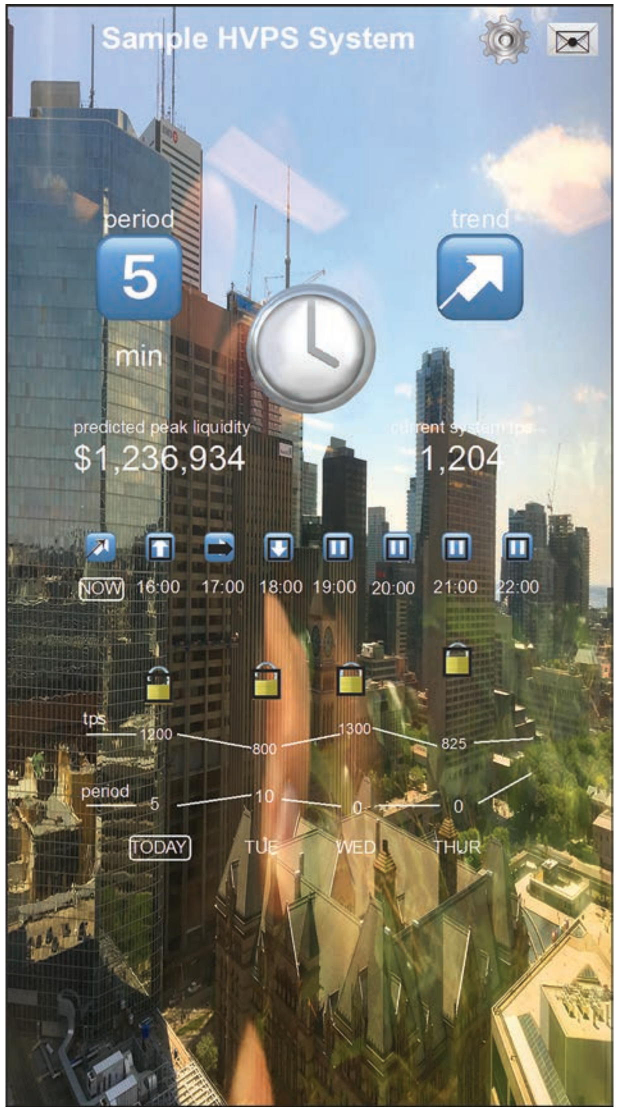

Figure 4 – Sample of Settlement / Liquidity Dashboard App

Summary

Due to the inherent risk and conservative behavior, adoption of new settlement methods by Central Banks and Regulators for their Market Infrastructures has always been cautiously slow. This often leads to solutions that, while considering of liquidity costs, are skewed out of necessity toward minimizing risk without balancing the need for minimizing collateral cost at the Participant's side. As monetary policies tighten across many jurisdictions, banks and financial participants will find the costs of collateral become increasingly punitive, making cost of direct participation in payments systems harder to justify for lower volume players. However, the increasing sophistication and maturity of Artificial Intelligence-based forecasting and data analytics systems suggest that we are already capable of delivering solutions for settlement models that would bring risk management and liquidity efficiency benefits for both Participants and Central Banks.

Moving to an AI-driven variable netting period model would enable maximum liquidity efficiency to be extracted from a real-time payment settlement system. Optimizing liquidity usage reduces the cost of participation in real-time payment settlement systems, which in turn would allow a greater number of organizations to participate directly and therefore reduce inherent costs associated with indirect connection models. With the risk balanced and costs minimized, real-time payments can be more effectively leveraged to drive the growth of electronic payments and reduce reliance on cash.

Introducing new technology and a new way of thinking into the traditional environment of payment settlement will bring greater efficiencies and lay the foundation for broader competition, which is to the benefit of all payment system users.

cgi.com