CORE
Metadata, citation and similar papers at core.ac.uk
Provided by AIS Electronic Library (AISeL)

Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2017 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

Summer 7-19-2017

How to Enable Future Faster Payments? An Evaluation of a Hybrid Payments Settlement Mechanism

Zhiling GUO
Singapore Management University, ZHILINGGUO@smu.edu.sg

Yuanzhi Huang
Singapore Management University, yzhuang@smu.edu.sg

Follow this and additional works at: http://aisel.aisnet.org/pacis2017

Recommended Citation
GUO, Zhiling and Huang, Yuanzhi, "How to Enable Future Faster Payments? An Evaluation of a Hybrid Payments Settlement Mechanism" (2017). *PACIS 2017 Proceedings*. 231.
http://aisel.aisnet.org/pacis2017/231

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 2017 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

How to Enable Future Faster Payments? An Evaluation of a Hybrid Payments Settlement Mechanism

Completed Research Paper

Zhiling Guo*
School of Information Systems
Singapore Management University
Singapore 178902
zhilingguo@smu.edu.sg

Yuanzhi Huang
School of Information Systems
Singapore Management University
Singapore 178902
yzhuang@smu.edu.sg

Abstract

In the era of Fintech innovation and e-commerce, faster settlement of massive retail transactions is crucial for business growth and financial system stability. However, speeding up payments settlement can create periodic liquidity shortfalls to banks which would incur high cost of funds in the settlement process. We propose a new hybrid settlement mechanism design that integrates features of real-time gross settlement, deferred net settlement, and central queue management structure. The hybrid mechanism is managed by an intermediary and is particularly suitable to settle large volume of small-value retail payments. We evaluate the mechanism using computer experiments and simulation. We find that central-queue netting is an effective means to achieve high system performance. Our results also show that the intermediary plays an important role in coordinating multilateral central-queue netting and supplying liquidity as needed to banks. We offer some policy insights into future faster payments settlement mechanism design and implementation.

Keywords: Liquidity, mechanism design, payment system, queue management, simulation

* The author names are listed in alphabetical order.

Introduction

The emerging trend in the retail payments market over the past decade has been towards processing an increasingly high volume of relatively low-value payments, in forms such as checks, Giro credit transfers, automated clearing house transactions and electronic funds transfers at the point of sale. The rising demand in e-commerce and electronic payments requires the financial services industry to develop innovative and sustainable retail payments solutions supported by new technologies (CPSS 2012a). The whole financial services industry is now looking for new ways to speed up and streamline payments processing. At present, many countries have been working on modernizing the payment infrastructure and developing real-time payment capabilities. For example, VocaLink, the UK's leading payments infrastructure company, has delivered the largest real-time payment system in the world – the Faster Payments Service. This has been supporting UK businesses and consumers with real-time bank account to account transfers on digital platform, who benefit from faster cash flow, immediate funds access, and better customer experiences. For traditional payment system and network operators as well as individual retail banks, failure to provide much-needed faster payments services could make them vulnerable in industry competition.

The movement to faster (or even real-time) payments poses serious technical and regulatory challenges in the global transformation of banking infrastructure. Retail banks would bear deferment risks related to unsettled payments if they still adopt the traditional batched processing of settlement. Even though a few faster payment schemes have been enabled by the introduction of ISO 20022 and XML schema, it is yet a reality in most countries. In the battle for digital dominance (especially in new areas such as mobile business), current solutions still depend on the banks' traditional infrastructure to move the money and finish retail transactions. Most existing payment methods are costly, inefficient and lack of transparency desired by both consumers and regulators. How should innovative faster payments services be designed? How to upgrade and leverage the existing payments infrastructure to speed up payments processing? Most important of all, how to effectively handle massive small-value retail payments settlement without incurring significant settlement costs or risks? All these pose key settlement mechanism design challenges and open up tremendous opportunities in the next wave of Fintech innovations in the payments industry.

In handling payments, clearing refers to the promise or commitment to settle a transaction, and settlement refers to the transfer of funds between the payer's bank and the payee's bank. These functions are normally carried out by a credible settlement intermediary (i.e. a central bank or a clearing house). Interbank funds transfer systems can be classified in several ways. In common, there are two types of settlement mechanisms: *deferred net settlement* (DNS) that has been used with small-value payments and *real-time gross settlement* (RTGS) that focuses on large-value payments (CPSS 2012b). The typical range of US dollar values for payments settled via DNS is on the order of $2,000 to $5,000, while for RTGS, it is about $2 million to $5 million. Historically, DNS mechanisms have been used to handle the relatively small-value retail payments in many clearinghouses, where settlement of funds transfers (these are mainly retail transactions) occur periodically on either a bilateral or multilateral net basis. Net settlement (netting) is an efficient approach to reduce the liquidity demands of a payments network. Under DNS, however, delays in settlements are undesirable and create financial vulnerabilities due to accumulated unsettled funds. To mitigate the risks involved in large-value time-critical transactions between commercial banks, central banks have implemented the RTGS mechanism, where settlement of funds may occur near continuously on a transaction-by-transaction basis.

More recently, leading clearing houses in a few developed countries have implemented new platforms to support faster retail payments. Bankgirot, the Sweden clearing house for retail payments, launched a new payment system called "Payments in Real Time" in 2012 to support mobile payments. While interbank transfers normally take one or several days, the new system enables instant funds transfers between multiple bank accounts. Although RTGS has been recognized as attractive, faster settlement of funds is uneasy. Infrastructure costs, operational risks, liquidity management concerns, banks' incentives, social and business benefits all require careful consideration. So far various designs exist in retail payments industry (McIntosh et al. 2014), based on which we believe a hybrid mechanism that combines RTGS and DNS would be the most promising choice.

In this research paper, we propose a new faster payments settlement design by integrating the unique features of various traditional payment systems solutions. We develop an innovative hybrid payments settlement mechanism which enables interbank funds transfer via (near) real-time settlement, liquidity-saving queue management (via settlement prioritization), and central netting functions under a unified framework, which is centrally operated by an intermediary who is responsible for payments network management. We show the performance improvement of such a mechanism over traditional DNS that is most commonly used in the retail industry. We also evaluate the positive impacts of central queue management (to enable periodic netting) on settlement cost minimization. This is conducted via simulation and computer experimentation.

Literature Review

Generally, payment systems can be classified as large-value payment systems (LVPS), such as Fedwire Funds and CHIPS in the US, and retail payment systems, such as Automated Clearing House FedACH and EPN. LVPS have been thoroughly studied in the literature, and widely adopted by monetary authorities such as the European Central Bank, Bank of Japan, and Bank of England (Norman 2010). Economic models and theoretical frameworks are often envisaged for evaluation of financial policies imposed on LVPS. For an overview of LVPS in several countries/regions, we can refer to Nakajima (2012) and Tompkins and Olivares (2016).

Most LVPS has RTGS features, which promote faster settlement but increase banks' liquidity risks. Substantial risks can spread within and across networks (Ferrara et al. 2016). To reduce the risk level, the central banks' LVPS in some developed countries incorporates a liquidity-saving mechanism (LSM), a theoretical evaluation of which is done in Willison (2004) and Martin and McAndrews (2008) based on economic models and Nash equilibrium. LSM enables banks to defer settlement until an appropriate time in the future, which would relieve their funds pressure. In this case, banks can choose between RTGS and deferred settlement via LSM, which is what we also consider in our paper. However, there are no specific LSM described in the above theoretical studies. This is our focus in this research. In particular, our new mechanism is suited for retail payment systems instead of LVPS studied in the literature. It also has enhanced functions on both internal queue and central queue clearing, which is used to speed up deferred settlement in our case. In addition, our proposed system is operated by an intermediary who can provide liquidity as needed to selected banks and help increase the network settlement performance. Thus, we treat RTGS as a function of the mechanism instead of an alternative mechanism, to timely handle high-priority payments without incurring settlement deferment.

Successful implementation of a new payment system is often difficult in practice (Leonardi et al. 2016; Venkatesh et al. 2016). It depends on mechanism designers' efforts to accommodate existing infrastructure/investments and financial regulations. Hence, when we design and introduce the new mechanism for multichannel retail payments, on the one hand, we make it as flexible as possible so that its base design can be easily adapted to meet unique requirements for different systems or networks. On the other hand, we add a number of optional new features such as a pricing model and a set of settlement prioritization rules to the mechanism we propose. Differential pricing enables banks to offer a range of financial services to their clients, which is justified in Clemons et al. (2002). We take account of this so as to derive a mechanism that aims to minimize the settlement costs incurred from interbank funds transfer. Meanwhile, banks could face settlement/default risks propagating across the network (Hu et al. 2012). A suitable settlement prioritization function would enable banks to choose between RTGS and deferred settlement, thus to influence settlement timing and better control risk themselves. This also makes queue management more flexible because banks do not have to follow the classical first in, first out rule. Banks can then settle time-critical transactions in real-time and defer non-urgent transactions to a later time. These optional features make our novel mechanism design a unique and distinguishable solution for future faster retail payments.

Classical empirical simulation tools, such as BoF-PSS2 (Koponen and Soramäki 1998, Leinonen and Soramäki 2003), has been used for a wide range of central bank research on LVPS. Hellqvist and Laine (2012) and Laine (2015) present a collection of such simulation-based studies from different areas of central bank research on RTGS and financial risk control. Johnson et al. (2004) and Afonso and Shin (2011) have done simulation experiments with real transaction data from Fedwire Funds. This has been a sophisticated and sensible method for empirical evaluation of mechanism designs (Guo et al. 2012), though we focus on retail payment system in this paper. To assess the benefits of our design, we also refer to the agent-based model in Galbiati and Soramäki (2008). Such kind of a model enables us to investigate the individual banks' behaviors within an assumed network.

An Innovative Settlement Mechanism

How the Mechanism Works

Bank clients can be classified based on abundant transaction record and predictive analytics (Bapna et al., 2011, Martens et al., 2016). Likewise, it makes sense to distinguish the heterogeneous behaviors of banks that act on behalf of their clients. We assume a domestic retail payments network that includes a number of settlement banks. The banks conduct interbank funds transfer with each other in real time. We then assume a private sector system operated by an intermediary (i.e. clearing house). Each bank maintains a settlement account in the system for processing payments.

Figure 1 illustrates the process flow when a payment request from a bank's client arrives at the bank and when a real-time settlement from another bank arrives. As shown in the Figure, generally our mechanism involves three phases. To handle masses of funds transfer requests (payments), a bank honors RTGS when it has sufficient funds. The beneficiary bank immediately receives the funds from all the real-time settlements. When RTGS is impossible (due to lack of funds), the funds transfer requests would be queued in the bank's book (the bank's internal queue). At an undesignated future time, once the bank has recovered funds (e.g., receiving real-time payments from other banks), deferred settlement of those queued requests can occur (internal-queue clearing). Traditional RTGS and DNS do not have this queue management function we introduce. At some predefined and designated time, all banks submit their internal queues to the intermediary for a multi-lateral central-queue netting, which is a DNS optimization function for bank reconciliation.

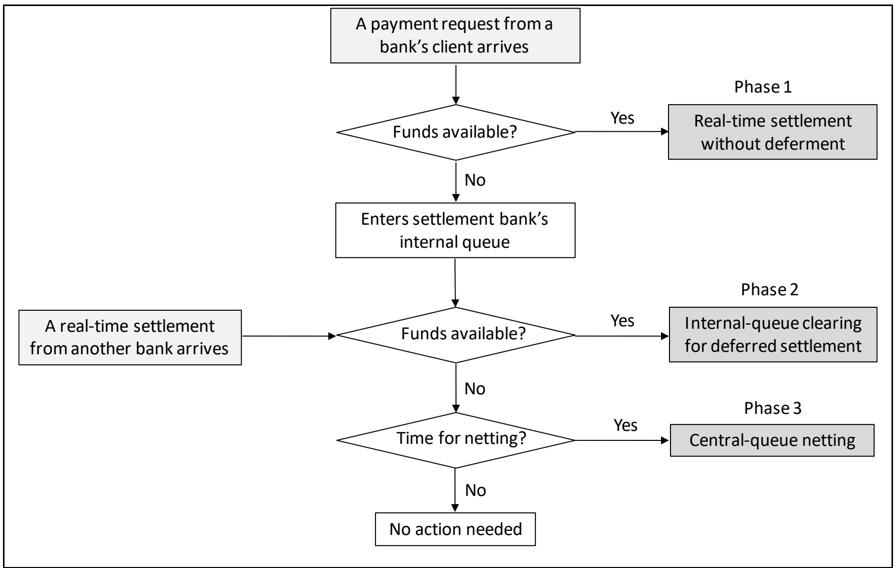

Figure 1. The Three-Phase Settlement Framework

Key Design Elements

In terms of the time horizon, the payment system is defined as the number of settlement sessions. The length of a settlement session can be defined in terms of how often the central-queue netting is scheduled. For example, if we define each session as one minute, then we can use 60 sessions to model one-hour operation. If we define each session as one hour, then we can use 24 sessions to model a settlement network's continuous daily operation.

In terms of the network structure, a payments network consists of a number of banks trading with each other. We use two parameters to model the network:

- **Number of participating banks.** Banks are divided into two categories: active and inactive, where active banks have more payment requests to process than inactive banks.
- **Bank-level payment density.** This refers to the probability that a bank submits one or more requests within a settlement session, which is larger for each active bank.

Actual retail payments settlements in most countries only occur several times a day, based on DNS. When we introduce the hybrid settlement mechanism, it is believed that there would be an increase in the liquidity demands and the cost of funds to cover faster settlements (and possible overdrafts). In our mechanism, two sources of funds would affect the overall system liquidity.

- **Initial balance requirements for individual bank accounts.** Before the start of the first settlement session, individual banks are required to maintain a certain level of initial funds to their settlement accounts. In the case that the bank's account balance turns negative, the bank would start to incur overdraft penalties.
- **Reserve funds from the intermediary.** We advise the intermediary to hold some reserve funds in the payment system. When banks experience illiquidity on occasion, our mechanism would allow them to borrow funds from the intermediary at a cost.

The hybrid settlement mechanism we propose combines RTGS/DNS features as well as new functions such as settlement prioritization and queue management, which makes it different from other mechanisms discussed in the LVPS literature.

- **Internal queue management.** Depending on the urgency and liquidity requirement, individual banks decide real-time settlements and clear their own internal queues when funds are available.
- **Central queue management**. At the end of each session, the intermediary will merge all internal queues temporarily. The central-queue netting enables multilateral offsetting to clear some deferred payments from the banks' internal queues. The settlement decision is made by the central-queue intermediary, who should act on behalf of all individual banks.

We focus on evaluating three options of the central queue management structure embedded in the mechanism design, in order to quantify the benefits brought by the intermediary's participation. Because bank-level transaction data are highly sensitive and often include too much background noise, real-world data are not suitable for our computer experiment. We thus construct and use different simulated datasets (30 in each case) to evaluate the three design alternatives (M1, M2, M3). Table 1 shows the various parameter choices we assume.

Treatments/Parameters	**Values**
Central queue management *M*	**M1**: Central queue with liquidity provision **M2**: Central queue without liquidity provision **M3**: No central queue
Number of settlement sessions *T*	60
Number of banks *I*	{3 active banks, 3 inactive banks}
Initial bank account balance *B*	{500K for active banks, 50K for inactive banks}
Initial central-queue balance *C*	50K
Payment density *p*	{50% for active banks, 25% for inactive banks}
Cost coefficient α	{0, 0.05, 0.3, 0.95, 1}
Coefficient associated ratio r	{20%, 60%, 20%}
Cost of funds per unit of overdraft β	1
Cost of funds per unit of loan γ	1.05

Table 1. Conditions and Treatments of the Computer Experiment

We assume that there are six banks in a default payments network. To take the heterogeneous behaviors and characteristics of individual banks into account, we assume three of them are more active than the other three. To match their different settlement demands, those active banks should maintain higher initial account balance (500K for each) than inactive banks (50K for each). We assume the central queue reserve funds of 50K in total, which should be sufficient for supporting the circulation of funds in the network. We focus on a simulation span of 60 settlement sessions.

In each session, a number of random payment requests can be submitted by individual banks. We operationalize the banks' settlement service demands in terms of the bank-level payment density, which refers to a fixed probability that a bank would submit one or more requests in one session (50% for all active banks and 25% for inactive ones). In the case that a bank has submitted, the number (value) of the corresponding requests are drawn from a Uniform (Pareto) distribution in random. Besides, each request is associated with a cost coefficient, the unit cost that would be incurred from deferring the settlement of this request to the next session (the total cost is the cost coefficient times the transaction value). This indicates the settlement urgency of each request. We assign small cost coefficients 0 (0.05) to those whose value is over 50K (between 10K and 50K), as we give priority to small-value retail transactions. Faster settlement of large-value payments would consume a high level of liquidity and thus are treated to be non-urgent in our case study. With probabilities 20%, 60%, and 20%, we then randomly assign the cost coefficients of 0.3, 0.95 and 1 to all requests below or equal to 10K. Higher cost coefficients for payments of lower value (≤10K) indicate our mechanism's preference to settle as many small-value retail payments as possible.

Under the general hybrid payments settlement mechanism we describe in Figure 1 (M1), banks can borrow reserve funds from the central queue during multilateral netting. An alternative design is that banks are not allowed to borrow (M2). Under the third mechanism without the DNS function (M3), the central queue would not even be allowed to exist such that the central netting function is disabled too. M3 acts as the benchmark in our mechanism design.

In the event that an overdraft occurs, the per unit overdraft cost is 1, which must be no higher than the maximal cost coefficient. This implies that all payments with cost coefficient 1 can be settled in real time (upon arrival), regardless of the bank' current account balance. In the case of M1, per unit (central-queue) loan cost is 1.05. This is higher than the overdraft cost such that the bank would settle their loans in time and thus central queue is able to preserve sufficient funds.

Speeding up settlement without incurring high costs of funds is the key objective of our mechanism design. We define performance measures as follows:

- **Settlement ratio (SR)**. Calculated at the end of the final (i.e. 60th) session, this is the ratio of the settled payment requests to the total number of requests.
- **Normalized deferment index (NDI).** Settlement deferment time is the difference between the time a payment request is submitted and the time it is settled (i.e. measured by the number of sessions in our simulation). Taken the whole network into account, NDI is a ratio between 0 and 100%. Its numerator is the total of the deferment time incurred (with the transaction value as the weight), under the settlement mechanism design M we evaluate. The denominator of NDI is the total of the deferment time incurred (with the transaction value as the weight) using the traditional end-of-all-sessions DNS as the reference. To reduce random variation, our calculation of NDI excludes transactions whose value are over 50K, the cost coefficients of which are zero.
- **Deferment cost index (DCI).** Based on the formula for NDI, we introduce the cost coefficient of each payment request as the second weight in addition to the transaction value. While DCI is similar to NDI, it estimates the overall deferment costs incurred, which is an inherent component of the total settlement cost.
- **Average funds required.** Based on the source of fund costs, we calculate the mean balance overdraft (MBO) and mean borrowed credits (MBC) per bank per session.

For a pure DNS mechanism, both NDI and DCI should be 100%; both MBO and MBC are 0. In contrast, for a pure RTGS mechanism, both NDI and DCI should be 0, the same as the MBO and MBC. Banks are required to maintain extremely high balance across all sessions, since all transactions should be real-time (default is not allowed). Our hybrid mechanism achieves a balance between these two traditional mechanisms. Our experiments will show we can achieve a small NDI and DCI ratios at the expense of not too high MBO and MBC costs, realizing near real-time settlement for most transactions in our network while effectively controlling network liquidity risk.

Results

Our computer experiment is based on 30 datasets of simulated retail payment requests (i.e. transfer of funds of a certain value from one bank to another bank in a certain session). Among these datasets, the mean number of requests is 48,700 (Min: 41,649. Max: 55,200). We plot the settlement ratio for all 30 simulated datasets under mechanisms (M1, M2, and M3 in Figure 2(a). To evaluate the treatment effect of *B*=500K/50K, we consider an extreme case when all the initial balance levels are zero (i.e. no initial funds are available). The settlement ratios in this case are shown in Figure 2(b).

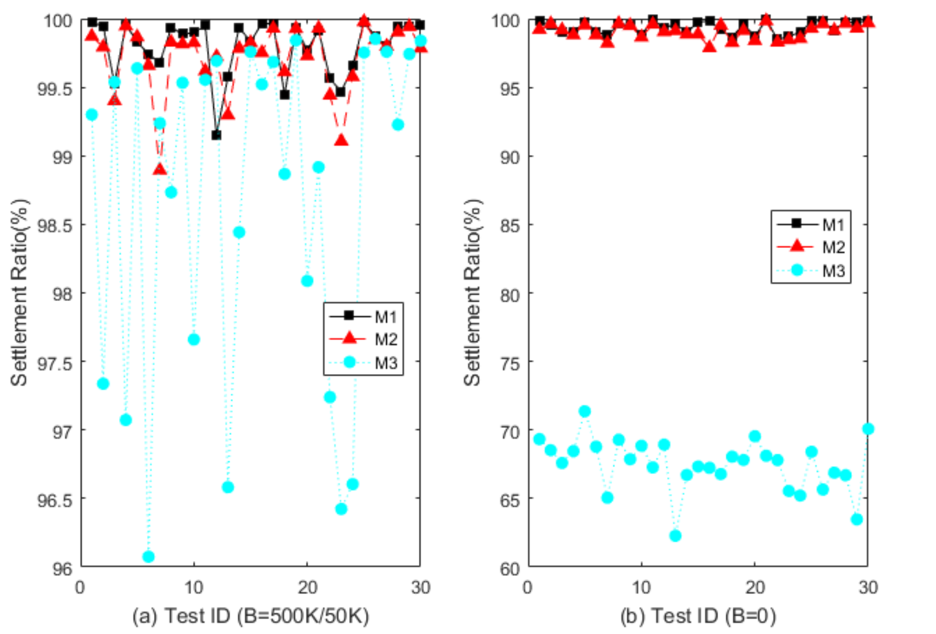

Figure 2. Overall Settlement Ratio when (a) *B*=500K/50K and (b) *B*=0

Most unsettled requests are associated with large value or low settlement priority (small cost coefficient). The fact that M3 is the worst under both *B*=500K/50K and *B*=0 scenarios shows the importance of efficient central queue management. The settlement ratio is over 95% in all cases unless there is no central queue (M3) and banks have no initial funds for their settlement accounts (*B*=0). When banks are highly liquidity-constrained (*B*=0), however, introducing the central queue function (under M1 or M2) can improve the settlement ratio from around 70% to above 98%.

When banks increase the initial balance to *B*=500K/50K, the number of real-time settlements and deferred settlements via internal-queue clearing increases. This reflects the benefits in injecting additional funds to individual banks' settlement accounts. Under M1, 5.41% of the requests are settled via central-queue netting. However, if the banks' initial balance are all zero, 42.88% of requests are settled via central queue. While those urgent requests are still settled in real-time, due to liquidity constraint, banks are more dependent on deferred settlement and central queue management. Via actively supplying liquidity, the intermediary plays an important role and contributes more to banks' faster settlement of payments.

In addition to the total number of settlements, we are also interested in the tradeoff between settlement time and costs. The mean values of NDI and DCI at both the network level and class levels are shown in Table 2. As the six banks can be divided into two classes based on their financial activities within the network, the deferment index measures at the class level help us understand the heterogeneous behaviors and differences between active banks and inactive ones.

	Overall Network		Class of Three Active Banks		Class of Three Inactive Banks	
	NDI	DCI	NDI	DCI	NDI	DCI
M1	1.08	0.39	0.75	0.29	6.60	2.10
M2	1.31	0.48	0.87	0.35	8.65	2.83
M3	4.21	1.45	3.79	1.30	11.03	4.02

Table 2. Mean Normalized Deferment Index (%) and Mean Deferment Cost Index (%) under the Three Mechanisms

Recall that the NDI of a DNS mechanism is fixed at 100%, in which case no settlements can be made in the 60 sessions. In Table 2, the mean NDI based on different datasets is below 5% under each mechanism, which indicates a dramatic improvement over DNS. The mean NDI under M1 is the lowest at 1.08%. As such, the extent of deferment under M3 is about four times as serious as that under M1. In addition, under current conditions, the three inactive banks would suffer more deferments and should have increased their initial account balance in order to ease the situation.

Effective central queue management is beneficial to banks in terms of queue clearing and relieving settlement deferments. In addition, if banks do not inject sufficient funds into their settlement accounts, the central-queue optimization/netting is even more critical for the settlement mechanism we propose (M1). Compared with the *B*=500K/50K scenario (the first column in Table 2), in the case of *B*=0, the mean NDI under M1, M2, M3 are found to be 3.62%, 4.88%, 52.91%, respectively.

We might also consider a network with 12 participating banks, for instance. In this case, six are active and the other six are inactive. This network handles twice the volume of payment requests. As a result, the mean NDI under the three mechanisms are 1.58%, 3.57%, 5.94%, with around 50-100% increase in average settlement deferment time. Nevertheless, the hybrid settlement mechanism (M1) still functions well, in which case providing liquidity to banks during netting is found to be a more effective measure (comparing M1 to M2). While we focus on the 6-bank network, this sensitivity study has demonstrated the strength and robustness of our mechanism design and results.

For the 6-bank network with initial bank account balance *B*=500K/50K, Figure 3 (Figure 4) illustrates the direct ratio and difference between NDIs (DCIs) in pairs. Because M1 achieves the highest average settlement performance, we contrast M1 to the alternative mechanisms (M2, M3) in order to quantify its benefit. As such, most of the calculated ratios should be lower than 100% whereas the differences should be negative. As we see, the occurrence and extent of settlement deferment is far more often without the central queue (M1/M3 and M1-M3). Dramatic improvements are observed when central-queue netting has been done in each session (M1/M2 and M1-M2). It is worth noting that the difference between NDIs is significant as a result of paired statistical T-test (the P-value is smaller than 0.001 in both cases). The use of T-test is justified here because of the small number of tests (i.e. 30) and mechanism comparisons based on identical datasets. With respect to DCI, effective central queue optimization/netting reduces the deferment costs banks bear such that network operating is more economical. This would benefit all parties involved in retail transactions, which is attractive banks and their clients who demand improved financial services.

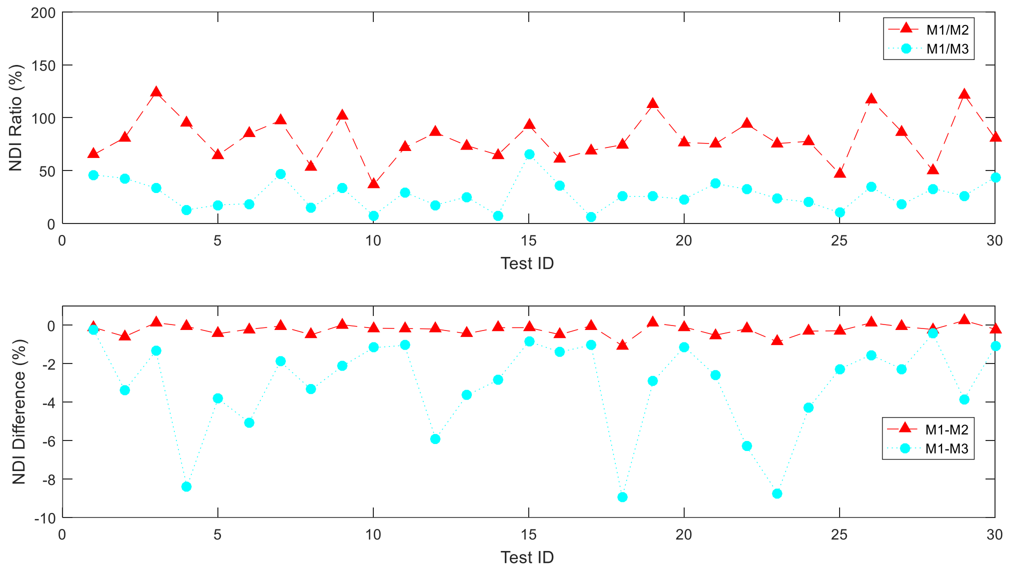

Figure 3. Ratio and Difference between NDIs in Pairs

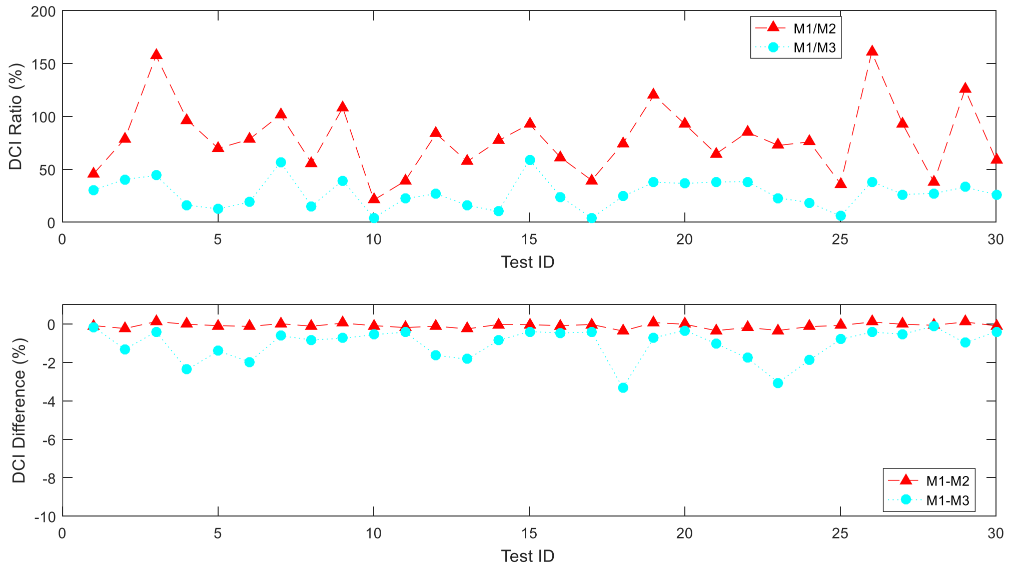

Figure 4. Ratio and Difference between DCIs in Pairs

At last, we check the mean balance overdraft (MBO) per session per bank under each mechanism. The sum of the initial funds of the six banks is 1.65M (1.5M of which is for those three active banks). On average, a bank has 275K in its settlement account at the start of the first session. We present the mean balance overdraft (MBO) and mean borrowed credits (MBC) in Table 3, which are based on the average of the 30 different tests too. Both measures lead to small values in comparison to 275K.

	Mean Balance Overdraft (MBO)	Mean Borrowed Credits (MBC)
M1	837.18	334.79
M2	191.93	0.00
M3	213.83	0.00

Table 3. Mean Balance Overdraft and Mean Borrowed Credits per Bank per Session

MBO is below 1,000 units of credit, thus the individual banks' overdraft levels are well-controlled and manageable in practice. As most banks' overdraft risks are low, the total of overdraft penalties should be small as well. As expected, the overdraft level is higher under M1, when average settlement speed is fastest. Besides, when banks are able to borrow credits from the intermediary's central queue (under M1), we also trace the loan costs of each bank. The calculated MBC is 334.79 under this scenario, which is less than half of the MBO, which would be manageable.

Based on our experimental results, we can offer a few practical insights. First, it is apparent that M2 outperforms M3 in terms of speeding up settlement at lower overdraft cost. It implies that central-queue netting is an essential design component in achieving faster settlement of payments. Second, compared with M2 and M3, M1 achieves the lowest deferment at the expense of slightly higher cost of funds. This is the inherent tradeoff between the speed of settlement and cost of funds. If speeding up settlement is a key design consideration, the mechanism designer needs to incentive the intermediary and banks to provide liquidity as needed to the system. The intermediary may charge an appropriate fee for its liquidity provision as a reward to its contribution to the network.

Conclusions

The evolution of payment settlement systems has been driven by economic advancement, changing consumer expectations and the development of new standards and financial technologies. Faster retail transactions are critical for sustainable economic advance and rapid expansion of new (e-) businesses and technologies. Smooth settlement of funds transfer requests among banks and other financial institutions is so important for constructing a stable financial network. Speeding up cash flow would enable banks and their (corporate) clients better reallocate resources (i.e. funds) and plan ahead of time, contributing to better logistics and operational management. However, under a RTGS mechanism, this creates issues such as imbalance of cash inflows/outflows and periodic bank liquidity shortfalls. To balance speed and risk, we propose an innovative hybrid payments settlement mechanism, which aims to reduce the liquidity risk and high costs of funds. Our mechanism design combines real-time settlement, DNS optimization with central-queue liquidity provision, as well as an internal queue clearing process. As such, this flexible solution could benefit both banks and their clients to achieve faster retail transactions under a relatively stable financial network.

To evaluate the proposed mechanism, we design computer experiments for quantification of the economic benefits of faster settlement (such as increased number of settlements and reduced deferment time/costs) as well as the costs of funds incurred (MBO and MBC, in addition to the individual banks' initial funds). Comparison with alternative design options (M2, M3) provides important insights regarding how the different functions of the mechanism works and how to streamline the traditional settlement process. Our results show that, in comparison to RTGS, in which case both NDI and DCI are 0, the liquidity demand under the new hybrid mechanism we assume is far lower. The network would enjoy lower settlement risk overall and be less concerned about future cash flow uncertainties. Under the mechanism M1 we recommend, both NDI (1.08%) and DCI (0.39%) are found to be close to 0. Even under the extreme scenario when the initial balance of all banks are set to zero, the two indices are still low (NDI: 3.62%. DCI: 1.55%) in relative to the DNS benchmark (NDI = DCI = 100%). This results from effective settlement prioritization and (internal and central) queue management. On the other hand, in comparison to DNS, instead of deferring all settlements till the end of the last session, the new hybrid mechanism is far superior in minimizing deferment. Though banks can incur overdraft penalties and interests from loans, these costs are found to be small. Overall, the benefits of the hybrid mechanism is overwhelming.

Amid challenge there are new business opportunities for all network participants. Currently, most intermediaries (except central banks which operate large-value payment systems) do not bear settlement risk implicated in real-time settlement and end-of-session netting. This could restrict their functions in terms of servicing banks and facilitating faster settlement. Our findings in this paper suggest that the network performance can be much improved if the intermediary plays a more active role under the hybrid payments settlement mechanism, such as to implement a robust DNS optimization solution with the liquidity provision function. Adopting such a mechanism may however require the intermediary to revise its current settlement policies and upgrade infrastructure to support contemporary financial technologies for a faster payment platform. The intermediary (either a central bank or third-party platform operator) should also coordinate the efforts among the banks and their clients, who need to see direct benefits from faster settlements. Depending on the background details of a (national or international) network, financial regulations and implementation, as well as the common patterns of retail payments data, the mechanism designers (e.g. intermediaries) can implement a variant of the new hybrid mechanism we envisage here. It would be straightforward to customize the theoretical framework described by adjusting the treatments or parameter values in Table 1. The simulator we develop for this research (and computer experiment) can also be used to evaluate real-world data for policy analysis of any settlement mechanism. It helps provide practical insights into alternative solutions of faster payments and other emerging financial services. Apart from investigating the robustness of this hybrid payments settlement mechanism under different networks and market conditions, our future research may include case studies and applications adopting the feasible variants of this new mechanism.

References

Afonso, G., and Shin, H. S. 2011. "Precautionary demand and liquidity in payment systems," *Journal of Money, Credit and Banking* (43:s2), pp. 589-619.

Bapna, R., Goes, P., Wei, K. K., and Zhang, Z. 2011. "A finite mixture logit model to segment and predict electronic payments system adoption," *Information Systems Research* (22:1), pp. 118-133.

Clemons, E. K., Hitt, L. M., Gu, B., Thatcher, M. E., and Weber, B. W. 2002. "Impacts of e-commerce and enhanced information endowments on financial services: a quantitative analysis of transparency, differential pricing, and disintermediation. *Journal of Financial Services Research* (22:1-2), pp. 73-90.

Committee on Payment and Settlement Systems (CPSS), 2012a. "Innovations in retail payments," Basel, Switzerland: Bank for International Settlement.

Committee on Payment and Settlement Systems (CPSS), 2012b. "Payment, clearing and settlement systems in the CPSS countries," Vol. 2, Basel, Switzerland: Bank for International Settlement.

Ferrara, G., Langfield, S., Liu, Z., and Ota, T. 2016. "Systemic illiquidity in the interbank network," London, UK, Bank of England.

Galbiati, M., and Soramäki, K. 2008. "An agent-based model of payment systems," London, UK: Bank of England.

Guo, Z., Koehler, G.J., and Whinston, A.B. 2012. "A Computational Analysis of Bundle Trading Markets Design for Distributed Resource Allocation," *Information Systems Research* (23:3), pp. 823 – 843.

Hellqvist, M., and Laine, T. (eds.) 2012. *Diagnostics for the Financial Markets - Computational Studies of Payment System*, Helsinki, Finland: Bank of Finland.

Hu, D., Zhao, J. L., Hua, Z., and Wong, M. C. S. 2012. "Network-based modeling and analysis of systemic risk in banking systems," *MIS Quarterly* (36:4), pp. 1269-1291.

Johnson, K., McAndrews, J. J., and Soramäki, K. 2004. "Economizing on liquidity with deferred settlement mechanisms," New York: Federal Reserve Bank of New York.

Koponen, R., and Soramäki, K. 1998. *Intraday Liquidity Needs in a Modern Interbank Payment System - A Simulation Approach*, Helsinki, Finland: Bank of Finland.

Laine, T. (ed.) 2015. *Quantitative Analysis of Financial Market Infrastructures: Further Perspectives on Financial Stability*, Helsinki, Finland: Bank of Finland.

Leinonen, H., and Soramäki, K. 2003. "Simulating interbank payments and securities settlement mechanism with the BoF-PSS2 simulator," Helsinki, Finland: Bank of Finland.

Leonardi, P. M., Bailey, D. E., Diniz, E. H., Sholler, D., and Nardi, B. A. 2016. "Multiplex appropriation in complex systems implementation: the case of Brazil's correspondent banking system," *MIS Quarterly* (40:2), pp. 461-473.

Martin, A. McAndrews, J. 2008. "Liquidity-saving mechanisms," *Journal of Monetary Economics* (55: 3), pp. 554-567.

McIntosh, G., Whisler, E., Koninckx, M., Hartley, M., and Gardiner, W. 2014. "Flavours of fast - a trip around the world in immediate payments," Jacksonville, Florida: FIS.

Martens, D., Provost, F., Clark, J., and Junqué de Fortuny, E. 2016. "Mining massive fine-grained behavior data to improve predictive analytics," *MIS Quarterly* (40:4), pp. 869-888.

Nakajima, M. 2012. "The evolution of payment systems," *The European Financial Review*, http://www.europeanfinancialreview.com/?p=2032.

Norman, B. 2010. "Liquidity saving in real-time gross settlement systems - an overview," London, UK, Bank of England.

Tompkins, M., and Olivares, A. 2016. "Clearing and settlement systems from around the world: a qualitative analysis," Ottawa, Canada, Bank of Canada.

Venkatesh, V., Bala, H., and Sambamurthy, V. 2016. "Implementation of an information and communication technology in a developing country: a multimethod longitudinal study in a bank in India," *Information Systems Research* (27:3), pp. 558-579.

Willison, M. 2004. "Real-Time Gross Settlement and hybrid payment systems: a comparison," London, UK, Bank of England.