Harry Leinonen (ed.)

Simulation studies of liquidity needs, risks and efficiency in payment networks

Proceedings from the Bank of Finland Payment and Settlement System Seminars 2005–2006

SUOMEN PANKKI · FINLANDS BANK

EUROJÄRJESTELMÄ
EUROSYSTEMET

Scientific monographs
E:39 · 2007

Harry Leinonen (ed.)

Simulation studies of liquidity needs, risks and efficiency in payment networks

Proceedings from the Bank of Finland Payment and Settlement System Seminars 2005–2006

EUROJÄRJESTELMÄ
EUROSYSTEMET

Scientific monographs
E:39 · 2007

ISBN 978-952-462-360-5
ISSN 1238-1691
(print)

ISBN 978-952-462-361-2
ISSN 1456-5951
(online)

Edita Prima Oy
Helsinki 2007

Abstract

This publication consists of nine separate studies on payment and settlement systems conducted using simulation techniques. Most have been carried out using the payment and settlement system simulator BoF-PSS2 provided by the Bank of Finland. The preliminary versions were presented at the annual simulator seminars arranged by the Bank in 2005 and 2006. The main focus of the analyses is on liquidity requirements, settlement speed, gridlock situations, gridlock resolution methods, liquidity economising, systemic risk and the impact of shocks on system performance as well as network analysis and modelling of payment systems. The studies look at systems in several countries and cover both RTGS and netting systems as well as securities settlement systems.

Keywords: simulation, payment and settlement system, payment networks, liquidity, gridlock, systemic risk, counterparty risk

JEL classification numbers: C15, C61, D53, G10, G18, G28

Tiivistelmä

Tämä julkaisu koostuu yhdeksästä erillisestä maksu- ja selvitysjärjestelmiä koskevasta tutkimuksesta, jotka on suoritettu simulointimenetelmillä. Useimmat näistä tutkimuksista on tehty käyttäen Suomen Pankin maksu- ja selvitysjärjestelmäsimulaattoria BoF-PSS2. Alustavat versiot tutkimuksista on esitelty Suomen Pankin järjestämien vuosittaisten simulaattoriseminaarien yhteydessä vuosina 2005 tai 2006. Pääpaino tutkimuksissa on ollut likviditeettitarpeiden selvittämisessä, katteensiirron nopeudessa, lukkiutumistilanteissa ja niiden avaamiseen liittyvissä metodeissa, likviditeetin käytön tehostamisessa, systeemiriskeissä sekä poikkeustilanne- ja vastapuoliriskeissä ja tämän lisäksi maksujärjestelmäverkkojen analyyseissa ja mallinnuksissa. Tutkimukset koskevat eri maissa toimivia järjestelmiä, ja niiden joukossa on RTGS-järjestelmiä ja nettoutusperiaatteella toimivia maksujärjestelmiä sekä arvopaperikauppojen selvitysjärjestelmiä.

Asiasanat: simulointi, maksu- ja selvitysjärjestelmä, maksuliikeverkot, likviditeetti, lukkiutumistilanne, systeemiriski, vastapuoliriski

JEL-luokittelu: C15, C61, D53, G10, G18, G28

Preface

Payment and settlement systems are an integral part of modern economies. With continuous progress in globalisation and rapid technological advances, payment systems are in a state of transition, and dependencies due to technology and integration are increasing continuously. There are new demands on real-time liquidity and risk management. Banks and authorities will face new types continuity requirements in a complex network of business relationships. It is important to gain a good understanding of the systems and their interdependencies in order to increase their efficiency under normal circumstances and their resilience in any abnormal situations that could arise.

Payment and settlement systems are a complicated area into which simulation techniques provide a good way to penetrate sufficiently deeply. Models can be built which closely replicate the real operating environment, and these models can be used for testing and observing scenarios not normally found in real operating environments.

The Bank of Finland has a long tradition of economic research and economic modelling, and modern payment and settlement systems have been one of the focal areas. Research using simulation models for payment systems was initiated around the time Finland joined the Economic and Monetary Union as it proved to be an excellent tool for studying changing liquidity needs and system risks under the new EMU regime. Based on positive results and feedback, the Bank of Finland decided to develop a diversified simulator designed especially for external use and international distribution, called BoF-PSS2. It was completed in spring 2004 and is available for research purposes free of charge. This service is under continuous development, with the latest additions made this year being a network analysis module and a stochastic input generator. Currently the simulator has over 60 users worldwide and on every continent. The users are mainly central banks but interest in the simulator has increased among academics and private infrastructure organisations during recent years.

Investment in the simulator and the service it offers attracted great international interest and generated a variety of research and studies by different central banks. The Bank of Finland arranged four annual international payment and settlement seminars and workshops from 2003 to 2006, and the fifth seminar is scheduled for later in 2007. The main goals of the seminars and workshops are to stimulate simulation-based payments and securities settlement research, share research results and experiences among members of the user community and receive ideas and feedback on simulator development needs. The presentations of the first two seminars were published in the first simulator publication (BoF publication E31:2005) and the presentations from the third and fourth seminar are included in this publication.

I would like to thank all the authors for their contributions to this publication, which I trust will provide a good introduction to the simulation analysis of payment and settlement systems and stimulate further research to enhance our understanding and improve the models and methodologies in the years ahead.

I would like to thank all other contributors, sponsors, commentators and users of the simulator for all the help they have provided during different periods of work on and with the simulator.

The simulator has recently been enhanced with payment network analysis features, for which I would like to thank the National Infrastructure Simulation and Analysis Center (NISAC, a programme of the US Department of Homeland Security's Infrastructure Protection/Risk Management Division comprised of a core partnership of Sandia National Laboratories and Los Alamos National Laboratory) for providing parts of Loki for use by the BoF-PSS2. Loki is a collaborative software project containing core functionalities for the creation and analysis of networks and for the development of a wide array of application-specific agent-based models. I would like to especially thank Walter E Beyeler and Robert J Glass at Sandia National Laboratories and Kimmo Soramäki from Helsinki University of Technology for their help on integrating Loki functionalities with those of BoF-PSS2.

At the same time I want to extend an acknowledgement to Ville Ruoppi at MSG Software for continuous involvement in the IT developments of the simulator. A detailed list of acknowledgements can be found on the simulator website and in the user manual.

For the finalisation of the publication we are indebted to Päivi Nietosvaara for text editing and Teresa Magi for printing administration. We are also indebted to the editorial board of the publication, which consists of Ari Hyytinen, Harry Leinonen and Jouko Vilmunen.

I hope users of the simulator will continue to be active and that the simulator will attract new users and sponsors. It is a great pleasure for me to present, via this second simulation-related publication, the fruits of this continuing productive cooperation between central banks.

Helsinki, May 2007
Matti Louekoski

Contents

Chapter 1

Introduction

Harry Leinonen

1 Introduction

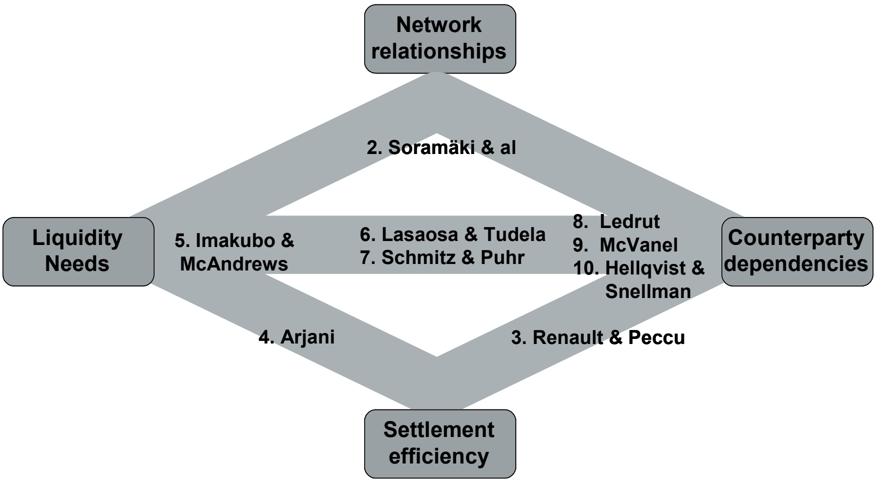
The last decade have seen a growing interest among central banks in deepening the analysis made in oversight assessments. Different kinds of stress-testing scenarios and tools have emerged in order to study concretely 'what if'-type failure and crisis situations. Among the factors behind the growing interest in this area are international cooperation and globalisation in general, technological advances and a growing dependency on payment systems and their important relationships to each other. Payment systems have become large networks or popularly 'webs' of relationships. Network analysis has therefore become a growing type of payment research. We seem to be on the verge of a new era of truly global payment and settlement systems operated completely through network-based interfaces and connections. The key official and international bodies for oversight and supervisory cooperation – the payment system committee of G10 central banks (CPSS), the Payment and Settlement System Committee of the Eurosystem (PSSC) and the committees and other groups of the European Commission – will need to increase and deepen their cooperation in order to fulfil the task of overseeing and supervising the web of international payment systems.

Payment system simulations started to attract interest in the 1990s and were to a large extent used to discover how new settlement conventions would affect the participants and the overall system and to find out the hidden credit and liquidity risks, especially in cross-border systems. These studies viewed the payment systems mostly as static machineries for booking transactions. Currently there is an emphasis on trying to better understand the network design of payment systems and the impact of the behaviour of individual participants or a group of participants. Customers and banks are particularly likely to react to external stimuli and change their payment behaviour in crisis situations. Agent-based simulation models are one way to try to catch the behavioural patterns in the payment infrastructure. More research is clearly needed in order to find out more about agent reactions in changing circumstances.

The Bank of Finland has hosted annual simulator seminars since 2003. This is the second publication of studies presented at these seminars and mainly consists of the outcome of seminars held in 2005 and 2006. It is interesting to view development over the years and see how analysis is becoming deeper and more profound step by step. Counterparty risks and dependencies in payment systems are of continuous concern to central banks and stress-testing systems are becoming a common tool for overseers. Algorithms for real-time settlement efficiency have improved over the years and become standardised features of RTGS systems. Network analysis is a new expanded area, which tries to describe the network structure of payment systems, for example the strong and weak dependencies in the system. The aim is at some point to be able to describe the behavioural patterns in payment networks.

The articles of this publication cover four main topics or views of payment and settlement systems as depicted in Figure 1.1.

Figure 1.1 **The main topics of the articles**



This publication starts with chapters on network analysis of payment systems, continues with articles about liquidity issues and ends with a smooth gradual transition to risk issues and different kinds of failure scenario in particular. Each chapter provides an individual stand-alone analysis, but some clearly build on earlier analyses. Each chapter is contributed by named authors.

Chapter 2 (Soramäki, Beyeler, Bech and Glass) presents new directions for simulation research in interbank payment systems that integrates network topology, network dynamics and agent-based modelling of bank behaviour. In the process it also reviews literature in the field and presents applications of the ideas presented. While the focus of the article is on systemic risk in interbank payment systems, the concepts and models presented are applicable for addressing questions related to other payment systems and topics such as liquidity flow efficiency as well.

Chapter 3 (Renault and Pecceu) studies the effects of queuing rules and optimisation algorithms for queued payments and especially FIFO (first-in first-out) rules compared to other processing orders. The authors aim to quantify to what extent non-FIFO optimisation algorithms can be more efficient than FIFO algorithms based on the algorithms used in the French large-value payment system PNS and in the planned TARGET2 system. Finally, the impact of the different optimisation algorithms is also investigated by simulating the complete PNS system using real data. In the context of a liquidity crisis created by a technical default of the largest participant of the system, the use of some non-FIFO algorithms is shown to reduce the number of rejected payments at the end of the day.

Chapter 4 (Arjani) explores a fundamental trade-off that occurs between settlement delay and intraday liquidity in the daily operation of large-value payment systems (LVPS), with specific application to Canada's Large Value Transfer System (LVTS). Intraday liquidity and settlement delay can be costly for LVPS participants, and improvements in the trade-off are desirable. The analysis shows that increased use of the LVTS central queue (which contains a complex queue-release algorithm) reduces settlement delay associated with each level of intraday liquidity considered, relative to a standard queuing arrangement.

Chapter 5 (Imakubo and McAndrews) describes the changes in liquidity characteristics due to the implementation of the next-generation RTGS of the BOJ-NET Funds Transfer System in Japan. Under the project, the new system will have liquidity-saving features and will incorporate large-value payments that are currently handled by two private-sector designated-time net settlement systems, the Foreign Exchange Yen Clearing System and the Zengin System. The authors analyse the characteristics of optimal funding levels under the new features using simulation analysis and find that optimal funding levels can be described with the total balances in the system, the distribution of the total balances across participants, and the timing of funding.

Chapter 6 (Lasaosa and Tudela) quantifies the tiering effects within the UK large-value payment system (CHAPS) by analysing node, credit and liquidity risks for different tiering scenarios. The results show that node risk would rise substantially in what is already a highly concentrated system. As for credit risk, the size of intraday exposures compared with settlement banks' capital is very small and therefore the likelihood of contagion is remote. The increase in credit risk brought to the system by settlement banks leaving CHAPS bears little relationship to the values settled by each individual bank.

Increasing the degree of tiering in CHAPS would lead to substantial liquidity savings within CHAPS.

Chapter 7 (Schmitz and Puhr) presents a statistical analysis of liquidity, risk concentration and network structure in the Austrian Real Time Interbank Settlement system and quantifies the contagion effect of an operational incident at one of participants' sites on other participants. The main results are that in general the value and number of payments received and submitted were quite concentrated among the top three banks and that the contagion effect in the payment system was substantial.

Chapter 8 (Ledrut) assesses the effect of counterparties' reaction to an operational failure at one of the biggest participants in the Dutch interbank payment system in which counterparties react according to two basic rules: they stop sending payments to the stricken bank either after some pre-determined time or after their exposure to the stricken bank reaches a certain level. Based on historic liquidity levels and payment flows, reacting seems to be more cost-effective when determined by the individual exposure of the stricken banks' counterparties than when triggered by the elapsed time after the disruption, but also depends on the degree of reciprocity among participants.

Chapter 9 (McVanel) provides an empirical analysis of participants' robustness to defaults in Canada's Large Value Transfer system (LVTS) by creating unanticipated defaults in LVTS. According to the findings all participants are able to withstand their loss allocations in all cases of large defaults created using actual LVTS data.

Chapter 10 (Hellqvist and Snellman) presents stress-testing results from HEXClear, the Finnish securities settlement system for equities, for four different scenarios: failure of the entire settlement process, failure of individual participants, failure of certain ICT connections and failure of the most common settlement algorithms. According to the results the current settlement system is a robust construction.

These research projects and studies have resulted in a deeper knowledge of payment systems and of the internal and external factors and parameters that affect them. The articles also show the learning curve brought about by international research cooperation which has helped the design of more efficient and stable operational systems. The new features tested via simulations have been implemented in several real-life systems. The Bank of Finland is grateful for the rewarding cooperation behind this kind of publication.

The Bank of Finland intends to continue with its successful research service in the form of simulator software development and support, arranging simulator seminars and stimulating simulation-based payment systems research and publications. A payment system is a network of cooperating participants and payment systems are currently forming links between each other, resulting in a multilayered network structure. Network analysis techniques will probably also become interesting new research topics and we hope that the new network analysis module in the simulator will stimulate research in this area. An interesting research topic is the use of behavioural and adaptive input models, such as agent-based models, which can be used to study behavioural and adaptive responses among agents in a payment systems. We hope to be able to develop the simulator in this area for the next larger update. There seems to be a growing interest in simulating securities settlement systems, and this is an important research area as the risks are less studied and known than those of basic payment systems. One interesting possibility would also be to expand the system participants/agents to include end-users such as companies and consumers. This would provide a possibility to simulate payment flows in the economy and study the effects of changing customer payment patterns on the banking sector and perhaps even the intriguing relationships between payment flows and monetary value. We hope that this publication will inspire new studies in this multidimensional business area.

Chapter 2

New approaches for payment system simulation research

Kimmo Soramäki – Walter Beyeler – Morten Bech – Rober Glass

The authors would like to thank Marco Galbiati (Bank of England) and Jeffrey Arnold (Federal Reserve Bank of New York) for comments and their input in the papers summarized in this article.

Soramäki: Helsinki University of Technology, Finland. Corresponding author: www.soramaki.net
Beyeler and Glass: Sandia National Laboratories, Albuquerque, NM, USA. Supported through the National Infrastructure Simulation and Analysis Center (NISAC), a program of the Department of Homeland Security's Infrastructure Protection/Risk Management Division and comprised of a core partnership of Sandia National Laboratories Los Alamos National Laboratory. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.
Bech: Federal Reserve Bank of New York, New York, NY, USA. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

2 New approaches for payment system simulation research

Abstract

This article presents new directions for simulation research in interbank payment systems that integrates network topology, network dynamics and agent-based modelling of bank behaviour. In the process it also reviews literature in the field and presents applications of the ideas presented. While the focus of the article is on systemic risk in interbank payment systems, the concepts and models presented are applicable to address questions related to other payment systems and topics such as liquidity flow efficiency as well.

2.1 Introduction

At the apex of the financial system is a network of interrelated financial markets by which domestic and international financial institutions allocate capital and manage their exposure to risk. Critical to the smooth functioning of these markets are a number of financial infrastructures that facilitate clearing and settlement. The events of 11 September 2001 underscored both the resiliency and the vulnerabilities of these financial infrastructures to wide-scale disruptions. Any interruption in the normal operations of these infrastructures may seriously impact not only the financial system but also the economy as a whole.

A growing body of policy-oriented research is available. One segment of the literature focuses on simulating the default of a major participant and evaluating the effects on other institutions in payments[1] and securities settlement systems[2]. Another segment presents detailed case studies on the responses of the US financial system to shocks such as the 1987 stock market crash and the attacks of 11 September 2001.[3] Much of the research has been conducted

---

[1] See Humphrey (1986), Angelini et al (1996), Kuussaari (1996), Bech et al (2002), Northcott (2002), Bech and Soramäki (2005), Bedford et al (2005) and Mazars and Woelfel (2005).

[2] See Hellqvist and Koskinen (2005) and Devriese and Mitchell (2006).

[3] See Bernanke (1990), McAndrews and Potter (2002) and Lacker (2004).

using data from real operating environments with the given payment flows and settlement rules of the respective systems. As such they are useful for assessing the operation of the particular system under disruptions, but the results are difficult to generalise to systems with other characteristics. Little research has focused on explaining the relationship between the characteristics of the system and its performance during and following disruptions. Also the behaviour of participants has been generally exogenously defined or assumed unchanged (or to change in a predetermined manner) when the policy parameters of the system change or when a bank changes its settlement behaviour as a consequence of operational or financial problems. Such assumptions are unlikely to hold in the case of real disruptions.

This article argues that three aspects are important for answering the still unanswered questions on what makes a payment system and its participants robust or fragile towards disruptions, and what are the most efficient measures to reduce the likelihood and magnitude of disturbances. First, understanding the pattern of liquidity flows among the system participants. Second, understanding how the rules of the system affect the dynamics of liquidity flows. Third, the ability to evaluate likely behavioural changes of the participants before, during and following disruptions or as a consequence of policy changes.

This article presents new approaches at answering the above questions. It is organised as follows. Section 2.2 discusses how payment system interactions can be described by means of network topology and presents empirical results for the US Fedwire system. Section 2.3 describes dynamics that can take place in interbank payment systems and presents a simple model of a payment system based on simple rules of settlement. Section 2.4 presents some possible directions for modelling participant behaviour in payment systems. Section 2.5 concludes.

2.2 Modelling interbank payment flows

A payment system can be treated as a specific example of a complex network (see eg Newman, 2003). In recent years, the physics community has made significant progress towards understanding the structure and functioning of complex networks. The literature has focused on characterising the structure of networked systems and how the properties of the observed topologies relate to stability, resiliency and efficiency in case of perturbations and disturbances.

From a technical perspective, most payment systems are star networks where all participants are linked to a central hub (the operator) via a proprietary telecommunications network. From a payment processing perspective, payment systems are generally complete networks as all nodes (participants) are linked in the sense that they can send and receive payments from each other. However, these representations do not necessarily reflect the actual behaviour of participants that controls the flow of liquidity in the system and thus the channels for contagious transmission of financial disturbances. In common with other of social networks mediated by technology (such as email or telephone calling), the networks formed by actual participant behaviour are of more interest than the network structure of the underlying communication system.

2.2.1 Network representation of payment systems

Networks have been modelled in several disciplines such as in mathematics and computer science under graph theory, in applied mathematics and physics under network theory and in sociology under social network analysis. While the terminologies and research questions in the different traditions vary, common to all is the representation of the topic under study as (at minimum) two types of elements: nodes and connections between them, ie links. The following paragraphs summarise the main concepts.

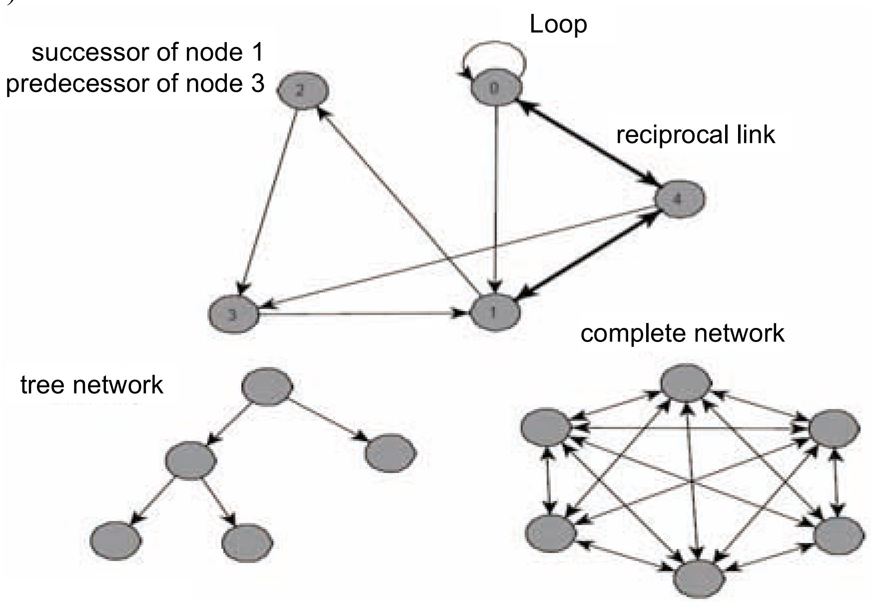
Links can be either undirected or directed. Links can have weights attached to them representing the importance of the relationship between nodes. The strength of a node can be calculated as the sum of the weights of all the links attached to it. For a directed network, strength can be defined over both the incoming and outgoing links.

A link from a node to itself is called a loop. The neighbours of a node are all the nodes to which it has a link. The predecessors of a node are the nodes that have a link to the node and the successors are the nodes that have a link from the node. A walk is a sequence of nodes in which each node is linked to the next. A walk is a path if all its nodes are distinct. The length of a path is measured by the number of links. If the start node and the end node of a path are one and the same, then it forms a cycle.

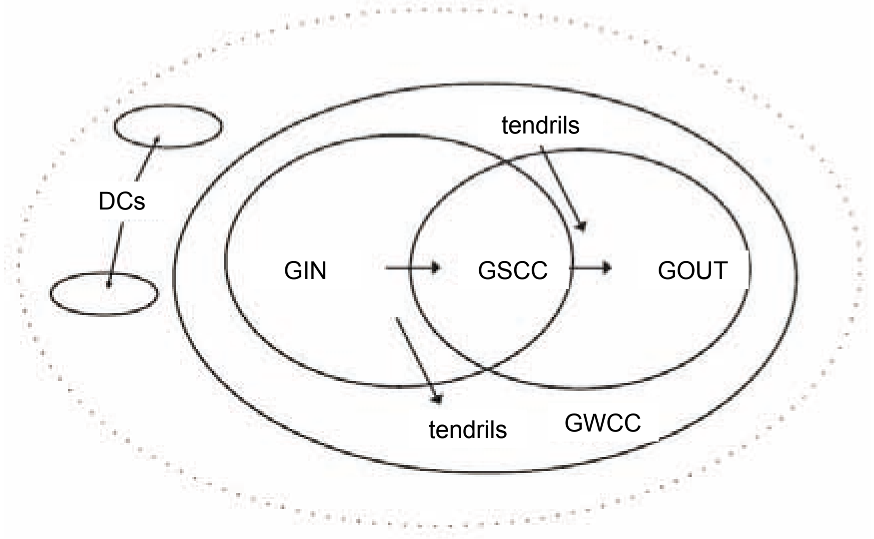
A complete network is a network where all nodes have a link to each other. A tree is a network in which any two nodes are connected by exactly one path. A connected network is a network where any two nodes can be joined by a path while a disconnected network is made up of two or more connected components or sub-networks. These concepts are illustrated in Figure 2.1a.

Figure 2.1 **Network modelling**

a)



b)



The most basic properties of a network are the number of nodes n and the number of links m. The number of nodes defines the size of a network while the number of links relative to the number of possible links defines the *connectivity* of a network. The *degree* of the network is the average number of links for each node in the network.

A starting point for the quantitative analysis of a network is to partition the set of nodes into *components* according to how they connect with other nodes. Dorogovtsev et al (2001) divide a network into a single giant weakly connected component (GWCC) and a set of *disconnected components* (DCs). The GWCC is the largest component of the network in which all nodes connect to each other via undirected paths. The DCs are smaller components for which the same is true. The GWCC consists of a *giant strongly connected component* (GSCC), a *giant out-component* (GOUT), a *giant in-component* (GIN) and *tendrils*. The GSCC comprises all nodes that can reach each other through a directed path. A node is in the GOUT if it has a path from the GSCC but not to the GSCC. In contrast, a node is in GIN if it has a path to the GSCC but not from it. Tendrils are nodes that have no directed path to or from the GSCC. They have a path to the GOUT or a path from the GIN (see Figure 2.1b).

Application of the component analyses to liquidity flows between banks provides insights on the structure of these flows within the payment system and gives clues with respect to the relative importance and vulnerability of banks in the system in case of disruptions. As banks in GOUT only receive funds from other banks in the GSCC, a disruption by a bank in GOUT would only affect other banks in that component. Banks in GIN are affected only by disruptions in the same component, and not by banks in other components as their payment processing is not dependent on incoming liquidity from these banks. Banks outside the GSCC tend to be smaller whereas all money center banks belong to the GSCC.

Two important characteristics of a node in a directed network are the number of links that originate from the node and the number of links that terminate at the node. These two quantities are referred to as the *out-degree* and *in-degree* of a node respectively. The average degree of a node in a network is the number of links divided by the number of nodes, ie <k>=m/n. Networks are often categorised by their degree distributions. The degree distribution of a classical random network (ER-network, Erdõs and Rényi, 1959) is a Poisson distribution. Many real networks have fat-tailed degree distributions and a large number have been found to follow the power law

$P(k_i = x) \sim k^{-y}$ for large-degree nodes. Networks with a power-law distribution are sometimes referred to as scale-free networks[4]. Scale-free networks have been found to remain better connected when subjected to random failures than other types of networks. Albert et al (1999) and Crucitti et al (2004) find that the connectedness of scale-free networks is robust to random failures but vulnerable to targeted attacks. However, one must be a bit careful here as the process acting on the network influences such analyses of robustness and vulnerability.

Simply put, banks that have a low in-degree and high weights for these links are likely to be more vulnerable to disturbances than other banks as the removal of one link will severely limit the amount of incoming funds. Conversely, banks with high out-degree have ceteris paribus the potential to affect more counterparties if their payment processing is disrupted. Understanding the topology of payment flows is likely to be important in assessing the resiliency of a payment system to wide-scale disruptions.

It is also common to analyse *distances* between nodes in the network. The distance from node i to node j is the length of the shortest path between the two nodes. The average distance from a node to any other node in a strongly connected network is commonly referred to as the average path length of a node. If the network is not strongly connected, paths between all nodes may not exist. In a payment network the path length may be important due to the fact that the shorter the distances between banks in the network, the easier liquidity can re-circulate among the banks. On the other hand, a payment system where liquidity flows over short paths is also likely to be more vulnerable to disruptions in these flows.

Sociologists have long studied *clustering* in social networks, ie the probability that two nodes which are the neighbours of the same node themselves share a link. This is equivalent to the observation that two people, each of whom is your friend, are likely to be friends with each other. One way of measuring the tendency to cluster is the ratio of the actual number of links between the neighbours of a node over the number of potential links among them. A tree network has a clustering coefficient of zero, and a complete network a coefficient of one. In a classical random network, the clustering coefficient is the unconditional probability of connection, ie $\langle C \rangle = p$.

---

[4] This is because the power law distribution is the only scale-free distribution, ie if the scale by which x is measured is increased by a factor, the shape of the distribution p(x) is unchanged, except for an overall multiplicative constant (see Newman, 2005).

In a payment network, the clustering coefficient measures the prevalence of payments between a bank's counterparties. In terms of resilience one could hypothesise that disturbances in banks with a higher clustering coefficient might have a compounding impact on their counterparties, as some of the disturbance may be passed on by the bank's neighbours to each other – in addition to the direct contagion from the source of the disruption.

There are various measures of the centrality that indicate the relative importance of nodes in a network. Four measures of centrality are commonly used in network analysis: degree, closeness, betweenness, and eigenvector centrality. The first three were described in their current form by Freeman (1979) while the last was proposed by Bonacich (1972). Degree centrality takes into account only the immediate neighbourhood of the node, ie it is simply the number of links the node has. Closeness centrality as defined by Freeman is the sum of shortest paths from all other nodes. Betweenness centrality may be defined loosely as the number of times that a node is on the shortest path between any pair of nodes. Eigenvector centrality encapsulates the idea that the centrality of a node depends also on the centrality of the nodes that it is linked by (or links to). A famous commercialisation of this centrality measure is the PageRank algorithm by Google (Brin and Page, 1995). In general, the importance of the node will depend on process taking place in the network. Borgatti (2005) provides a good overview of alternative processes in networks and centrality measures applicable for their analysis.

Finally, a key question in the study of networks is how the topologies that are seen in reality have come into being. There are two classes of network formation models some times referred to as equilibrium and non-equilibrium models (Dorogovtsev and Mendes, 2003). Equilibrium models have a fixed set of nodes with randomly chosen pairs of nodes connected by links. Erdõs and Rényi (1959) develop a basic model of a n node network, with each pair of nodes connected by a link with probability p. This type of network is commonly referred to as a classical random network. Non-equilibrium network models grow a network by successively adding nodes and setting probabilities for links forming between the new nodes and existing nodes and between already existing nodes. Many of these models, notably the Barabasi and Albert (1999) model (BA model), are based on preferential attachment. Preferential attachment assigns a probability of a link forming with a node that is increasing with the number of prior links of the node.

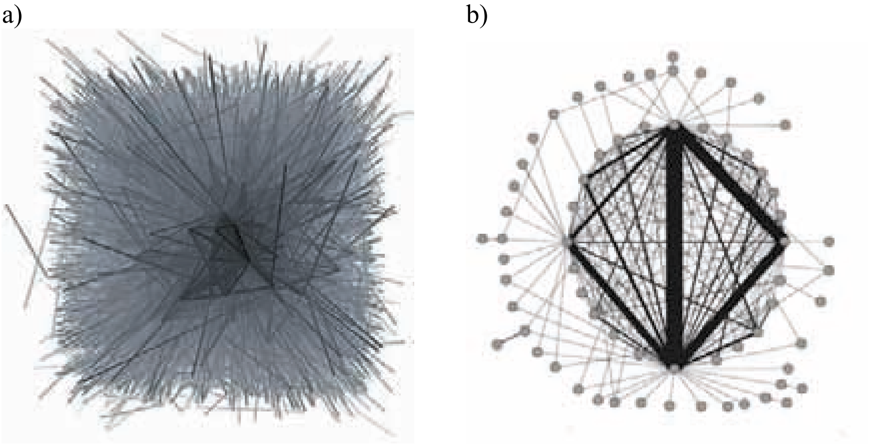
2.2.2 Fedwire as an example of a complex network

Soramäki et al (2007) analyse the topology of daily networks formed by the payment flows between commercial banks over Fedwire for a period of 62 consecutive business days. Apart from a few holidays, the statistics characterising the network were quite similar from day to day. These networks shared many characteristics with other empirical complex networks, such as a scale-free degree distribution, high clustering coefficient and the small world phenomenon (short path lengths in spite of low connectivity). Like many other technological networks, high-degree nodes tend to connect to low-degree nodes. Similar conclusions can also be reached from analysis on BoJ-NET by Inaoka et al (2005).

Moreover, Soramäki et al (2007) report that the topology of the network was significantly altered by the attacks of 11 September 2001. The number of nodes and links in the network and its connectivity was reduced, while the average path length between nodes was significantly increased. Interestingly, these alterations were of both similar magnitude and direction to those that occurred on several of the holidays contained within the period.

Figure 2.2a shows liquidity flows in Fedwire as a visual graph. The figure includes over 6,600 nodes and more than 70,000 links. Each link between two banks is shaded by the value of payments exchanged between them, with darker shades indicating higher values. Despite the appearance of a giant fur ball, the graph suggests the existence of a small group of banks connected by high value links. To gain a clearer picture of this group, a subset of the network where the focus is on high value links is displayed in Figure 2.2b. This graph shows the largest undirected links that comprise 75% of the value transferred. The network consists of only 66 nodes and 181 links. The prominent feature is a densely connected sub-graph, or clique, of 25 nodes to which the remaining nodes connect. By itself it is almost a complete graph. A small number of banks and the links between them thus dominate the value of all payments sent over the network.

Figure 2.2 **Visualisation of the liquidity flow network (Soramäki et al, 2007)**



The analysis finds that payment networks have characteristics similar to other social and technological networks. An unanswered question is why the network has the structure it does: the network may grow over time by a logic that is very general or that is particular to payment systems, or to specific policies of a given system. This is an interesting topic for future research. The network structure has also implications for its robustness. Robustness of the network, however, also depends on the processes taking place in it. This is the topic of the next sections.

2.3 Modelling payment system dynamics

2.3.1 Network dynamics

A number of payment system simulations carried out in recent years have used actual or generated payment data. These simulations have studied the actual dynamics of payment systems, where system rules have varied from simple real-time gross settlement to complex hybrid settlement mechanisms with offsetting and multilateral settlement capabilities. The research can be summarised as trade-off questions between liquidity, speed of settlement and risks. The impact of bank behaviour has not been taken endogenously into account in these simulations. A summary of this line of research is provided in Leinonen (2005) and is not presented here.

From a network perspective, the performance of banks (nodes) is often dynamically dependent on the performance of other banks within the network and upon the structure of linkages between banks. A failure by one node in the network, for example, may hinder flows in the network and adversely impact the performance of the other nodes as the disturbance propagates in the network.

One branch of network literature has investigated the resilience of different network topologies in terms of a connectivity threshold (ie percolation threshold)[5] at which a network dissolves into several disconnected components. A well-known finding is that scale-free networks are more robust to random failures than other types of networks. However, they are very susceptible to the removal of the very few highly connected nodes. These static failure analyses may be applicable to some networks if the interest is the availability of paths between nodes in the network – but are less applicable to networks of monetary flows which contain both flows via the shortest paths as well as longer walks within the network.

Another branch of the literature has studied the impact of perturbations that cascade through the network on the basis of established theoretical or domain-specific rules[6]. In these dynamical models nodes generally have a capacity to operate at a certain load and, once the threshold is exceeded, some or all of the node's load is distributed to neighbouring nodes in the network (Bak et al, 1987). While the detailed dynamics depend on the rules applied for the cascades, generally the most connected nodes (or nodes with highest load in relation to overall capacity) are more likely than average nodes to trigger cascades. Increased heterogeneity makes the system more robust to random failures, but more susceptible to targeted attacks that may cause global cascades.

Cascade models have been applied by physicists to systems within fields ranging from geology to biology to sociology (eg Jensen, 1998). This research has demonstrated that models made of very simple agents, interacting with neighbouring agents, can yield surprising insights about system-level behaviour. In the spirit of these cascade models, Beyeler et al (2007) formulate a simple agent-based model for liquidity flows within a payment system.

---

[5] Eg Bollobas (1985), Moore and Newman (2000) and Callaway et al (2000).

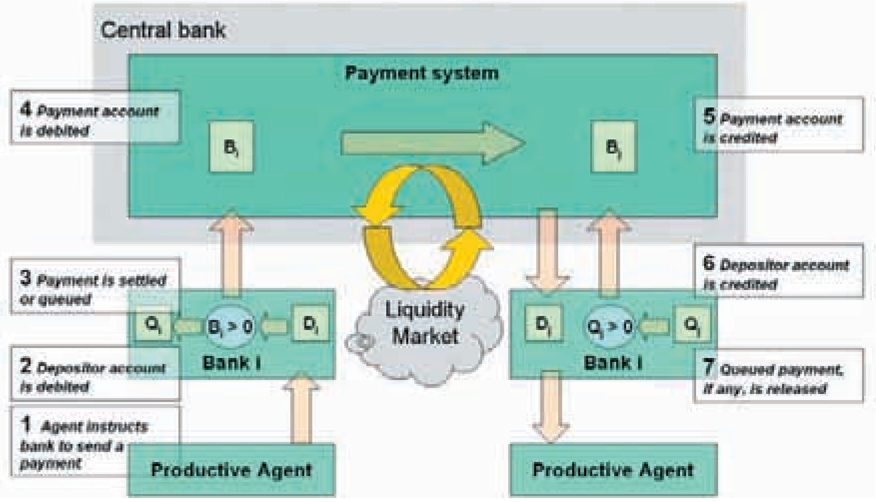
[6] Eg Watts (2002) and Crucitti et al (2004b) for random and complex networks, respectively, and Sachtjen et al (2000) and Kinney et al (2004) for power networks.

2.3.2 Simple payment system model

The model of Beyeler et al includes only the essential processes of a payment system and its accompanying liquidity market. A set of banks exchange payments through a single common payment system. All payments occur only along the links of a scale-free network – as was shown to be representative of Fedwire liquidity flows. Banks' customers randomly instruct them to make a unit payment to a neighbouring connected bank. Banks are reflexively cooperative: they submit the payment if the balance in their payment system account allows; otherwise they place the instruction on a queue for later settlement.

If the receiving bank has instructions in its queue, the payment it just received enables it to remove a queued instruction and submit a payment in turn. If the bank that receives that payment is also queuing instructions, then it can make a payment, and so on. In this way a single initial payment made by a bank can cause many payments to be released from the queues of the downstream receiving banks. This is an example of the cascade processes typically studied in other models of self-organised criticality. Statistics on these settlement cascades are an indicator of the extent of interdependence of the banks, and in the model they are a controlled by two parameters: the overall liquidity and market conductance.

Figure 2.3 **Simple payment system model (Beyeler et al, 2007)**

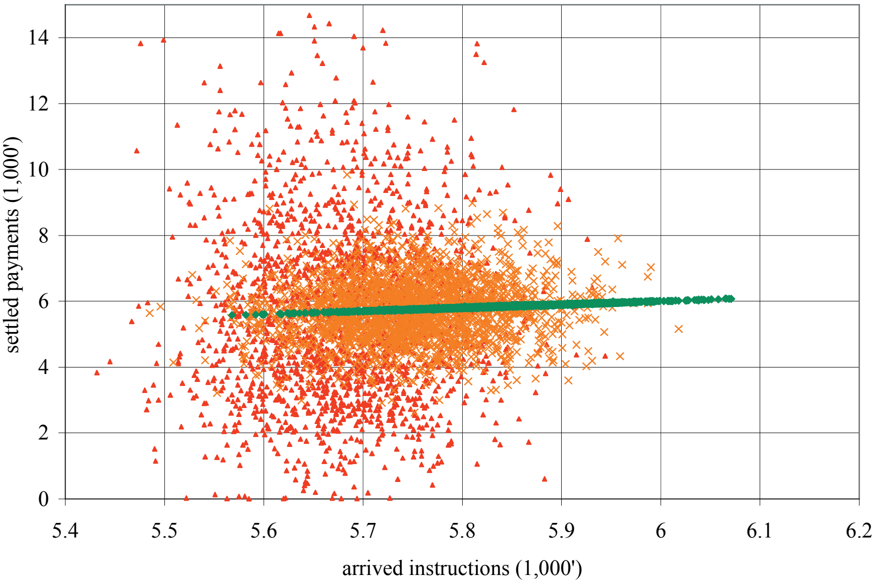


In the absence of a liquidity market, only abundant liquidity allows banks to operate independently; reducing liquidity increases the likelihood that a given bank will exhaust its balance and begin queuing payments. A bank that has exhausted its balance must wait for an incoming payment from one of its neighbours. When liquidity is low a bank's ability to process payments becomes coupled to its neighbours' ability to process. The output of the payment system as a whole is no longer determined by overall input, but instead becomes dominated by the internal dynamics of the system. Figure 2.4a shows how the correlation between arriving instructions and submitted payments degrades in the model as liquidity is reduced (1: high liquidity; 2: medium liquidity; 3: low liquidity). A settlement cascade, that is the release of queued payments as a result of a single initiating payment, can comprise hundreds of queued payments as illustrated in Figure 2.4b.

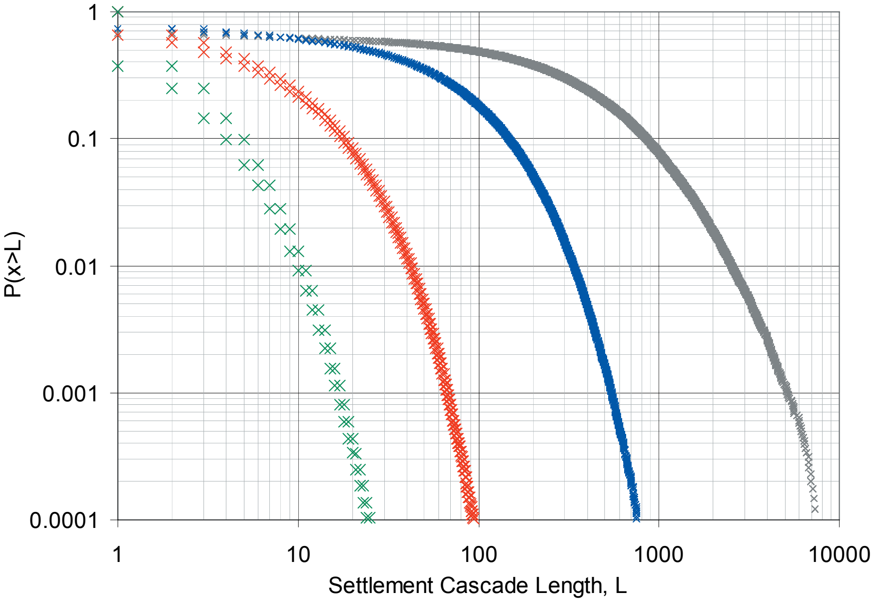
To explore how liquidity markets reduce coupling among network neighbours and thereby reduce congestion, market transactions were represented as a diffusive process where a bank's balance plays the role of a potential energy or pressure. Banks with high balances tend to contribute liquidity to the market, while banks with low balances tend to draw liquidity from the market. There is no decision-making or price-setting in this simple market model, but it reflects two essential features of a real market: liquidity flows from banks with surplus funds to banks that need funds, and liquidity can flow from any bank to any bank – flows are not confined to the links of the payment network. It creates a separate global pathway for liquidity flow. The ease of liquidity flow through the market is described by a single conductance parameter.

Figure 2.4 **Instruction and Payment Correlation (a) and Settlement Cascade Length Distribution (b).**

a)



b)



With a liquidity market included, the number of payments closely tracks the number of instructions as the coupling between banks is weakened and the size of the settlement cascades is reduced. The rate of liquidity flow through the market relative to the rate of flow through the payment system was surprisingly small. The performance of the system can be greatly improved even though less than 2% of the system through-put flows through the market.

2.4 Modelling bank behaviour

2.4.1 Decision-making, learning and adaptation

Wide-scale disruptions may not only present operational challenges for participants in the interbank payment system, but they may also induce participants to change the way they conduct business. The actions of participants have the potential to either mitigate or exacerbate adverse effects. Hence, understanding how participants interact and react when faced with operational adversity will assist operators and regulators in designing countermeasures, devising policy, and providing emergency assistance, if necessary.

The first approach to study bank behaviour in payment systems has been to use standard game theory. Angelini (1998) and Kobayakawa (1997) use a setup derived from earlier literature on precautionary demand for reserves. Angelini (1998) shows that in a RTGS system, where banks are charged for intraday liquidity, payments will tend to be delayed and that the equilibrium outcome is not socially optimal. Kobayakawa (1997) models the intraday liquidity management process as a game of uncertainty, ie a game where nature moves after the players. Kobayakawa (1997) shows that both delaying and not delaying can be equilibrium outcomes when intraday overdrafts are priced. McAndrews and Rajan (2002) study the timing and funding of transfers in the Fedwire funds transfer system. They show that banks benefit from synchronising their payment pattern over the course of the business day because it reduces the overdrafts. Bech and Garratt (2003) develop a stylised two-period-two-player model with imperfect information. They analyse the strategic incentives under different intraday credit policy regimes employed by central banks and characterise how the Nash equilibria depend on the underlying cost parameters for liquidity and delays. It turns out that two classical paradigms in game theory emerge: the Prisoner's Dilemma in the case where intraday credit is provided against collateral and the Stag Hunt coordination game in the case where the central bank charges a fee. Hence, many policy issues can be understood in terms of well-known conflicts and dilemmas in economics.

Other approaches that have been applied to similar problems of repeated interaction among a large number of players are evolutionary game theory and reinforcement learning (such as Q-Learning by Watkins et al, 1992). Agents who learn about each others' actions through repeated strategic interaction is a leading theme in evolutionary game theory. In most of the existing literature it is customary to look at the players' asymptotic behaviour in situations where the payoffs are some known function of players' strategies. In one strand of the literature, this knowledge is a prerogative of the players, who can therefore use adaptive rules of the type 'choose a best reply to the current strategy profile'. In a second research line, the learning rules do not require knowledge of the payoff function on the part of the learners. Such rules are instead of the kind 'adopt more frequently a strategy that has given a high payoff'.

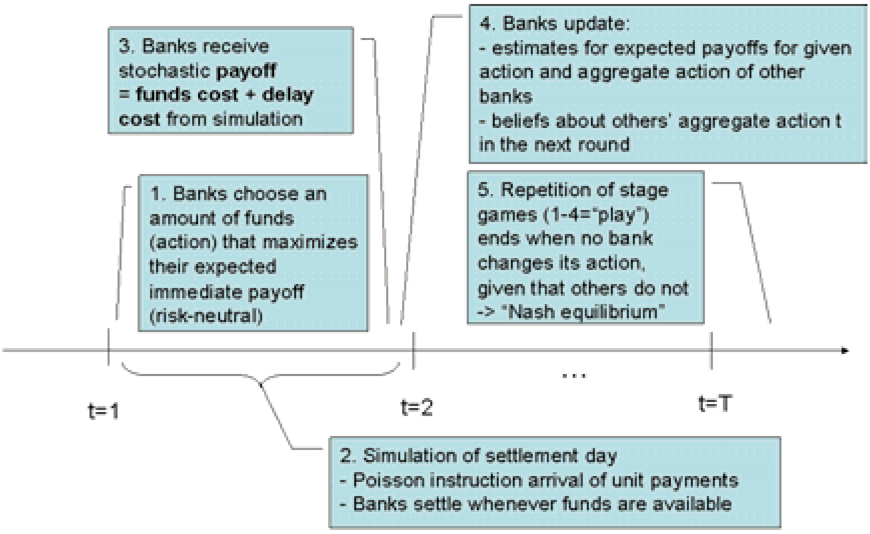
Galbiati and Soramäki (2007) use methods from reinforcement learning (Barto and Sutton, 1998) and fictitious play (Brown, 1951) to numerically solve a model with interactions among a large number of banks that settle payments on a continuous basis under imperfect information, stochastic payoffs and a finite but long sequence of settlement days. The model is summarised and discussed in more detail below.

2.4.2 Multi-agent model of bank behaviour

Galbiati and Soramäki (2007) develop a dynamic multi-agent model of an interbank payment system where payments are settled on the basis of pre-committed funds. In the model banks choose their level of committed funds on the basis of private payoff maximisation.

The model consists of a sequence of settlement days. Each of these days is a simultaneous-move game, in which each bank chooses the amount of liquidity to commit for payment processing and receives a stochastic payoff. Payoffs are determined by means of simulating the settlement day with the amounts of liquidity chosen by the banks. Instructions to be settled by the banks arrive on the basis of a Poisson process and are ex-ante unknown to the banks. As shown in Section 2.3.2, the relationship between instruction arrival and payment settlement is very complex and could not so far be described analytically. Adaptation takes place through reinforcement learning with Bayesian updating, with banks maximising immediate payoffs. Figure 2.5 shows the sequence of decisions, events and learning in the model.

Figure 2.5 **Overview of a multi-agent learning model of a payment system (Galbiati and Soramäki, 2007)**



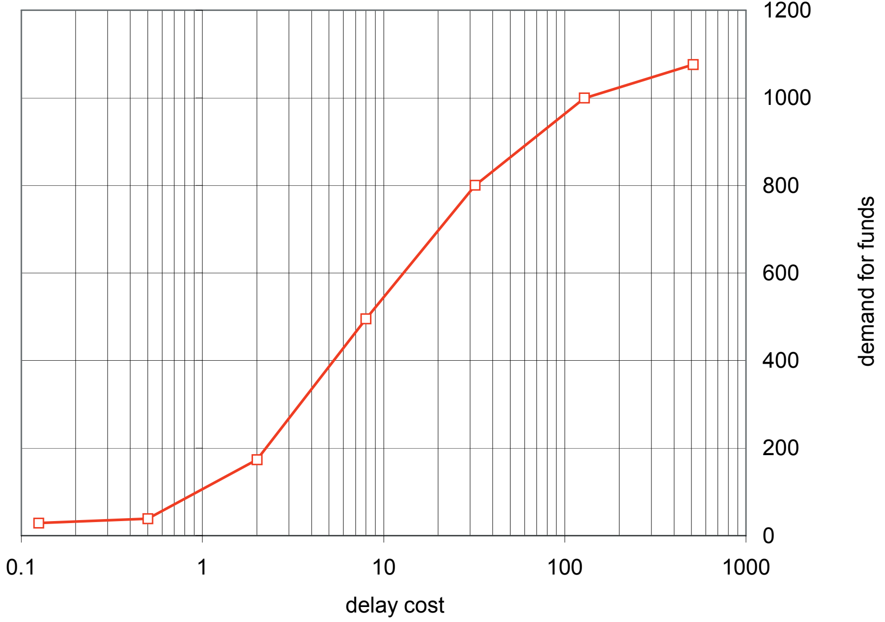
By the process of individual pay-off maximisation, banks adjust their demand for liquidity up (reducing delays) when delay costs increase and down (increasing delays), when they rise. It is well known that the demand for intraday credit is generated by a tradeoff between the costs associated with delaying payments and liquidity costs. Simulating the model for different parameter values, they find that the demand for intraday credit is an S-shaped function of the cost ratio between intraday credit costs and the costs associated with delaying payments[7] (see Figure 2.6a).

---

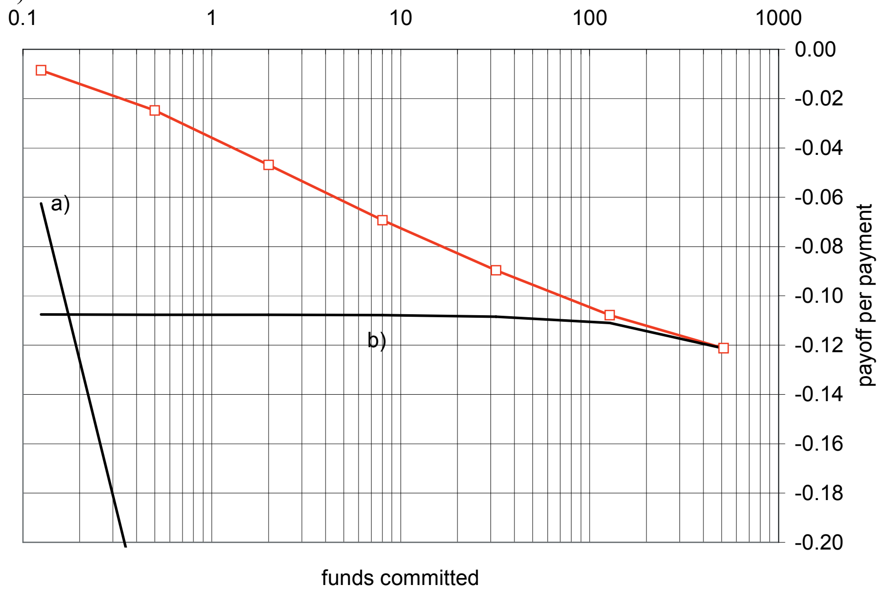
[7] In the model both costs are assumed to be linear.

Figure 2.6 **Demand for intraday credit (a), Payoff comparison (b)**

a)



b)



An interesting question is how good the performance of the banks is in absolute terms. To understand this we compare the payoffs received by the banks through adaptation with two extreme strategies:

a) delay all payments to the end of the day;
b) commit enough liquidity to be able to process all payments promptly.

The performance of these three strategies is shown in Figure 2.6b. For any level of the delay cost, the adaptive banks obtain better payoffs than either of the two extreme strategies as they manage to learn a convenient trade-off between delay and liquidity costs. On the contrary, the strategy under a) becomes quickly very expensive as delay costs increase, and the strategy under b) is exceedingly expensive when delays are not costly.

Ideally, banks should be taking into consideration the future stream of pay-offs as well. This would create a value of information to the banks as discounting expected future payoffs would create an explicit trade-off between exploitation (the use of actions that appear optimal in the light of the available information) and exploration (the use of seemingly sub-optimal actions, which might appear such because of lack of experimentation). Banks may also be risk-averse, interested not only in the expected pay-off but also its variability. These are among the topics for future research.

2.5 Conclusion

This article presented three elements of payment systems, new approaches for understanding and analysing them, and examples on how these approaches can be applied to specific research questions. It argues that performance of a payment system is a function of network topology, the ‘physics’ of the system and the behaviour of banks – one factor alone is not enough to evaluate efficiency or robustness.

First, the payment system can be understood as a network of liquidity flows and can be modelled as a graph. Each model of a payment system assumes some topology, be it random, complete or a topology closer to the system being modelled - such as the scale-free topology of Fedwire. Graph theory and social network analysis provide good tools for analysing the structure of interbank payment systems and their liquidity flows. Understanding how banks are connected in the payment network is important for analysing their robustness. The concepts developed in the field can help us structurally analyse payment flows in the system (see eg Newman, 2003). Measures of average path length can tell us how quickly disturbances are likely to reach other banks in the network. More research is clearly needed to identify measures that explain the connection between system topology and its robustness. Centrality measures can help us identify banks that are not only important through their size, but also due to their position in the network and due to their linkages to other banks (see eg Borgatti, 2005). A likely fruitful area in payment system research would be to use such approaches for the identification of important (and vulnerable) banks in networks representing RTGS or netting systems.

Second, payment systems have rules, procedures and technical constraints for the processing of individual payments that may produce emergent behaviour at the system level. An example of these is the settlement cascades that take place at low levels of liquidity and low market conductance. The model of payment system dynamics exhibits a transition from independent to highly interdependent behaviour and allows the study of factors that control system-wide interdependence. Complexity theory and models developed in statistical mechanics (see eg Bak, 1987, and Sachtjen et al, 2000) can help explain how simple local rules create emergent system-level behaviour.

Third, banks react to changes in the environment – be these changes in policy or disruptions to the system's operation or changes in the behaviour of other banks. Understanding how banks might react, and the impact of simultaneous reactions at the system level, greatly helps in evaluating risks and efficiencies of payment systems. While the incentives of banks may be analysed individually in isolation or when operating in a stipulated environment, their interaction in a system of banks with their own incentives necessitates a model. In modelling bank behaviour, methodologies developed under reinforcement learning (Sutton and Barto, 1998) and learning in games (Fudenberg and Levine 1998) may prove useful. As seen by the given example, mere simple 'intelligence' by agents can produce realistic behaviour and add value to the analysis of payment systems. In the development of more realistic behaviour for banks in settling payments, an important unanswered question is whether and what kind of bank behaviour can be identified from empirical payment data.

References

Angelini, P (1998) **An analysis of competitive externalities in gross settlement systems.** Journal of Banking Finance 22, 1–18.

Angelini, P – Maresca, G – Russo, D (1996) **Systemic risk in the netting system.** Journal of Banking and Finance 20, 853–868.

Albert, R – Jeong, H – Barabasi, A-L (1999) **Error and attack tolerance of complex networks.** Nature 406, 378–382.

Bak, P – Tang, C – Wiesenfeld, K (1987) **Self Organized Criticality: An explanation of 1/f Noise.** Physical Review Letters 59:4, 381–384.

Barabasi, A-L – Albert, R (1999) **Emergence of scaling in random networks.** Science 286, 509–512.

Barto, A G – Sutton, R S (1998) **Reinforcement Learning: an introduction.** MIT Press, Cambridge, Massachusetts.

Bech, M – Madsen, B – Natorp, L (2002) **Systemic Risk in the Danish Interbank Netting System.** Danmarks Nationalbank Working Paper No. 8.

Bech, M – Soramäki, K (2002) **Liquidity, gridlocks and bank failures in large value payment systems.** In E-money and Payment Systems Review. Central Banking Publications.

Bech, M – Soramäki, K (2005) **Systemic risk in a netting system revisited.** In Leinonen, H (ed.) Liquidity, risks and speed in payment and settlement systems – a simulation approach. Bank of Finland Studies E:31.

Bech, M – Garratt, R (2003) **The intraday liquidity management game.** Journal of Economic Theory, Vol. 109(2), 198–219.

Bech, M – Garratt, R (2006) **Illiquidity in the Interbank Payment System following Wide-Scale Disruptions.** Federal Reserve Bank of New York Staff Report 239.

Bedford, P – Millard, S – Yang, J (2005) **Analysing the impact of operational incidents in large-value payment systems: a simulation approach.** In Leinonen, H (ed.) Liquidity, risks and speed in payment and settlement systems – a simulation approach. Bank of Finland Studies E:31.

Bernanke, B (1990) **Clearing and Settlement during the Crash.** Review of Financial Studies, Vol. 3, Issue 1, 133–151.

Beyeler, W – Bech, M – Glass, R – Soramäki, K (2007) **Congestion and Cascades in Payment Systems.** Physica A. Forthcoming.

Bollobas, B (1985) **Random Graphs.** Academic Press, London, 123–136.

Bonacich, P (1972) **Factoring and weighting approaches to status scores and clique identification.** Journal of Mathematical Sociology 2, 113–120.

Borgatti, S P (2005) **Centrality and network flow.** Social Networks 27, 55–71.

Brin, S – Page, L (1998) **The anatomy of a large-scale hypertextual Web search engine.** Proceedings of the seventh international conference on World Wide Web 7, 107–117.

Brown, G W (1951) **Iterative Solutions of Games by Fictitious Play.** In Koopmans, T C (ed.) Activity Analysis of Production and Allocation. Wiley, New York.

Callaway, D – Newman, M – Strogatz, S – Watts, D (2000) **Network robustness and fragility: Percolation on random graphs.** Phys. Rev. Lett. 85, 5468–5471.

Crucitti, P – Latora, V – Marchiori, M – Rapisarda, A (2004a) **Error and attack tolerance of complex networks.** Physica A 340, 380–394.

Crucitti, P – Latora, V – Marchiori, M (2004b) **A model for cascading failures in complex networks.** Phys. Rev. E 69, 045104.

Devriese, J – Mitchell, J (2006) **Liquidity Risk in Securities Settlement.** Journal of Banking and Finance, Vol. 30, Issue 6, 1807–1834.

Dorogovtsev, S N – Mendes, J F F – Samukhin, A N (2001) **Giant strongly connected component of directed networks.** Phys. Rev. E 64.

Dorogovtsev, S N – Mendes, J F F (2003) **Evolution of Networks.** Oxford University Press.

Erdõs, P – Rényi, A (1959) **On random graphs.** Publicationes Mathematicae 6, 290–297.

Freeman, L C (1978) **Centrality in social networks: I. Conceptual clarification.** Social Networks 1, 215–239.

Fudenberg, D – Levine, D K (1998) **The Theory of Learning in Games.** MIT Press. Cambridge, Massachusetts.

Galbiati, M – Soramäki, K (2007) **A competitive multi-agent model of interbank payment systems.** arXiv:0705.3050v1 [cs.MA].

Hellqvist, M – Koskinen, J (2005) **Stress testing securities clearing and settlement systems using simulations.** In Leinonen, H (ed.) Liquidity, risks and speed in payment and settlement systems – a simulation approach. Bank of Finland Studies E:31.

Humphrey, D B (1986) **Payments Finality and Risk of Settlement Failure.** In Saunders, A and White, L J (eds) Technology and the Regulation of Financial Markets: Securities, Futures, and Banking. Heath: Lexington.

Inaoka, H – Ninomiya, T – Taniguchi, K – Shimizu, T – Takayasu, H (2004) **Fractal Network derived from banking transaction – An analysis of network structures formed by financial institutions.** Bank of Japan Working Papers No. 04-E-04.

Jensen, H J (1998) **Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems.** Cambridge University Press, ISBN 0-521-48371-9.

Kinney, R – Crucitti, P – Albert, R – Latora, V (2005) **Modelling cascading failures in the North American power grid.** Eur. Phys. J. B 46, 101–107.

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

Kuussaari, H (1996) **Systemic Risk in the Finnish Payment System: An Empirical Investigation.** Bank of Finland Discussion Paper 3.

Lacker, J M (2004) **Payment system disruptions and the federal reserve following September 11, 2001.** Journal of Monetary Economics, Vol. 51, No. 5, 935–965.

Leinonen, H (ed) (2005) **Liquidity, risks and speed in payment and settlement systems – a simulation approach.** Bank of Finland Studies E:31.

Mazars, E – Woelfel, G (2005) **Analysis, by simulation, of the impact of a technical default of a payment system participant.** In Leinonen, H (ed.) Liquidity, risks and speed in payment and settlement systems – a simulation approach. Bank of Finland Studies E:31.

McAndrews, J – Potter, S M (2002) **The Liquidity Effects of the Events of September 11, 2001.** Federal Reserve Bank of New York Economic Policy Review, Vol. 8, No. 2, 59–79.

McAndrews, J – Rajan, S (2000) **The timing and funding of Fedwire funds transfers.** Federal Reserve Bank New York Econ. Pol. Rev. 6, 17–32.

Moore, C – Newman, M E J (2000) **Exact solution of site and bond percolation on small-world networks.** Physical Review E 62: 7059–7065.

Newman, M E J (2003) **The Structure and Function of Complex Networks.** SIAM Review, Vol. 45, No. 2, 167–256.

Newman, M E J (2005) **Power laws, Pareto distributions and Zipf's law.** Contemporary Physics, Vol. 46, No 5, 323–351.

Northcott, C A (2002) **Estimating Settlement Risk and the Potential Contagion in Canada's Automated Clearing Settlement System.** Working Paper 2002-41. Bank of Canada.

Sachtjen, M L – Carrerras, B A – Lynch, V E (2000) **Disturbances in a power transmission system.** Physical Review E, 61:5, 4877–4882.

Soramäki, K – Bech, M L – Glass, R J – Beyeler, W E (2007) **The Topology of Interbank Payment Flows.** Physica A. Vol. 379, 317–333.

Watkins, C – Dayan, P (1992) **Technical note: Q-learning.** Machine Learning, 8(3/4), 279–292.

Watts, D J (2002) **A simple model of global cascades on random networks.** PNAS, April 30, 2002, Vol. 99, No. 9, 5766–5771.

Chapter 3

From PNS to TARGET2: the cost of FIFO in RTGS payment systems

Fabien Renault – Jean-Baptiste Pecceu

3 From PNS to TARGET2: the cost of FIFO in RTGS payment systems

Abstract

Most of the recent RTGS payment systems are equipped with various optimisation algorithms that are able to increase the settlement speed by resolving fully or partially some of the gridlock situations that arise in the system. Today, most of the optimisation algorithms in use follow – at least partially– the FIFO (First In First Out) rule, meaning that they always settle the queued payments in their order of arrival. While the FIFO rule may be desirable based on some other considerations, for example legal ground, it creates an additional constraint to the optimisation problem, potentially leading to a less efficient solution in terms of settled value. The aim of this paper is to try to quantify to which extent non-FIFO optimisation algorithms can be more efficient than FIFO algorithms.

In the first part of this paper, some simulations performed on randomly generated sets of payments are used to evaluate the efficiency of several FIFO and non FIFO optimisation algorithms. This analysis is conducted both in the case of bilateral optimisation and in the case of multilateral optimisation. The results show that in those conditions, some non-FIFO algorithms are able to improve significantly on their FIFO counterparts.

In a second part, the impact of the different optimisation algorithms is investigated further by simulating the complete PNS system using real data. In the context of a liquidity crisis created by the technical failure of the largest participant of the system, the use of some non-FIFO algorithms is shown to reduce the number of rejected payments at the end of the day.

3.1 Optimisation in RTGS

3.1.1 From net to hybrid systems

The last two decades have witnessed important transformations in the field of payment systems. Pure DNS (Deferred Net Settlement) systems, in which payment orders are stored throughout the day and the resulting net balances are settled only once at the end of the day, were the predominant form of LVPS (Large Value Payment Systems) in the 1980s. Although DNS systems are extremely efficient in terms of central bank money usage, the absence of intraday finality leading to potentially large intraday exposures raised some concerns in the context of ever-increasing values exchanged. Indeed if one participant fails to meet its end-of-day payment obligations in an unprotected DNS system, some or all payments involving this participant have to be unwound, potentially leading to the default of other participants and further unwinding. This potential domino effect can have unpredictable consequences on the final cash balances of each participant and on the number of rejected payments at the end of the day and thus undermines confidence in the payment system.

For these reasons, DNS systems were progressively replaced in the 1990s by RTGS (Real Time Gross Settlement) systems, in which payments are settled one by one as soon as the payment orders enter the system (and provided sufficient liquidity is available). Compared to DNS systems, RTGS systems tremendously reduce the risks associated with exchanging large value payments, but they also require significantly higher levels of central bank money to operate.

In order to reduce the central bank money usage of their participants, RTGS systems progressively adopted several payment-offsetting features.[1] Payments that cannot be settled immediately are held in a centrally-organised queue, and more or less sophisticated optimisation algorithms are used to try and simultaneously settle groups of queued payments that can not be settled individually.

Examples of such RTGS systems with offsetting mechanisms, sometimes referred to as ’hybrid systems’, include the French LVPS PNS (Paris Net Settlement) and the future pan-European system TARGET2. Besides offsetting algorithms, those two systems offer the participants the possibility to establish bilateral sending limits towards

---

[1] Here offsetting is to be understood as the gross execution of individual payments simultaneously within one legal and logical second. From a legal perspective, offsetting in RTGS is very different from the netting process in DNS.

their counterparties. A bilateral limit is the net amount of money a participant is willing to pay another participant before being paid back. This feature is helpful for risk management purposes and creates incentives to submit payments early into the system. Indeed, when intraday liquidity is scarce in a payment system, some participants might delay their payments in order to get a free ride on other participants' liquidity (see eg Bech and Garratt, 2003). When no bilateral sending limit feature is available, if bank A is not willing to grant bank B free intraday credit, the only solution bank A has is to retain its payments towards bank B in its own internal queue (located in its private IT infrastructure and invisible to the system and other participants). Conversely, if bank A can establish a bilateral limit towards bank B, bank A can submit payments towards bank B and let them be blocked by the RTGS system. Bank B is therefore incentivised to submit payments towards bank A. Doing so will not deplete bank B's liquidity stock because bank B's submission of payments towards bank A will trigger the release of bank A's payments towards bank B. Bilateral sending limits, together with offsetting mechanisms, thus transform intraday liquidity management from a competitive game (whoever submits his payments last wins) into a cooperative game (I will pay you at the exact time you pay me, so it is optimal for you to pay me early).

3.1.2 Optimisation and the FIFO rule

The benefits provided by offsetting algorithms in terms of lower liquidity needs in RTGS have been extensively investigated in recent years, notably thanks to the development of simulation tools for RTGS systems. Koponen and Soramäki (2005) and Leinonen and Soramäki (2005), among others, clearly showed how offsetting algorithms could for a given level of liquidity reduce the settlement delay and conversely reduce the liquidity needs for a given level of delay.

However, most of the analysis done until now relates to the use of optimisation algorithms that follow the First In First Out (FIFO) rule, meaning that payments have to be settled in the order they entered the system. While this constraint might be supported by some participants wishing to keep full control of their payment queue and might also be desirable from a legal point of view, it potentially lowers the efficiency of the optimisation algorithm in terms of settled value. Clearly, if a single very large payment is first in the queue, it might block many later-submitted smaller payments, and a FIFO algorithm will not be able to do anything about it.

The aim of this paper is to investigate other types of offsetting algorithms which do not necessarily follow the FIFO rule and to try and quantify to what extent non-FIFO optimisation algorithms can be more efficient than FIFO algorithms. In other words, we will try to calculate the cost of the FIFO rule for RTGS systems in terms of decreased efficiency of the optimisation mechanisms.

Bech and Soramäki (2001 and 2005) formalised the problem by introducing a clear distinction between the Gridlock Resolution Problem (GRP, ie the problem of optimisation under the FIFO constraint, as defined by Bech and Soramäki) and the Bank Clearing Problem (BCP, ie the free optimisation problem, as referred to by Güntzer et al, 1998).

3.1.3 Formalisation of the problem

The notations used in this section come from Bech and Soramäki (2001). We consider n banks ($i = 1 \ldots n$) participating in a RTGS system, each characterised by its initial amount of liquidity S_i. The queue of bank i contains m_i payments waiting to be settled. The kth payment sent by bank i is characterised by its value $a_{i,k}$ and the receiving bank designated by the integer $r_{i,k} \in \{1,2,\ldots,n\} \setminus \{i\}$.

In order to be able to characterise any subset of the queued payments, we will use the indicators $x_{i,k} \in \{0,1\}$. A value of 1 (respectively 0) attributed to $x_{i,k}$ simply means that the kth payment of bank i is included (respectively not included) in the considered subset.

Bech and Soramäki define the **Gridlock Resolution Problem** as finding the $(x_{i,k})_{\substack{i=1\ldots n \\ k=1\ldots m_i}}$ that maximise the total value settled $V = \sum_{i=1}^{n} \sum_{k=1}^{m_i} a_{i,k} x_{i,k}$ under the dual condition

$$\begin{cases} \forall i \in \{1 \ldots n\}, \quad S_i - \sum_{k=1}^{m_i} a_{i,k} x_{i,k} + \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{k=1}^{m_j} a_{j,k} x_{j,k} \delta_{r_{j,k}=i} \geq 0 \\ \qquad \forall i \in \{1 \ldots n\}, \quad \forall k \in \{1 \ldots m_i - 1\}, x_{i,k+1} \leq x_{i,k} \end{cases}$$

where $\delta_{r_{j,k}} = i$ is equal to 1 if $r_{j,k} = i$ and zero otherwise.

The first condition is the liquidity constraint. It simply states that a bank cannot have a negative cash balance within the considered payment system. The second condition is the sequence constraint. It simply translates the fact that bank i wants its payments settled in the chronological order in which they were received by the system.

The **Bank Clearing Problem**, as defined by Güntzer et al (1998) is similar to the Gridlock Resolution Problem with the difference that the second constraint (the sequence condition) is not present in the BCP.

3.1.4 Non-FIFO features in PNS and TARGET2

Neither the French LVPS PNS nor the future pan-European RTGS TARGET2 totally comply with the sequence constraint of the GRP problem, as explained in the previous section. Indeed, the FIFO rule is arguably breached on several occasions.

First, in both PNS and TARGET2, the FIFO principle is to be understood on a bilateral basis. A payment from bank A to bank B can be settled before a payment from bank A to bank C that entered the system earlier. Moreover, it is clear that such an exception to the FIFO rule will be present in all systems offering the participants the possibility to set bilateral limits towards their counterparties. Indeed, if the payment from bank A to bank C is queued because the bilateral limit bank A has set towards bank C has been reached, bank A will still want to be able to settle payments towards its other counterparties.

Furthermore in PNS, a low value payment (whose value is lower than EUR 1 million) from bank A to bank B will be settled directly by the entry mechanism of the system, provided bank A has the necessary funds and whether or not earlier submitted payments from bank A to bank B are present in the queue. The aim of this rule is to avoid a situation where a very large queued payment creates a blockage, unnecessarily delaying the settlement of many small payments. A similar feature exists in the entry mechanism of TARGET2: indeed when a normal priority payment[2] is submitted, '*it is not checked whether the normal [priority payments] queue is empty, because the FIFO principle can be breached for normal [priority] payments'*.[3]

---

[2] In TARGET2, the participants will be able to choose either normal priority or urgent priority for each payment they emit.

[3] TARGET2 User Detailed Functional Specifications, first book version 2.0, page 145.

Even when retaining a bilateral definition of the FIFO rule, PNS (for payments lower than EUR 1 million) and TARGET2 (for normal priority payments) do not comply with this rule in the entry mechanism.

Finally, another breach of the FIFO rule occurs in the multilateral optimisation algorithm of the PNS system which attempts to settle simultaneously all queued payments of all participants. In case it is impossible to settle all queued payments because one or several participants do not have sufficient liquidity, the algorithm will consider the participant having the largest net debit position and de-activate the smallest of its payments whose value exceeds the value of its net debit position (in case no payment exceeds the value of the net debit position, the biggest payment of the participant having the largest net debit position is de-activated). In this special case, the payments are then selected according to their value, and not according to the order they arrived in the system.

3.1.5 Objectives of the paper

We have just shown that the settlement process of PNS, in particular for low value payments, and of TARGET2 in the case of normal priority payments, breach the FIFO rule on several occasions. Moreover, in TARGET2 normal priority payments can by-pass other queued payments in the entry mechanism while they are treated according to a strict FIFO rule (in a bilateral-FIFO sense) in the bilateral optimisation algorithm. One can thus feel entitled to investigate the benefits non-FIFO optimisation algorithms could bring to the system.

There are several good reasons for a payment system to follow the FIFO principle: it makes the rules of the system easier and allows participants to keep full control of the order their payments are settled. For this last reason in particular, some treasurers are very fond of the FIFO principle. Moreover, FIFO optimisation algorithms are fast, simple to understand and easy to implement while efficient enough to solve many gridlock situations.

In theory the drawback associated with the lack of flexibility the FIFO principle represents is decreased settlement efficiency. All other things being equal, a pure FIFO RTGS is characterised by a higher settlement delay than a RTGS equipped with more advanced non-FIFO offsetting algorithms.

The aim of this paper is not to discuss whether or not the FIFO principle should continue to be applied in today's RTGS, as many other considerations may influence the conclusion that could be made regarding this topic. Instead, the objective of this contribution is to try and quantify the expected increase in settlement efficiency that would allow the use of non-FIFO offsetting algorithms.

Two types of optimisation algorithms co-exist in PNS and in the future TARGET2 system: bilateral optimisation and multilateral optimisation. We will examine them successively in a theoretical framework before moving to a 'real-life case' in the PNS system.

3.2 Bilateral optimisation

In this section we focus on bilateral optimisation, ie we examine two participants A and B and consider only queued payments from A to B and from B to A. The objective of a bilateral optimisation algorithm is to settle simultaneously a set of queued payments for as high a total cumulated value as possible (the number of settled payments is also of interest as a 'secondary objective', although the settled value is usually considered more important).

One may wonder why optimisation should be performed on a bilateral basis rather than directly on a multilateral basis, ie considering all queued payments of all participants at the same time. In theory, any solution provided by a bilateral optimisation algorithm could also be found by a multilateral optimisation algorithm while the opposite is not true. In practice, bilateral optimisation takes profit from the usually relatively high level of reciprocity of payment networks in order to drastically reduce the number of variables and the complexity of the problem. Another important element is the presence of bilateral sending limits (cf 3.1.1) which create a strong linkage between the payments exchanged by a pair of participants (A will pay B if and only if B pays A). While the treatment of bilateral limits is cumbersome in a multilateral optimisation algorithm, it is very easily implemented and effective in a bilateral optimisation algorithm.

For those reasons, bilateral optimisation and multilateral optimisation can be considered as complimentary and are both used in PNS and in TARGET2.

3.2.1 Bilateral optimisation in PNS and TARGET2

The two systems, PNS and TARGET2, rely on the same bilateral optimisation algorithm. This algorithm follows the FIFO rule in a bilateral sense. First, the algorithm tries to settle all payments queued between the two banks simultaneously. If this is not possible, the most recent payment from the participant lacking liquidity is de-activated. This process is iterated until all payments have been de-activated or until a solution has been found. The 'FIFO bilateral optimisation algorithm' is described in detail in Appendix 1.

The fact that PNS and TARGET2 rely on the same bilateral optimisation algorithm comes as no surprise. It is indeed quite easy to show that the bilateral optimisation algorithm used in PNS and TARGET2 is the best algorithm that abides by the bilateral-FIFO rule, in the sense that it will always provide the unique solution maximising both the volume and value settled (Bech and Soramäki, 2001).

3.2.2 The bilateral Greedy algorithm

The bilateral Greedy algorithm was proposed by Güntzer et al in 1998. Payments are not retained according to their arrival order but according to their value. As in the FIFO bilateral optimisation algorithm, the Greedy algorithm first tries to settle all payments queued between the two banks simultaneously. If this is not possible, all payments from the participant lacking liquidity are de-activated and are then re-activated whenever possible given the liquidity constraint in the decreasing order of their value. This process is iterated until all payments have been de-activated or until a solution has been found. The details of the algorithm can be found in Appendix 1. Compared to the FIFO algorithm used in PNS and TARGET2, bigger payments are favoured at the expense of payments that entered the system early. One of the advantages of the Greedy algorithm over the FIFO algorithm is that queues will not be blocked due to a single very large payment that would prevent all subsequent payments from settling.

A very interesting property of the Greedy algorithm is that it yields a solution that maximises the value of payments settled when the sequences of values of the queued payments are superincreasing, that is to say when every queued payment from A to B is larger than the sum of all the smaller queued payments from A to B and every queued payment from B to A is larger than the sum of all the smaller queued payments from B to A. A proof of this claim is presented in Appendix 3. In the case of the PNS system, it can be shown that any average set of 3 payments has a 95% chance of forming a superincreasing sequence. This probability drops to 65% if we consider a set of 5 payments and to only 2% if we consider a set of 10 payments. The ideal case of superincreasing sequences is therefore not unrealistic when there are only few queued payments between two given participants (as is often the case in PNS in normal working conditions). It is also important to keep in mind that the Greedy algorithm can very well provide the best solution even if the payment sequences are not superincreasing, although this is not guaranteed in this case.

Another interesting feature of the Greedy algorithm lies in its simplicity and speed. Indeed, once queued payments have been ordered according to their value, the number of operations to perform is only proportional to the number of queued payments, that is to say the Greedy algorithm is not slower than the simple FIFO algorithm used in PNS and TARGET2. The time needed to order a set of N payments is typically proportional to N.log(N) but such a task only needs to be performed once. Furthermore, the tests showed that compared to the FIFO algorithm, fewer iterations were needed for Greedy to produce a solution.

3.2.3 New ideas regarding bilateral optimisation

We present some new ideas regarding bilateral optimisation. The Greedy algorithm is already very efficient but is not guaranteed to give the best solution when payment values are not superincreasing. Is it possible to improve on Greedy?

Two distinct ideas were investigated. The first idea is to introduce some flexibility to Greedy, which always re-activates payments in the decreasing order of their value. We consider the problem of bilateral optimisation between bank A and bank B and denote payments from A to B as the $(a_i)_{i=1...N}$ where a_1 is the biggest payment and a_N is the smallest payment. It is clearly optimal to re-activate a payment a_i satisfying $a_i \geq \sum_{k=i+1}^{N} a_k$ as we know that the Greedy algorithm will yield the best answer for a superincreasing payment sequence (see Appendix 3). On the other hand, if the sequence is not locally superincreasing, ie if $a_i < \sum_{k=i+1}^{N} a_k$, it is unclear whether the payment a_i should be re-activated or not. The idea behind the Las Vegas Greedy bilateral optimisation algorithm is to try both solutions, stochastically. The algorithm is presented in more detail in Appendix 1. It is important to note that for superincreasing payment value sequences, the Las Vegas Greedy algorithm degenerates into Greedy.

The other idea investigated was to try and benefit from the ever-increasing computational power available to try more settlement possibilities than Greedy does. While this computational power might not be sufficient to try each of the 2^{N+M} possibilities involved (N is the number of queued payments from A to B and M is the number of queued payments from B to A), it is reasonable to consider that some of them might be tested to the limit of, say, a thousand cases. The key question is now how to select those cases to be tested. The idea behind Greedy++ is to run the Greedy algorithm, and after each iteration that does not yield a solution because one of the participants does not have the needed liquidity, to test all possibilities involving the 10 payments closest to the error (ie the $2^{10} = 1024$ cases obtained when considering the re-activation/de-activation of the 10 payments closest to the error, all other payments staying in the same state). If a solution is found, then the solution maximising the settled value will be chosen. When this treatment yields no solution, a Greedy iteration is applied, hence the name of Greedy++ for this algorithm. The algorithm is presented in more detail in Appendix 1.

3.2.4 A few basic examples

The aim of this section is to help understand concretely how the algorithms work on practical examples.

Example 1

		Bank A	Bank B
Cash balance		10	10
Queued payments, in order of arrival	1st	500	20
	2nd	20	20
	3rd	20	20
	4th	20	20

In this example a large payment from bank A to bank B (of value 500) is preventing subsequent payments from settling. Clearly, nothing can be settled with a FIFO algorithm as any solution would involve the by-passing of bank A's earliest-sent payment. The Greedy algorithm, as well as Greedy++ and the bilateral Las Vegas Greedy, will however find the value maximising solution (settle bank A's second, third and fourth payment together with three of bank B's payments).

Example 2

		Bank A	Bank B
Cash balance		0	10
Queued payments, in order of arrival	1	140	20
	2		20
	3		30
	4		100

This example is typical of non-superincreasing payment sequences (here $30 < 20+20$ so the sequence of Bank B's payments is not superincreasing). The Greedy algorithm will start by activating all payments, and as bank B has a negative virtual position (-30), will de-activate all payments from B to A and re-activate them in the decreasing order of their value. By re-activating the payment of value 30, Greedy will miss the trivial solution ($140 = 100 + 20 + 20$) and terminate without settling any payment.

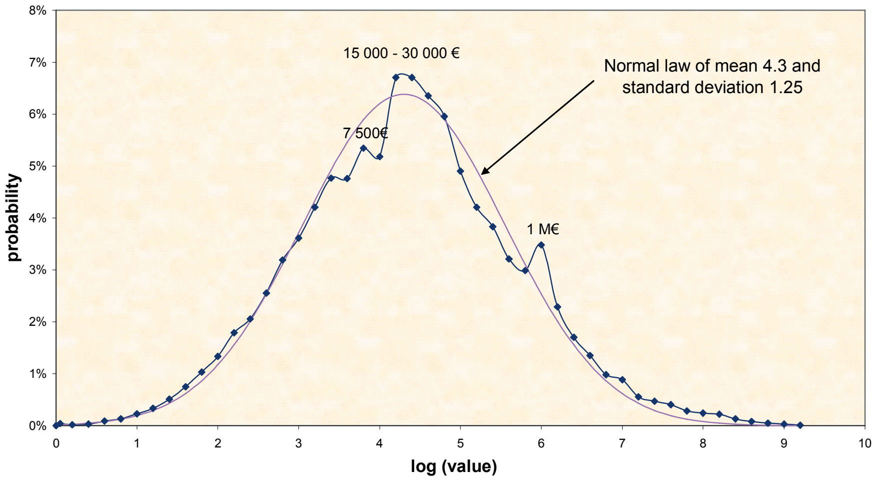
The Greedy++ algorithm will start by activating all payments, and as settlement is impossible, will examine all possibilities involving the 10 payments closest to the error (here the error is equal to 30, and as there are only 5 payments in the queue, the $2^5 = 32$ possibilities will be tried, and the value maximising solution will be retained). Greedy++ will therefore find the correct solution – as always when the number of queued payments is fewer than 10.

The bilateral Las Vegas Greedy algorithm will also start by activating all payments, and after noticing that B has a negative virtual position, will de-activate all payments from B to A. Payments from B to A will then be considered for re-activation in the decreasing order of their value, up to a total cumulated value of 140 (the sum of the activated payments from A to B + B's position). Bank B's biggest payment, of value 100, will be re-activated with a probability of 100% since the cumulated value of the lower payments, 70, is strictly lower than B's virtual position of 140. The payment of value 30 is then considered for re-activation. It will be re-activated with a probability equal to $\frac{30}{20+20} = 75\%$. If the algorithm is launched 10 times, the probability for the value-maximising solution to be found is then close to 95%.

3.2.5 Relative efficiency of bilateral optimisation algorithms

In order to compare the different bilateral optimisation algorithms presented in the previous pages, the following test was developed:

Figure 3.1 **Payment value distribution in the PNS system**



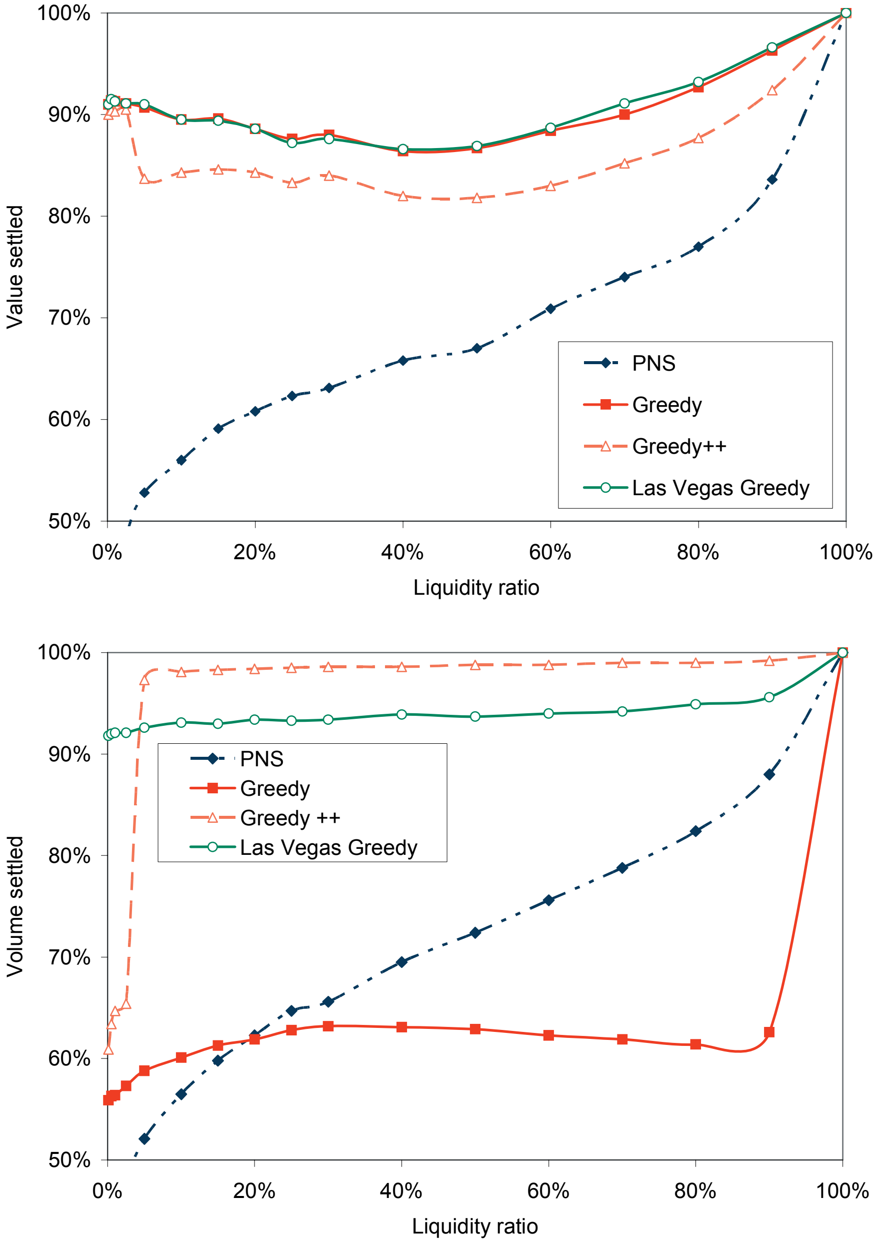
- We considered two banks, A and B. We assumed that there are N queued payments from bank A to bank B, the $(a_i)_{i=1...N}$, and N queued payments from bank B to bank A, the $(b_i)_{i=1...N}$. The value of each of these 2N payments was generated randomly according to the observed payment value distribution in the PNS system: as shown in the above graph, the payment distribution in PNS can be approximated by a log normal law of mean 4.3 and of standard deviation 1.25 with great accuracy.

- We can assume without any loss of generality that the sum of the values of the payments emitted by A, designated by $G = \sum_i a_i$ exceeds the sum of the values of the payments emitted by B, noted as $H = \sum_i b_i$. The starting balance of bank B, S_B is then set to zero, while the starting balance of bank A, the net emitter, is set to $S_A = \alpha(G - H)$, where α is a parameter ranging from 0 (no liquidity is present at all), to 1 (all queued payments can be settled simultaneously).

- The presented problem of bilateral optimisation was run with the PNS/T2 FIFO algorithm, Greedy, Greedy++ and Las Vegas Greedy bilateral algorithm. Regarding the Las Vegas Greedy bilateral algorithm, it was applied 5 times in a row (ie it was applied a first time to the initial problem, then it was applied a second time to what had not settled the first time, and so on…). The results were averaged over 5,000 different payment distributions generated randomly, according to the presented log normal law.

- The results obtained in terms of value and volume settled are shown in Figure 3.2. While the volume efficiency is defined simply as the ratio between the number of settled payments and the total number of queued payments 2N, it was thought more significant to define the value efficiency as the ratio between the cumulated value of settled payments and the maximum amount that can be settled if payments can be split: $2H + S_A$.

Figure 3.2 **Bilateral optimisation, value (top) and volume (bottom) settled versus liquidity**



3.2.6 Conclusion regarding bilateral optimisation

In terms of settled value, while the three presented non-FIFO algorithms perform significantly better than the standard FIFO algorithm, especially at low liquidity levels, the use of the most complicated algorithms (Las Vegas Greedy and Greedy++) does not yield better results than the use of the simple Greedy algorithm.

In terms of settled volume, however, the Greedy algorithm performs significantly worse than the standard FIFO algorithm, with only 60% of the total number of payments settled when 90% of liquidity is available. On the other hand, the Greedy++ algorithm is basically able to settle 99% of all queued payments whenever more than 5% of the liquidity needed to settle all payments is present.

The best overall performance is arguably realised by the Las Vegas Greedy bilateral algorithm, which performs extremely well both in terms of volume and value. If only the settled value matters, the simple Greedy algorithm – simpler and faster than Las Vegas Greedy and Greedy++ – is the natural choice. Finally, the bilateral optimisation algorithm implemented in TARGET2 and PNS provides satisfactory results given the strong constraint represented by the FIFO rule.

3.3 Multilateral optimisation

This section focuses on multilateral optimisation. This time, all participants and all queued payments are considered simultaneously. The aim of multilateral optimisation is to find a set of payments – as far as possible with the largest cumulated value – that can be settled simultaneously.

3.3.1 Multilateral optimisation in PNS and TARGET2

The multilateral optimisation algorithm of both PNS and TARGET2 starts by activating all queued payments. Of course, if all participants have a positive virtual cash balance,[4] all the payments are settled

---

[4] Throughout this paper, the virtual cash balance of a participant designates its cash balance if all the activated payments of all participants in the system are settled simultaneously. Clearly a necessary condition for all activated payments to be settled is that all participants should have a positive virtual cash balance.

simultaneously. In the opposite situation, the participant with the largest net debit position is considered by the multilateral optimisation algorithm of both TARGET2 and PNS. The approach followed is then slightly different in the two systems.

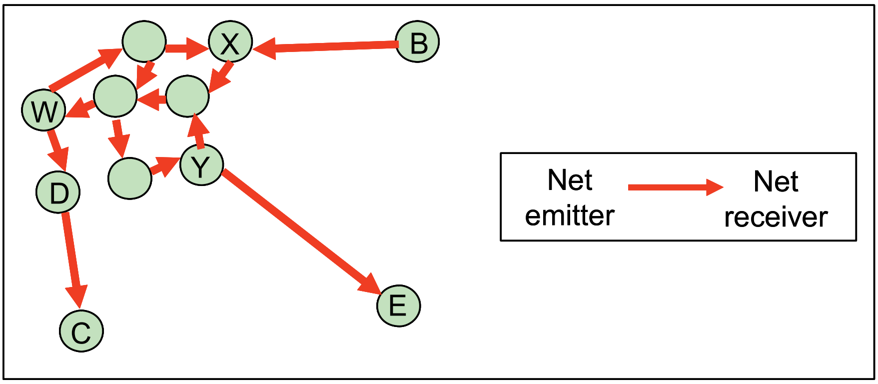
In TARGET2, the algorithm will simply de-activate the most recent payment of the participant with the largest net debit position.

In PNS, the algorithm will consider the participant with the largest net debit position but this time de-activate the smallest payment whose value exceeds the value of its net debit position. (In case no payment exceeds the value of the net debit position, the biggest payment of the participant having the largest net debit position is de-activated, then the second biggest, and so on until one payment exceeds the value of the participant's net debit position.)

3.3.2 A new concept: pre-conditioning

The concept of pre-conditioning is a new idea in the field of multilateral optimisation. The basic idea is to make the most of the existing liquidity by simply letting it flow towards the central core of the payment network. In order to do so, we de-activate as many as possible of the queued payments towards the peripheral participants, who only exchange payments with a single other bank. (In the sketch below, the peripheral participants are B, E and C. By recursion, once the payment from D to C has been de-activated, D will also become a peripheral participant and the payment from W to D will be de-activated.)

Figure 3.3 **Pre-conditioning algorithm**



There are two kinds of peripheral participants:

- The net emitters such as B (B is a net emitter because the cumulated value of queued payments from B to X is larger than the cumulated value of queued payments from X to B), which are a source of liquidity for the network. However, the reason for some payments between X and B being held in the queue is that B does not have the necessary liquidity to settle its net position. As B cannot receive liquidity from any other participant, the set of queued payments between X and B will never be settled as a whole.

- The net receivers (such as C and E) are liquidity traps for the network (C is a net receiver because the cumulated value of queued payments from C to D is smaller than the cumulated value of queued payments from D to C). Indeed the liquidity transmitted from Y to E will not be used again for further settlement.

We can then conclude that whatever their net position (net emitters or net receivers), peripheral participants always have a negative impact on the network. We can therefore try and improve the efficiency of a multilateral optimisation algorithm by removing them before the algorithm is launched.

In the example presented in Figure 3.3, the pre-conditioning algorithm will therefore de-activate all payments from or towards participants B, E, C and then D. Once the multilateral optimisation algorithm has been applied to the network, payments involving peripheral participants will be dealt with separately with the help of bilateral optimisation algorithms.

This pre-conditioning algorithm was implemented in the following algorithms presented in this paper: the Multilateral Greedy Las Vegas, the Multilateral PNS Las Vegas and the OPM1010 algorithm.

3.3.3 The multilateral Las-Vegas algorithms

As in bilateral optimisation, some algorithms trying to use randomly generated numbers to improve on the efficiency of standard algorithms were developed, such as the Multilateral Greedy Las Vegas and the Multilateral PNS Las Vegas algorithms.

The Multilateral PNS Las Vegas algorithm is based on the algorithm used in PNS. However, instead of de-activating the smallest payment that is larger than the deficit of the bank with the largest debit position, the algorithm randomly chooses which payment to de-activate. In order to so, each payment is affected by a certain 'de-activation probability' based on three different criteria: payments whose value is close to the net debit position of the emitter, payments whose de-activation allows the emitter to reach a net credit position and, finally, payments whose de-activation neither creates nor aggravates the deficit of another participant are de-activated with a higher probability. Appendix 2 provides more insight on the details of the algorithm.

The Multilateral Greedy Las Vegas algorithm is somewhat similar, with the exception that instead of de-activating payments of participants with a net debit position payment by payment, a 'Greedy approach' is followed. All payments originating from the considered participant with a net debit position are de-activated and are considered for re-activation in decreasing order of their value, as in Greedy, but also taking into account the position of the receiver of the payment (payments towards participants with a net debit position are re-activated with a higher probability).

As in bilateral optimisation, the use of random numbers is a way to create algorithms which can be run several times. In the following tests, the Las Vegas algorithms were applied five times in a row (ie the algorithm was applied a first time to the initial problem, then it was applied a second time to what had not settled the first time, and so on…).

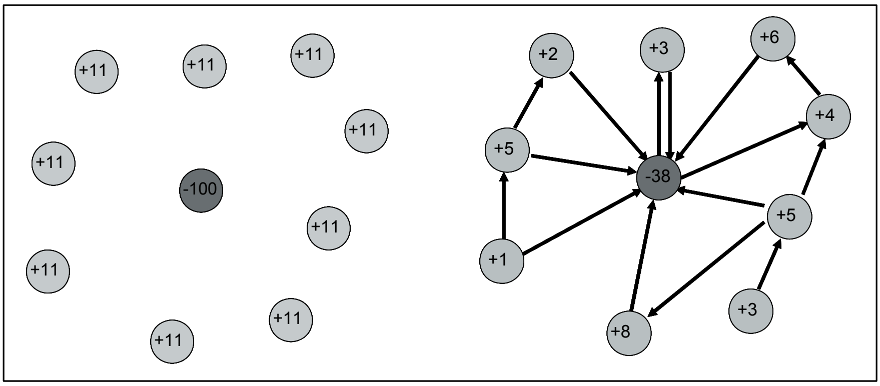
The last algorithm tested is the OPM1010 algorithm. It is quite close to the Multilateral PNS Las Vegas algorithm, with the difference that the payments are not de-activated randomly but in a deterministic way. For each payment, a 'de-activation score' is calculated by considering the net positions of the emitter and of the receiver, and the payment with the higher score is de-activated.

3.3.4 Relative efficiencies of multilateral optimisation algorithms

A test case was derived to assess and compare the settlement efficiency of the presented multilateral optimisation algorithms. We considered ten banks participating in a large value payment system and assumed that a severe operational problem affecting the payment system IT infrastructure had resulted in the unavailability of the banks' cash balances.[5] As a consequence, the cash position of every participant was considered to be zero until some fresh collateral was provided by the banks.

This liquidity shortage prevented a highly urgent 'all or nothing' ancillary system from settling. We assumed the net position of the banks within the ancillary system was as shown on the left part of Figure 3.4, with nine participants being equally long in the system with a net credit position of EUR 11 million, and only one short participant with a net debit position of EUR 100 million.

Figure 3.4 **Multilateral optimisation test case: settlement of an urgent Ancillary System in a LVPS**



The Central Bank operating the large value payment system wished to speed up as much as possible the settlement of the highly urgent ancillary system and to do so asked the participant with a net debit position in the ancillary system to provide some additional collateral. As fresh collateral might have been scarce in a period of crisis, the Central Bank was interested in trying to reduce the liquidity burden affecting the participant with a net debit position in the AS. To achieve this goal, the system operator could have made use of normal priority payments that were held in the queue due to the lack of available liquidity in the system. It is clear that simultaneously settling the pending AS with some normal priority payments from the long participants to the short participant could lower the amount of collateral the short participant has to find in order to be able to settle

---

[5] In the context of Target2, such situation could occur for example in case of a regional disaster.

the ancillary system. To do so, we can use a multilateral optimisation algorithm that 'locks the AS settlement', meaning that the algorithm can not settle any payment unless the highly urgent AS is settled simultaneously with it. Such an approach is in particular used in TARGET2, with algorithm 4[6] ('Partial optimisation with ancillary system').

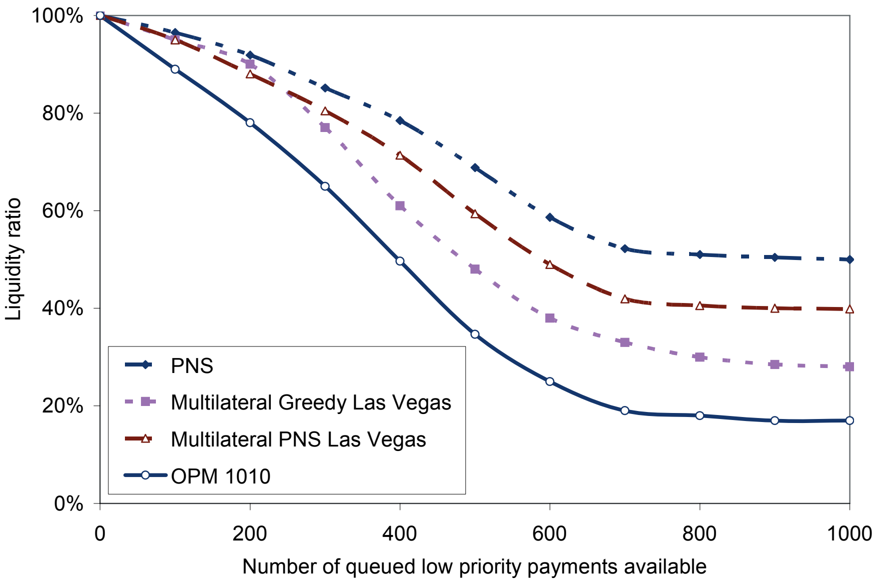
In our test case, we assumed a given number of low priority payments were queued between participants. The low priority payments were generated randomly according to the log normal law that describes the payments value distribution in the PNS system (see Section 3.2.5), and choosing the emitter and the receiver of the payments from the list of the participants with an equal probability.

This test case was run with the presented multilateral optimisation algorithms. The obtained results, averaged over 100 randomly generated low priority payment distributions, are shown in Figure 3.5. The liquidity ratio, defined as the ratio between the remaining collateral value that the short participant has to find, and its net debit position in the ancillary system (EUR 100 million), is plotted on the y-axis against the total number of available low priority payments at the beginning. As an example, in the graph provided in Figure 3.4, the obtained liquidity ratio is 38%. Clearly, when no low priority payments are present to offset the AS, the short participant has to provide the entire 100 millions and the liquidity ratio is one, whatever the algorithm used. When more low priority payments are available, the collateral needs of the short participant are reduced, to an extent that depends on the chosen algorithm.

---

[6] TARGET2 User Detailed Functional Specifications, first book version 2.0, page 161.

Figure 3.5 **Multilateral optimisation test case: Results**



The results clearly show the interest of multilateral optimisation for the settlement of ancillary systems. When many payments are available, the best algorithm is able to divide by five the value of fresh collateral the short participant has to provide. The best algorithm is OPM 1010, followed by the Multilateral Las Vegas Greedy Algorithm, the Multilateral Las Vegas PNS algorithm and the algorithm implemented in the PNS system.

3.4 Optimisation in PNS in case of an operational failure

3.4.1 The PNS system

PNS (Paris Net Settlement) is a French LVPS which operates alongside TBF, the French RTGS component of the TARGET system. It provides real-time settlement of transactions on central bank money accounts that must always remain in credit. In 2006, 17 banks and credit institutions were participating in the PNS system and exchanging an average of 27,000 payments on a daily basis, with a total value between EUR 45 and EUR 90 billion per day. A cash link established between PNS and TBF allows the participants to transfer liquidity between their TBF account and their PNS account at any time of the day, depending on their cash needs.

PNS is often presented as a hybrid system because it is equipped with efficient optimisation algorithms that are able to settle simultaneously several queued payments, thus allowing the system to operate at lower liquidity levels. The study of the PNS system is of special interest to the central banks of the Eurosystem because the algorithms implemented in the PNS system are extremely similar to the ones that will be used in the future pan-European TARGET2 system. Moreover as in TARGET2, bilateral sender limits (which can be defined as the maximum net amount a participant is willing to pay to another participant before being paid in return) can be set and modified freely by each participant of the PNS system vis-à-vis its counterparties.

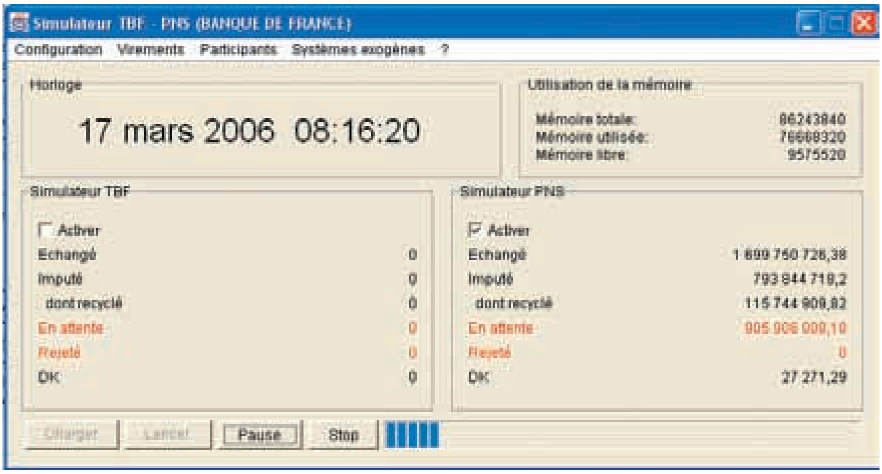
3.4.2 Simulating the technical default of the largest participant in the system

Following Banque de France's previous paper on the PNS system,[7] we investigated the role of optimisation mechanisms under special crisis circumstances. A previous study showed that an operational problem preventing a major participant from issuing payments could lead to a liquidity shortage within the PNS system and finally to the rejection of several queued payments at the closure of the system. Indeed, as the biggest participant is still able to receive payments, but can no longer issue payments, it turns into a 'liquidity trap', depriving the system of the liquidity needed to settle the pending payments. The settlement delay thus increases and eventually some payments can even be rejected at the end of the day. Being able to use advanced non-FIFO algorithms at this point could allow a significant reduction in the number of rejected payments. The case of the technical default of the biggest participant in PNS was therefore revisited after the algorithms presented in this paper had been implemented in Banque de France's PNS/TBF simulator.

---

[7] Analysis by simulation, of the impact of a technical default of a payment system participant (Liquidity, risks and speed in payment and settlement systems – a simulation approach, Bank of Finland Studies, 2005), Mazars, E and Woefel, G.

Figure 3.6 **Screenshot of Banque de France's PNS simulator**



17 mars 2006 08:16:20

Mémoire totale:	86243840
Mémoire utilisée:	76668320
Mémoire libre:	9575520

Simulateur TBF		Simulateur PNS	
☐ Activer		☑ Activer	
Echangé	0	Echangé	1 899 750 726,38
Imputé	0	Imputé	793 844 718,2
dont recyclé	0	dont recyclé	115 744 908,82
En attente	0	En attente	905 906 008,10
Rejeté	0	Rejeté	0
OK	0	OK	27 271,29

The simulator used for those simulations is able to reproduce the exact functioning of PNS (bilateral limits, optimisation algorithms, liquidity transfers between PNS and the TBF) and processes the operations one by one as in the real system. In contrast to the test cases presented in Section 3.2.5 (bilateral optimisation) and Section 3.3.4 (multilateral optimisation), real transactions data was used this time.

3.4.3 Results

The month of March 2006 was selected and for each day of the month the consequences of the technical default of the largest participant were investigated with Banque de France's PNS simulator. We assumed that the other participants would not retain their payments in reaction to the technical default of the biggest participant and would not change their behaviour in any way. The most severe consequences were observed for 17 March. Indeed, on this day, provided the default had no influence at all on the behaviour of the other participants, the technical default of the biggest participant would have resulted in 32 payments, representing a total value of EUR 14 billions or 28% of the total value of the submitted payments being rejected at the end of the day. The consequences of the technical default of the biggest participant appear, therefore, to be extremely strong. In reality, however, it is likely that the non-defaulting participants would have tried to mitigate the consequences of the crisis by injecting more liquidity into the system, thus reducing the number and value of the rejected payments.

The potential impact of the implementation of the presented advanced algorithms into the PNS system as replacements for the original algorithms was investigated with the PNS simulator for 17 March 2006.

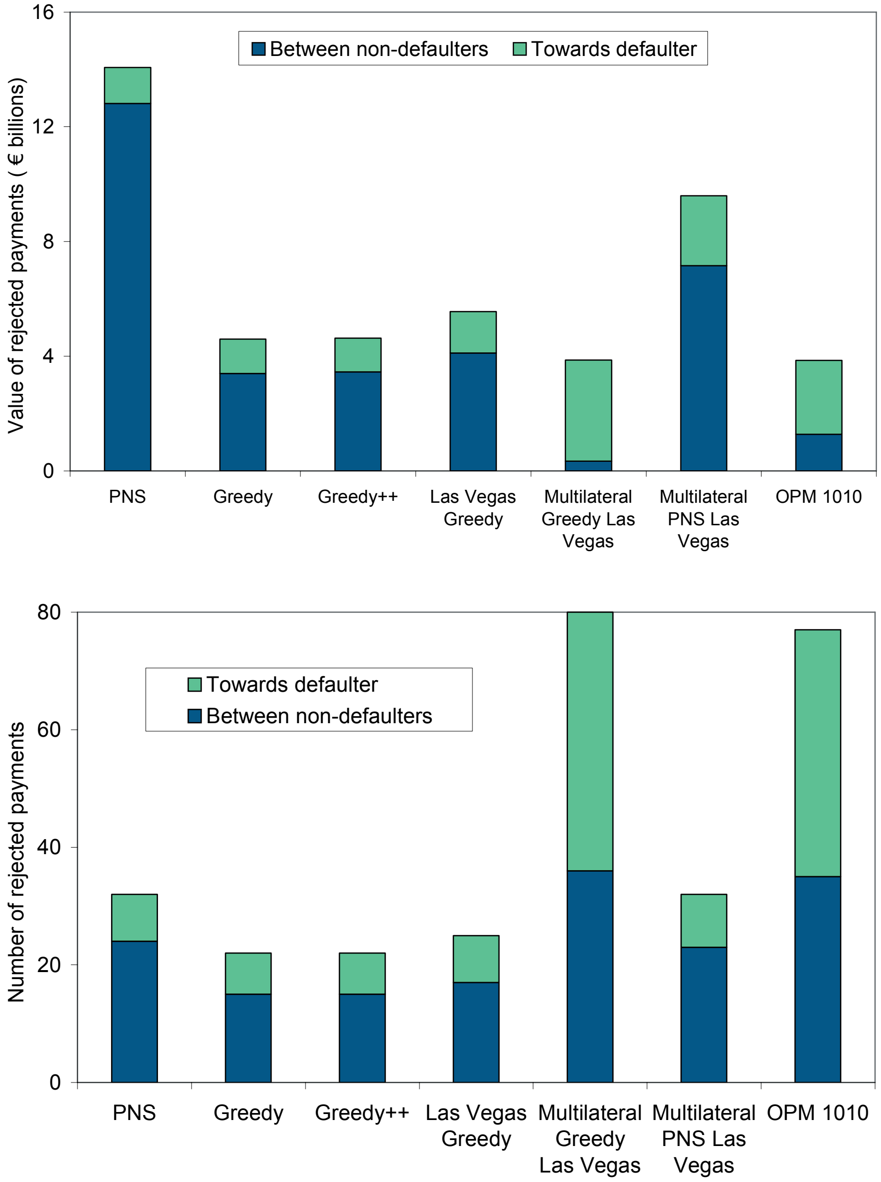
Simulations were made using:

- PNS bilateral optimisation algorithm and PNS multilateral optimisation algorithm (pure PNS);
- Greedy bilateral optimisation algorithm and PNS multilateral optimisation algorithm;
- Greedy++ bilateral optimisation algorithm and PNS multilateral optimisation algorithm;
- Las Vegas Greedy bilateral optimisation algorithm and PNS multilateral optimisation algorithm;
- PNS bilateral optimisation algorithm and the multilateral Las Vegas Greedy algorithm;
- PNS bilateral optimisation algorithm and the multilateral Las Vegas PNS algorithm;
- PNS bilateral optimisation algorithm and the OPM 10–10 algorithm.

The algorithms making use of random numbers for optimisation (bilateral Las Vegas Greedy, Multilateral Las Vegas Greedy, Multilateral PNS Las Vegas) were run five times. No significantly better results were found by increasing the number of iterations.

Figure 3.7 shows the impact of the various optimisation algorithms on the number and total value of the payments rejected at the end of the day. In this given case, it appears that non-FIFO algorithms presented in this paper perform significantly better than the algorithms used in the PNS system. We can also note that the chosen algorithm can also significantly shift the outcome of the settlement, either towards an outcome less favourable to the defaulter (with the multilateral Greedy Las Vegas algorithm) or characterised by a decreased average value of rejected payments (with OPM 1010 or the Multilateral Greedy Las Vegas algorithm).

Figure 3.7 **Effect of the technical default of the biggest participant in the PNS system. Rejected payments towards the defaulter (light) and between non-defaulters (dark) at the end of the day according to the algorithms implemented, in terms of value (top) and volume (bottom).**



The influence of optimisation algorithms on the settlement delay is shown Figure 3.8. In normal conditions (ie without any technical default), the use of non-FIFO optimisation algorithms lowered the settlement delay by about 50% in terms of value, while the settlement delay in terms of volume remained constant. Multilateral optimisation algorithms have a much smaller influence on the settlement delay, as in PNS the multilateral optimisation algorithm is called only three times a day, at 10:30, 12:30 and 16:00. When the biggest participant defaults, the edge given by the non-FIFO algorithms in terms of value becomes significantly bigger.

3.4.4 Payments rejected at the end of closure

In order to provide the reader with a clearer insight of the effect of the optimisation algorithms on settlement efficiency, Table 3.1 presents the list of payments rejected between two participants in the PNS system, designated here as participant A and participant B. It appears that the cumulated value of rejected payments between those two participants is extremely high, and represents the main part of the total value of rejected payments.

The PNS bilateral optimisation algorithm is unable to settle any of those payments, given the cash balances of participants A and B. However, it is easy to see that the Greedy bilateral algorithm will simultaneously settle the payments with a value of EUR 1,500 and EUR 2,000 million from A to B and EUR 3,500 million from B to A. In this situation, the Greedy++ algorithm will simultaneously settle the payment with a value of EUR 3,500 million from B to A and the payments with a value of EUR 313, EUR 956, EUR 2,000, EUR 51 and EUR 180 million. As none of those solutions complies with the FIFO rule, the PNS FIFO bilateral optimisation algorithm will not consider them.

In this situation, the use of an advanced optimisation algorithm results in a reduction of EUR 7 billion reduction in the total value of the payments rejected at the end of the day.

Figure 3.8 **Effect of the technical default of the biggest participant in the PNS system. Settlement delay in terms of value (top) and volume (bottom), in normal conditions (dotted line) and in case of the technical default of the biggest participant (solid line).**

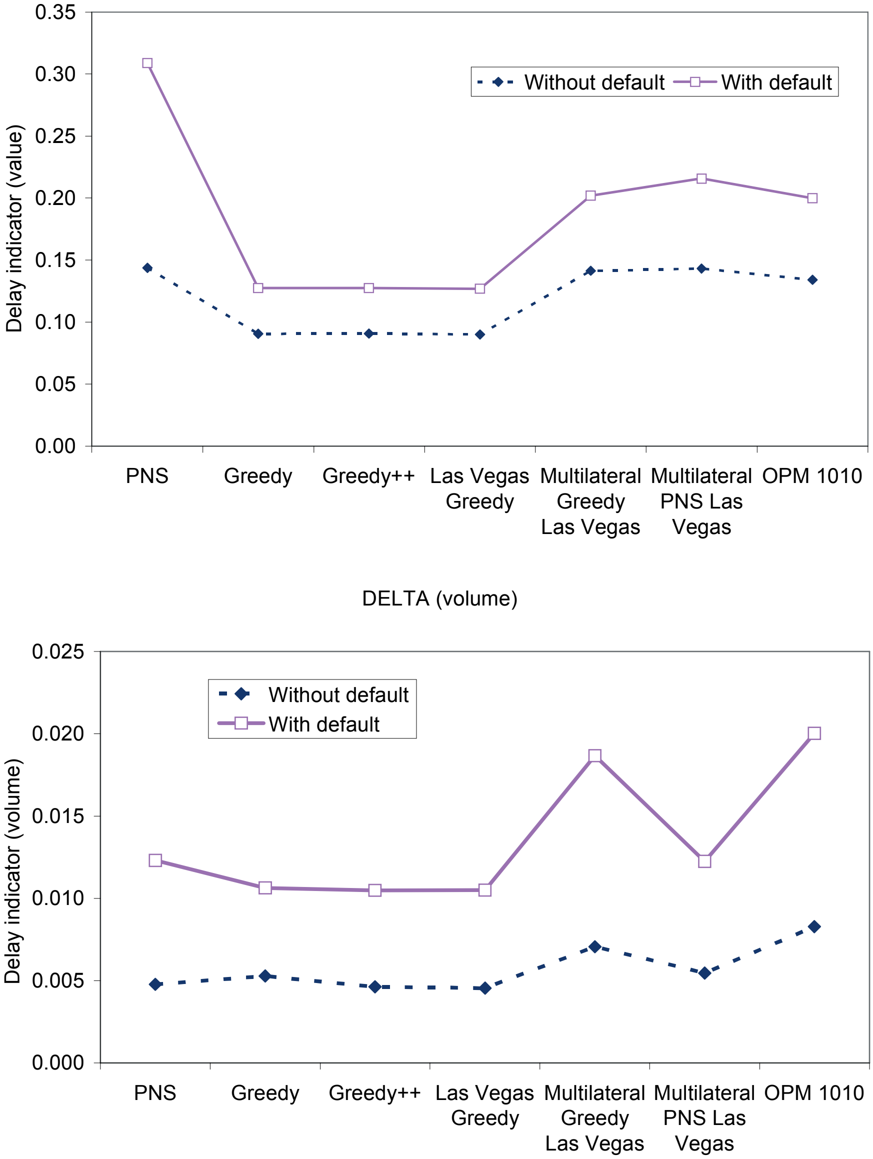


Table 3.1 **Simulation of the technical default of the biggest participant in the PNS system (17 March 2006, standard PNS algorithms), rejected payments between two selected participants.**

		Participant A	Participant B
Cash balance in EUR at closure		3.5 million	22.5 million
Queued payments between participants A and B rejected at closure, in order of arrival	1	160 million	1,000 million
	2	313 million	3,500 million
	3	956 million	87 million
	4	1,500 million	
	5	2,000 million	
	6	51 million	
	7	180 million	

3.5 Conclusion

In this paper several optimisation algorithms that do not follow the FIFO constraint, ie algorithms that are allowed to settle queued payments irrespective of their order of arrival, were presented and their efficiency was investigated in several tests. The results of these tests suggest that the simple Greedy algorithm of Güntzer et al and the suggested OPM1010 algorithm are able to improve respectively on their bilateral and multilateral FIFO counterparts.

Of course, the choice of an optimisation algorithm in a RTGS involves many other considerations than the mere settlement efficiency of the algorithms. In particular, the rules of the system have to be legally sound and have to match the needs of the users as much as possible. No definitive conclusion regarding the use of non-FIFO algorithms in RTGS can therefore be drawn from this paper.

The standard case of the technical default of the biggest participant in a RTGS was also revisited in the context of the PNS system and with several different optimisation algorithms. On the business day chosen for this exercise, (chosen as the ‘worst day’ of the month of March 2006 in terms of rejected payments resulting from the technical default of the biggest participant), the use of non-FIFO algorithms was shown to greatly reduce the value of rejected payments at the end of the day while shortening the settlement delay. However, when the same exercise was carried out for certain other days of the same month, the use of non-FIFO algorithms did not bring any improvement. It was even the case that the use of non-FIFO algorithms led to a slight deterioration of the situation at the end of the day. This is due to the fact that efficient algorithms tend to settle payments earlier, as shown by Figure 3.8. Sometimes a slightly less efficient algorithm stockpiles many payments in the queue during the day and is then able to profit from the optimisation opportunities created by the large number of queued payments, resulting in better end-of-the-day results.

This observation having been made, it could make sense to imagine an RTGS in which FIFO algorithms, which combine the advantages of being fast, reasonably efficient, predictable and perfectly transparent to the users, would be used throughout the day, while some more advanced, non-FIFO algorithms could be used in case of a liquidity shortage. In the case of TARGET2, for example, those algorithms would be launched at the closure of the system in case some payments remain in the queue. If the advanced algorithms are then able to settle some additional payments, the number and cumulated value of the rejected payments would be lowered. In the opposite case, the use of those algorithms would not have affected the functioning of the system.

Advanced non-FIFO algorithms could also be useful to accelerate the settlement of a highly urgent ancillary system in the context of a liquidity shortage, as presented in Section 3.3.4. Such specially designed non-FIFO algorithms would only be run in case the standard AS settlement procedure has failed and the settlement delay is creating concerns.

References

Bech, M – Soramäki, K (2001) **Gridlock resolution in interbank payment systems.** Discussion Paper 9/2001. Bank of Finland.

Bech, M – Garratt, R (2003) **The intraday liquidity management game.** Journal of Economic Theory 109, 198–219.

Bech, M – Soramäki, K (2005) **Gridlock resolution and bank failures in interbank payment systems, in Liquidity, risks and speed in payment and settlement systems – a simulation approach.** Bank of Finland Studies E:31.

Güntzer, M – Jungnickel, D – Leclerc, M (1998) **Efficient algorithms for the clearing of interbank payments.** European Journal of Operational Research 106, 212–219.

Koponen, R – Soramäki, K (2005) **Intraday liquidity needs in a modern interbank payment system – a simulation approach.** In Liquidity, risks and speed in payment and settlement systems – a simulation approach. Bank of Finland Studies E:31.

Leinonen, H – Soramäki, K (2005) **Optimising liquidity usage and settlement speed in payment systems.** In Liquidity, risks and speed in payment and settlement systems – a simulation approach. Bank of Finland Studies E:31.

Shafransky, Y – Doudkin, A (2006) **An optimisation algorithm for the clearing of interbank payments.** European Journal of Operational Research 171, 743–749.

Appendix 1

Bilateral optimisation algorithms

Bilateral optimisation: notations

We denote the payments from bank A to bank B (respectively from bank B to bank A) as the (a_i) and the (b_i). The (x_i) and the (y_i) are two vectors of indicators. For each k, $x_k = 0$ (resp. $x_k = 1$) means that the payment a_k is not activated (resp. activated); similarly for each k, $y_k = 0$ (resp. $y_k = 1$) means that the payment b_k is not activated (resp. activated).

S_A is the initial cash balance of bank A and S_B is the initial cash balance of bank B. For given (x_i) and (y_i), bank A's virtual cash balance is equal to $B_A = S_A - \sum_i a_i x_i + \sum_i b_i y_i$ and bank B's virtual cash balance is $B_B = S_B + \sum_i a_i x_i - \sum_i b_i y_i$.

FIFO bilateral optimisation algorithm (PNS, TARGET2 …)

```
Activate all payments between the two considered banks.
WHILE the simultaneous settlement of all activated payments is impossible
    • De-activate the most recent activated payment from the deficient bank.
END WHILE
Settle all activated payments
```

LAS VEGAS GREEDY bilateral optimisation algorithm

```
Activate all payments between the two considered banks.

WHILE one of the two banks has a negative Virtual Cash Balance
    • De-activate all payments from the deficient bank (let us suppose it is bank A).
    • Go through the payments of bank A, the $(a_i)_{i=1...N}$ from A's biggest payment $a_1$ to A's smallest payment $a_N$. When considering payment $a_k$ for re-activation:
        o IF $a_k$ > Bank A's Virtual Cash Balance THEN $a_k$ can not be re-activated.
```

```
    o ELSE:
        • Calculate $R_k = \sum_{i=k+1}^{N} a_i$ sum of all the payments smaller than $a_k$
        • IF $R_k$ < Bank A's Virtual Cash Balance THEN $a_k$ is activated
        • ELSE:
            • Let $p_k = \min\left(\frac{a_k}{R_k};1\right)$
            • Re-activate payment $a_k$ with a probability of $p_k$
    o Next $a_k$
END WHILE

Settle all activated payments.
```

The point of this algorithm is to launch it several times. In the tests presented in this paper, it was applied 5 times in a row (ie it was applied a first time to the initial problem, then it was applied a second time to what had not settled the first time, and so on). Another possible use is to run it a certain number of times on the initial problem and to retain the best solution.

GREEDY bilateral optimisation algorithm (Güntzer et al, 1998)

```
Activate all payments between the two considered banks.

WHILE one of the two banks has a negative Virtual Cash Balance
• De-activate all payments from the deficient bank (let us suppose it is bank A).
• Go through the payments of bank A, the $(a_i)_{i=1...N}$ from A's biggest payment $a_1$ to A's smallest payment $a_N$. When considering payment $a_k$ for re-activation:
    o IF $a_k$ > Bank A's Virtual Cash Balance THEN $a_k$ can not be re-activated.
    o ELSE re-activate $a_k$
    o Next $a_k$
END WHILE

Settle all activated payments
```

GREEDY++ bilateral optimisation algorithm

```
Activate all payments between the two considered banks, bank A and bank B.

WHILE one of the two banks has a negative virtual cash balance

• Let $G = \sum_i a_i x_i$ and $H = \sum_i b_i y_i$
• The error is defined as: $\Delta = \left| G - H - \frac{1}{2}(S_A - S_B) \right|$
• Pick up the 10 payments (from either bank, selected or not) closest to the
  error $\Delta$ (we pick up the $a_i$ and $b_i$ that minimise $\left| \log\left(\frac{a_i}{\Delta}\right) \right|$)
• Try all possibilities involving the 10 picked up payments (1024
  possibilities)
• IF at least one of the possibilities allows settlement
    o THEN choose the possibility that maximises the value settled.
    o ELSE:
        • De-activate all payments from the deficient bank (let us
          suppose it is bank A).
        • Go through the payments of bank A, the $(a_i)_{i=1...N}$ from A's
          biggest payment $a_1$ to A's smallest payment $a_N$. When
          considering payment $a_k$ for re-activation:
            • IF $a_k$ > Bank A's Virtual Cash Balance THEN $a_k$ can
              not be re-activated.
            • ELSE re-activate $a_k$
            • Next $a_k$

END WHILE
Settle all activated payments
```

Appendix 2

Multilateral optimisation algorithms

Multilateral PNS Las Vegas

```
Activate all payments

1. Attempt to settle all queued payments simultaneously ('all or nothing')

2. As long as there is a peripheral participant:
   De-activate all payments to or from a peripheral participant.

3. WHILE there is a participant with a negative Virtual Cash Balance

   3.1 Randomly choose a participant with a negative Virtual Cash Balance (Uniform law).
   3.2 The chosen participant, bank i, has a negative Virtual Cash Balance $B_i$. We are then going to de-activate one of bank i's outgoing payments. In order to do so, for each activated payment k emitted by bank i, we calculate the coefficient $b_{i,k} = \gamma^{suf}_{i,k}\gamma^{def}_{i,k}\gamma^{cre}_{i,k}$ where:
       • $\gamma^{suf}_{i,k} = 2$ if the inactivation of payment k makes bank i's Virtual Cash Balance positive, else $\gamma^{suf}_{i,k} = 1$.
       • $\gamma^{def}_{i,k} = \max\left[\min\left(\frac{|B_i|}{p_i^k};\frac{p_i^k}{|B_i|}\right);0.1\right]$ so that payments whose value are close to the deficit are de-activated with a higher probability.
       • $\gamma^{cre}_{i,k} = 4$ if the inactivation of payment k does not create nor aggravate the deficit of another participant, else $\gamma^{cre}_{i,k} = 1$.
       • We then randomly select one of bank i's outgoing payments so that payment k has a probability $\frac{b_{i,k}}{\sum_k b_{i,k}}$ to be de-activated.
   3.3 If bank i now has a positive Virtual Cash Balance, attempt to re-activate some of bank i's outgoing payments in the order of their decreasing amount.

END WHILE

4. When all participants have a positive Virtual Cash Balance, all activated payments are settled.
```

Re-activate all de-activated payments (including those involving peripheral participants) and go through all bilateral relations, from the most balanced to the most unbalanced and run the Las Vegas Greedy bilateral optimisation algorithm.

The point of this algorithm is to launch it several times. In the tests presented in this paper, it was applied 5 times in a row (ie it was applied a first time to the initial problem, then it was applied a second time to what had not settled the first time, and so on). Another possible use is to run it a certain number of times on the initial problem and to retain the best solution.

Multilateral Greedy Las Vegas

Activate all payments

1. Attempt to settle all queued payments simultaneously (‘*all or nothing*’)

2. As long as there is a peripheral participant:
   De-activate all payments to or from a peripheral participant.

3. WHILE there is a participant with a negative Virtual Cash Balance

   3.1. Consider all the banks with a negative Virtual Cash Balance in the increasing order of the number of participants they send payments to, then in the decreasing order of their deficit (we then start by the banks which emit payments towards a single counterparty).

   3.2 The considered bank i, has a negative Virtual Cash Balance B_i. De-activate all its outgoing payments, then consider them for re-activation in the decreasing order of their value, under the constraint that the virtual position of bank i remains positive. We then have, for payment number l of bank i, p_i^l:
   - a. IF $p_i^l > B_i$ then payment number l can not be re-activated.
   - b. ELSE IF $R_i^l < B_i$ then payment number l is re-activated (where $R_i^l = \sum_{k \geq l+1} p_i^k$ is the cumulated value of Bank i’s payments smaller than p_i^l)

c. ELSE the payment p_i^l is re-activated with a probability equal to $\min\left(\frac{p_i^l}{R_i^l} b^{\pm};1\right)$, where:

- $b^{+} = \frac{m^{+} + m^{-}}{2m^{+} + m^{-}}$ $\quad b^{-} = \frac{m^{+} + m^{-}}{m^{-}} b^{+}$
- m^{+} (resp. m^{-}) is the number of participants receiving payments from bank i whose Virtual Cash Balance is positive (respectively negative).
- $b^{\pm} = b^{+}$ if the receiver of the payment has a positive Virtual Cash Balance, else $b^{\pm} = b^{-}$.

END WHILE

4. When all participants have a positive Virtual Cash Balance, all activated payments are settled.

Re-activate all de-activated payments (including those involving peripheral participants) and go through all bilateral relations, from the most balanced to the most unbalanced and run the Las Vegas Greedy bilateral optimisation algorithm.

The point of this algorithm is to launch it several times. In the tests presented in this paper, it was applied 5 times in a row (ie it was applied a first time to the initial problem, then it was applied a second time to what had not settled the first time, and so on). Another possible use is to run it a certain number of times on the initial problem and to retain the best solution.

Algorithm OPM 1010

```
Activate all payments

1. Attempt to settle all queued payments simultaneously ('all or nothing')

2. As long as there is a peripheral participant:
   De-activate all payments to or from a peripheral participant.

3. WHILE there is a participant with a negative Virtual Cash Balance

   3.1 Go through all banks with a positive Virtual Cash Balance and re-activate the payments that can be re-activated.
   3.2 Randomly choose a bank with a negative Virtual Cash Balance.
       WHILE the chosen bank i has a negative Virtual Cash Balance $B_i$:
           Calculate for each activated payment k sent by bank i, the coefficient $b_{i,k} = \gamma^{suf}_{i,k}\gamma^{def}_{i,k}\gamma^{cre}_{i,k}$ where:
           • $\gamma^{suf}_{i,k} = A$ if the inactivation of payment k makes bank i's virtual position positive, else $\gamma^{suf}_{i,k} = 1$.
           • $\gamma^{def}_{i,k} = \max\left[\min\left(\frac{|B_i|}{p_i^k};\frac{p_i^k}{|B_i|}\right);0.1\right]$ in order to favour the payments whose value is close to $|B_i|$ the net debit position of bank i.
           • $\gamma^{cre}_{i,k} = C$ if the inactivation of payment k does not create nor aggravate the deficit of another participant, else $\gamma^{cre}_{i,k} = 1$.
           • A sensitivity study performed at several levels of liquidity concluded that the OPM alrorithm have better results with A = C = 10, hence the name of OPM1010 for this given variation of the algorithm.
           De activate the payment with the highest coefficient $b_{i,k}$.

       END WHILE

       If there are some of bank i's de-activated payments can be re-activated, re-activate them in the decreasing order of their value.

END WHILE

4. When all participants have a positive Virtual Cash Balance,
   4.1 Go through all banks and re-activate the payments that can be re-activated in the decreasing order of their value.
```

5. Settle all activated payments.

Re-activate all de-activated payments (including those involving special participants) and go through all bilateral relations, from the most balanced to the most unbalanced and run the Las Vegas Greedy bilateral optimisation algorithm.

The point of this algorithm is to launch it several times. In the tests presented in this paper, it was applied 5 times in a row (ie it was applied a first time to the initial problem, then it was applied a second time to what had not settled the first time, and so on). Another possible use is to run it a certain number of times on the initial problem and to retain the best solution.

Appendix 3

The Greedy algorithm of Güntzer et al[8] and superincreasing payment values distributions

Superincreasing sequences

Let p be a strictly positive integer. A sequence of positive reals $(u_i)_{i=1...p} \in \Re^p$ is said to be superincreasing when: $\forall k \in \{1...p-1\}$, $u_{k+1} > \sum_{i=1}^{k} u_i$. For a central banker, a good example of a superincreasing sequence is the sequence of the values of the euro banknotes (5 euros, 10 euros, 20 euros, 50 euros, 100 euros, 200 euros and 500 euros). Indeed, any banknote is worth more than the sum of the smaller banknotes. This highly desirable property ensures that a cashier can minimise the number of banknotes to be given back to a customer by simply following a Greedy type of algorithm, that is to say by always using the biggest banknote whose value is lower than the remaining amount of money to be handed back. Should a 400-euro banknote be introduced, the Greedy solution (500+200+100, 3 banknotes) would be beaten by a non-Greedy solution (400+400, 2 banknotes) if 800 euros had to be handed back by the cashier. This property is actually closely related to the aim of this demonstration.

Notations

- Let there be two banks A and B, characterised by their respective liquidity S_A and S_B. There are N queued payments from A to B and M queued payments from bank B to bank A.

- We assume that the sequences of the queued payments from A to B and from B to A, respectively the $(a_i)_{i=1...N}$ and $(b_i)_{i=1...M}$ are superincreasing sequences, that is to say that we have $\forall i \in \{1...N-1\}, a_i > \sum_{k=i+1}^{N} a_k$ and $\forall i \in \{2...M-1\}, b_i > \sum_{k=i+1}^{M} b_k$ (a_1 is

---

[8] Güntzer, M – Jungnickel, D – Leclerc M (1998) Efficient algorithms fort he clearing of interbank payments. European Journal of Operational Research 106, 212–219.

therefore the biggest payment from bank A to bank B, and a_N is the smallest).

- The $(x)_{i=1...N} \in \{0,1\}^N$ and the $(y)_{i=1...M} \in \{0,1\}^M$ are two vectors of indicators. For each k, $x_k = 0$ (resp. $x_k = 1$) means that the payment a_k is not activated (resp. activated); similarly for each k, $y_k = 0$ (resp. $y_k = 1$) means that the payment b_k is not activated (resp. activated).
- The Greedy algorithm is as defined in Appendix 1.

Lemma

Provided the sequence of the payment values is superincreasing, the Greedy algorithm re-activates at each iteration the payments whose cumulated value is maximal.

Proof

Without any loss of generality, we can assume that the bank in deficit is bank A. At the beginning of an iteration, all payments emitted by bank A are de-activated and are then considered for re-activation in the decreasing order of their value. It is clear that the total cumulated value of the re-activated payments can not exceed a ceiling of $S_A + \sum_{i=1}^{M} b_i y_i$ where y_i indicates whether the i^{th} payment of bank B is activated.

The Greedy algorithm first considers bank A's biggest payment a_1 for re-activation. If $a_1 > S_A + \sum_{i=1}^{M} b_i y_i$ then a_1 can clearly not be re-activated, whatever the algorithm used. Let us now suppose that $a_1 \leq S_A + \sum_{i=1}^{M} b_i y_i$, the Greedy algorithm will therefore re-activate payment a_1. Any algorithm which would choose not to re-activate this payment would yield a poorer solution than Greedy's since as $a_1 > \sum_{k=2}^{N} a_k$ (because the sequence is superincreasing), any solution not retaining a_1 would be worse than any solution retaining a_1.

By induction, the same result applies to all of bank A's payments, hence we can conclude that for a given iteration, the value of the payments re-activated by the Greedy algorithm is maximal.

Proposition

The Greedy algorithm is the most efficient in terms of settled payment value provided the sequence of the payment values is superincreasing.

Proof

Let $G_k = \sum_{i=1}^{N} a_i x_i$ be the cumulated value of the activated payments from A to B after the kth iteration of the *Greedy* algorithm where bank A is in deficit. Similarly let $H_k = \sum_{i=1}^{M} b_i y_i$ be the cumulated value of the activated payments from B to A after the kth iteration of the *Greedy* algorithm where bank B is in deficit.

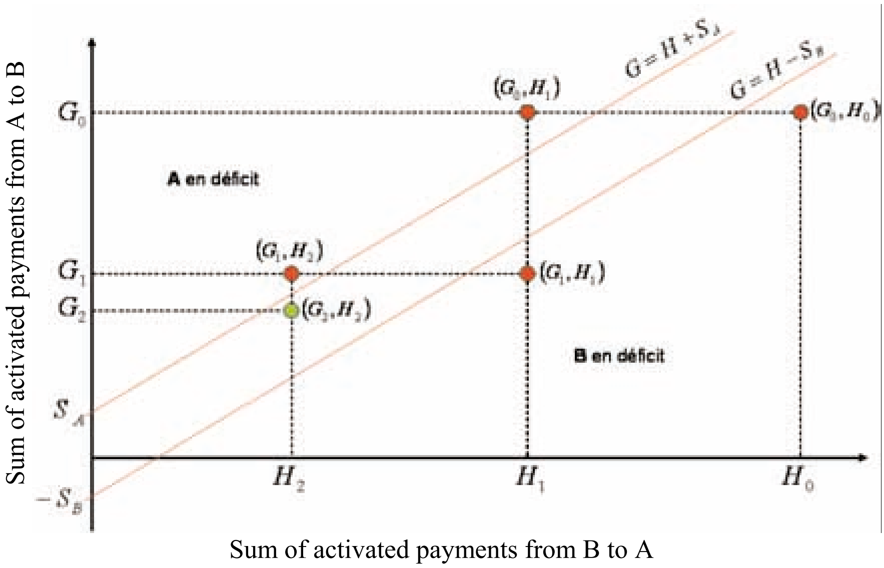
The settlement condition can be written as the dual inequality: $-S_B \leq G - H \leq S_A$.

At the start of the algorithm, all payments are activated, hence $G_0 = \sum_{i=1}^{N} a_i$ and $H_0 = \sum_{i=1}^{M} b_i$. Without any loss of generality, we can assume that bank B will be the first bank to be in deficit. The pairs of payment flows that will be considered will then be: (G_0, H_0), (G_0, H_1), (G_1, H_1) ...

It is easy to demonstrate that the Greedy algorithm will terminate, by noticing that the G_k and H_k are two strictly decreasing sequences taking only a finite number of positive values. We denote as t the subscript of the last iteration of the Greedy algorithm. The final state will therefore be either (G_t, H_t) or (G_{t-1}, H_t).

Given the characteristics of the Greedy algorithm, we already know that we will have: $G_0 > G_1 > .... > G_t$ and $H_0 > H_1 > .... > H_t$. The underneath sketch shows how the pairs (G_k, H_k) converge towards a solution satisfying the settlement condition (the pairs satisfying the settlement condition are located between the two parallel red lines).

Figure 3.9 **Illustration of the convergence of the Greedy algorithm**



Let G* and H* be the values of the payment flows characterising the solution maximising the settled value. This value maximising solution trivially exists (at worse we have G* = H* = 0).

Let us show by induction that $\forall k \in \{0...t\}, \begin{cases} G_k \geq G^* \\ H_k \geq H^* \end{cases}$

Basis: trivially, we have $\begin{cases} G_0 \geq G^* \\ H_0 \geq H^* \end{cases}$

Inductive step: Let be $k \in \{0...t\}$. Suppose that $\begin{cases} G_k \geq G^* \\ H_k \geq H^* \end{cases}$

As we assumed that B was initially in deficit, after k iterations on G and k iterations on H, B is still the bank in deficit. Greedy then has to evaluate the new cumulated payment flows of bank B, H_{k+1}. Let us show that $H_{k+1} \geq H^*$.

According to the lemma, H_{k+1} is the highest possible value that can take the cumulated sum of the activated payments of bank B under the constraint: $H_{k+1} \leq S_B + G_k$.

Now we also have $G_k \geq G^*$ according to our inductive hypothesis and we know in addition that H^* verifies the settlement condition $H^* \leq S_B + G^*$, since the pair (G^*, H^*) is a solution to the problem. That gives us the inequality $H^* \leq S_B + G_k$ and as H_{k+1} is the highest possible value lower than $S_B + G_k$ we can then conclude that $H_{k+1} \geq H^*$.

Now A is in deficit and the same demonstration applies to prove that $G_{k+1} \geq G^*$. We can then conclude.

We have then shown that $\forall k \in \{0...t\}, \begin{cases} G_k \geq G^* \\ H_k \geq H^* \end{cases}$

In particular we have $\begin{cases} G_t \geq G^* \\ H_t \geq H^* \end{cases}$

G* et H* being by construction the best solution, we have $\begin{cases} G_t = G^* \\ H_t = H^* \end{cases}$

When the payment value sequences are superincreasing, the Greedy algorithm thus yields the solution that maximises the settled value.

Chapter 4

Examining the tradeoff between settlement delay and intraday liquidity in Canada's LVTS: a simulation approach

Neville M Arjani

4 Examining the tradeoff between settlement delay and intraday liquidity in Canada's LVTS: a simulation approach

Abstract

The paper explores a fundamental tradeoff occurring in the daily operation of large-value payment systems (LVPS) – between settlement delay and intraday liquidity – with specific application to Canada's Large-Value Transfer System (LVTS). To reduce settlement delay, participants generally must maintain greater intraday liquidity in the system. Intraday liquidity and settlement delay can be costly for LVPS participants, and improvements in the tradeoff are desirable. The replacement of standard queuing arrangements with a complex queue-release algorithm represents one such improvement. These algorithms are expected to lower intraday liquidity needs and speed up payment processing in an LVPS. Simulation analysis is used to empirically test this proposition for the case of Canada's LVTS. The analysis is conducted using a payment system simulator developed by the Bank of Finland, called the BoF-PSS2. It is shown that increased use of the LVTS central queue (which contains a complex queue-release algorithm) reduces settlement delay associated with each level of intraday liquidity considered, relative to a standard queuing arrangement. Some important issues for discussion emerge from these results.

4.1 Introduction

A well-functioning large-value payment system (LVPS) is an integral component of any advanced financial system. In a market economy such as Canada's, virtually all economic transactions ultimately involve a transfer of funds between a buyer and a seller. An LVPS provides the electronic infrastructure necessary to facilitate such an exchange of funds between financial institutions in order to discharge large-value payment obligations on behalf of their own business and that of their customers. There are different designs of LVPS currently operating around the world, with each achieving a different balance between the minimisation of systemic risk, the speed of payment settlement, and the liquidity and operational costs of settlement.

This paper examines a fundamental tradeoff occurring in the daily operation of an LVPS – between settlement delay and intraday liquidity – with particular application to Canada's LVTS.[1] Settlement delay refers to a potential time lag occurring between a participant's intended submission of a payment to the system and when it is processed by the LVPS with finality.[2] Intraday liquidity refers to a participant's ability to meet its outgoing payment obligations immediately when intended. Generally speaking, to achieve shorter settlement delay participants must maintain greater intraday liquidity in the system. When sufficient intraday liquidity is not maintained, payments will be queued and will be released only when the participant's liquidity position improves. Settlement delay, then, reflects the amount of time that a payment is queued before being processed by the system.

Intraday credit is an important source of liquidity. To control credit risk, grantors of intraday credit (typically central banks) usually require eligible collateral, which is likely to entail a cost for participants. At the same time, settlement delay may also be expensive for participants. The cost of settlement delay may be borne both internally by the participant that delays sending the payment and externally by the receiving participant. Participants generally must tradeoff the cost of settlement delay and the cost of intraday liquidity in conducting their daily payment operations. It follows that a reduction in the amount of intraday credit provision to participants will entail both a benefit and cost. The benefit is that participants' liquidity (ie collateral) cost can be reduced, but possibly only at the expense of a higher settlement delay cost.

A simple graphical framework of the general risk-efficiency tradeoff in payment systems, inspired by Berger, Hancock and Marquardt (1996), is useful when thinking about the nature of the tradeoff between settlement delay and intraday liquidity in an LVPS. Given the cost to participants of both settlement delay and intraday

---

[1] The LVTS is owned and operated by the Canadian Payments Association (CPA). For a more thorough description of the LVTS, including an overview of the Bank of Canada's multiple roles within the system, see Dingle (1998) and Arjani and McVanel (2006).

[2] Use of the term 'intended' is made so that this definition of settlement delay could apply to LVPS designs with and without a central queue. Under the latter design, a participant may intend to submit a payment to the LVPS at a certain time but, due to lack of intraday liquidity and the absence of a central queue, must hold the payment internally until it can be successfully processed by the system.

liquidity, improvements in the tradeoff are desirable. An improvement in the tradeoff is characterised by this paper as reduced settlement delay associated with each level of intraday liquidity, for the same value of payment activity. Innovations in LVPS design may make this possible. The replacement of standard queuing arrangements with a complex queue-release algorithm represents such an innovation. The potential benefit of such algorithms includes both lower liquidity needs for the release of queued payments and thus faster processing of these payments by the LVPS.

A simulation approach is used to empirically test the proposition that a complex queue-release algorithm can lower liquidity costs and speed payments processing relative to a standard queuing arrangement – that is, improve the tradeoff between settlement delay and intraday liquidity. Using actual intraday transaction and credit limit data, simulation analysis is employed to quantify the current tradeoff between settlement delay and intraday liquidity in the Canadian LVTS. Then, improvements in this tradeoff are sought by simulating an alternative LVTS environment in which current restrictions on use of the LVTS central queue are relaxed. The LVTS queue employs a complex queue-release algorithm that seeks to partially offset batches of queued payments on a multilateral basis throughout the day. However, under current system rules, participants' excessive use of the central queue is not encouraged.[3] Instead, standard internal queuing arrangements are typically employed by participants.

The analysis reveals that a tradeoff does indeed exist between settlement delay and intraday liquidity in Canada's LVTS. Moreover, the results indicate that increased use of the central queue will reduce settlement delay in the LVTS for each level of intraday liquidity considered according to three different settlement delay measures. Some important discussion points also emerge from these results.

The remainder of this paper is as follows. Section 4.2 discusses the nature of the tradeoff between settlement delay and intraday liquidity in greater detail. The graphical framework is presented in Section 4.3, and potential improvements in the tradeoff are also discussed in that section. Section 4.4 contains relevant background information on the LVTS. Section 4.5 provides an overview of the simulation

---

[3] See LVTS Rule No. 7. There are several hypothesised reasons for this. Perhaps the foremost reason pertains to the issue of whether queue transparency may cause participants to take on credit risk by crediting clients' accounts with expected incoming funds prior to these payments actually being received. This was a major concern of central banks at the time the LVTS was being developed. See RTGS (1997) and discussion in Section 4.6.2.

methodology as well as a description of the data. Section 4.6 presents results from the simulations and related discussion. Section 4.7 offers concluding remarks and some caveats to the analysis.

4.2 Settlement delay and intraday liquidity in an LVPS

Participants in an LVPS typically maintain a daily schedule of payments which they must send through the system on behalf of their own business and that of their customers. Included in this schedule is the time that each payment is due to be sent. For example, certain payments are considered 'time-sensitive' and thus have to be sent by a specific time during the day. The remaining majority of payments is considered 'non-time-sensitive' and simply must be sent by the end of the day. In practice, however, participants generally do not wait until the end of the day to submit all of their non-time-sensitive payments for reasons that will be outlined below.

In Real-Time Gross Settlement (RTGS) and RTGS-equivalent LVPS (such as Canada's LVTS), participants must maintain intraday funds in the system to send a payment to another bank. Hence, the concept of intraday liquidity in an LVPS specifically refers to a participant's ability to access sufficient intraday funds to meet its outgoing payment obligations in a timely manner. There are two main sources of intraday funds available to an LVPS participant: 1) funds acquired from other participants due to either regular transaction activity or through an interbank loan arrangement and 2) funds acquired through an intraday credit extension. Incoming funds from regular transaction activity are the cheapest source of liquidity for participants, and it is expected that participant banks will try to use these funds as much as possible to finance their own payment activity.[4] For various reasons (eg the differing nature of individual participants' business), however, it may not always be possible for participants to coordinate their daily payment activity so that incoming payments largely finance their outgoing payment needs.

The inability of participants to perfectly coordinate their incoming and outgoing payment activity creates a role for the provision of intraday credit. Martin (2005) emphasises the importance of intraday

---

[4] See McAndrews and Rajan (2000) and McAndrews and Potter (2002) for discussion and identification of this type of coordination behaviour among participants in the US Fedwire system.

credit as a source of intraday funding for participants. The author argues that the coordination of incoming payments to meet outgoing obligations is often difficult (especially for time-sensitive payments), and therefore a well-designed LVPS should allow participants to acquire funds when necessary through intraday credit. Where intraday credit is available to participants on a free and unlimited basis, participants can borrow funds any time that a payment is due, thus eliminating potential settlement delay in the LVPS. However, although settlement delay would cease to exist in this case, lenders of intraday credit (typically central banks) could face large risk exposures vis-à-vis borrowers, which is not desirable from a public policy perspective. Consequently, intraday credit in RTGS and equivalent systems is not free and unlimited, but rather is often subject to net debit caps, (eligible) collateral requirements which typically entail an opportunity cost, and in certain cases an explicit interest charge, eg the US Fedwire system. Maintaining intraday liquidity in the system can therefore be costly for participants.

Where a participant does not have sufficient funds available to meet a payment obligation upon intended submission, processing of the payment by the LVPS will be delayed. Settlement delay can be defined as a time lag occurring between a participant's intended submission of a payment to the LVPS, and when the payment is processed by the LVPS with finality, ie when intraday funds are exchanged between participants on an unconditional and irrevocable basis in order to discharge the payment obligation.[5] Payments that cannot be processed because of a participant's lack of intraday liquidity may be held in that participant's internal queue. Alternatively, these payments could be submitted to the LVPS and held in the system's central queue if one is available. Under standard queuing procedures, internally and centrally queued payments are released and processed by the LVPS on an *individual* basis when a sending participant's intraday liquidity improves to the extent that these payments can be passed.[6] The settlement delay associated with an individual payment essentially reflects the amount of time that the payment must wait in the queue before being processed by the LVPS.

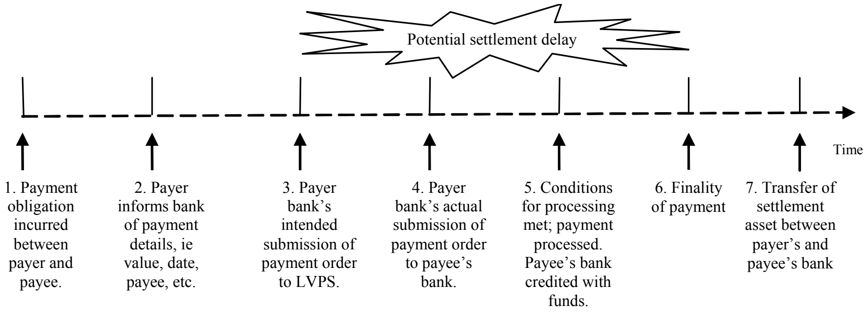
---

[5] A key feature of RTGS and equivalent LVPS is that these systems offer immediate intraday finality. Payments in these systems are considered final upon being processed.

[6] This liquidity improvement could occur as a result of the participant receiving a payment, or gaining access to more intraday credit.

Figure 4.1 provides a graphical characterisation of settlement delay within the context of the life-cycle of a large value payment.[7]

Figure 4.1 **The life-cycle of a large-value payment**



Just as there is a cost associated with maintaining intraday liquidity in the system, given the high speed and high value of daily payments processed by an LVPS, settlement delay may also entail a significant cost for participants. Further, the nature of this delay cost is likely to depend on whether a payment is time-sensitive or not. Time-sensitive payments may include those related to the final funds settlement of other important national and international clearing and settlement systems, large government receipts and disbursements, and also payments related to the daily implementation of monetary policy. A participant that is unable to meet a time-sensitive payment obligation when due may therefore face large internally borne costs because of the delay, such as reputation damage with its peers and, possibly, a loss of its clients' business. Explicit penalty charges may even be imposed by the system operator since the delay of these payments could cause a disruption elsewhere in the financial system.

For the remaining majority of (non-time-sensitive) payments, there is no formal intraday deadline to submit these payments. It is not expected that a participant will incur an (immediate) reputation loss or penalty charge, nor a loss of its clients' business, if processing of these

---

[7] The paper recognises that achieving payment finality need not encompass the transfer of the settlement asset. Therefore, the notion of settlement delay applies equally to RTGS and RTGS-equivalent LVPS, where this transfer occurs on a multilateral net basis at the end of the day in the latter.

payments is delayed until the end of the day.[8] However, there may be other external costs imposed on the system in this case. Despite being non-time-sensitive, intended receiving banks may be expecting these payments by a certain time of day, and such a delay will result in a shortfall in their intraday funds position. If these participants are planning on using these funds to send their own payments, then they may have to incur additional costs in order to replace these funds on short notice. Where they cannot find other funds in time to meet their obligations, additional settlement delay is created in the system. Settlement delay created by one participant in an LVPS could quickly spread to others in the system. Moreover, a comparable disruption to the liquidity position of a receiving bank's client may also occur (where a delayed payment is ultimately intended for this customer), resulting in potentially broader consequences for economic activity.

Prolonged delay of these payments may also intensify the potential losses associated with other risks in the system, such as operational risk. An operational event (such as a computer outage that prevents one or more participants from sending payments) will likely have a larger impact in a case where a number of payments remain unprocessed at the time that the incident occurs.[9] At the same time, a large backlog of payments being submitted all at once to the LVPS late in the day could increase the potential likelihood that an operational event occurs in the first place. Lastly, where the potential for settlement delay could discourage use of an LVPS in favor of systems that are not as well risk-proofed, the existence of settlement delay may translate to higher systemic risk in the broader financial system.

It follows that, to eliminate the potential costs associated with settlement delay, participants will likely have to borrow a large amount of intraday credit and thus incur high liquidity costs. Conversely, participants need not incur any intraday liquidity cost, but will then have to bear (possibly along with other participants in the

---

[8] Prolonged delay of non-time-sensitive payments is unlikely to cause reputation loss immediately, but such a loss could occur if repeated over time. In a relatively concentrated payment system like Canada's LVTS, participants maintain frequent communication with each other throughout the day and are able to develop fairly accurate forecasts of certain incoming payment flows based on historical payment patterns with other participants. Thus, a participant that often delays its non-time-sensitive payments in favor of lower liquidity costs is unlikely to go unnoticed among its peers in the system.

[9] Conversely, an operational disruption could also lead to settlement delay in an LVPS since it may result in a participant's inability to send payments through the system. For this reason, contingency measures are usually available in an LVPS for the release of time sensitive payments in the event of a disruption.

system) the costs of the accompanying settlement delay. It is unlikely that participants will not maintain sufficient liquidity to meet their time-sensitive payment obligations since the cost of delaying these payments is very high. Consequently, the discussion of a tradeoff between settlement delay and intraday liquidity may not apply to time-sensitive payments in practice. However, for non-time-sensitive payments, the tradeoff is likely to exist. Since settlement delay may entail costs and repercussions for the system as a whole, any innovation in LVPS design that can increase settlement speed for a given level of intraday liquidity is desirable.

4.3 A simple graphical framework

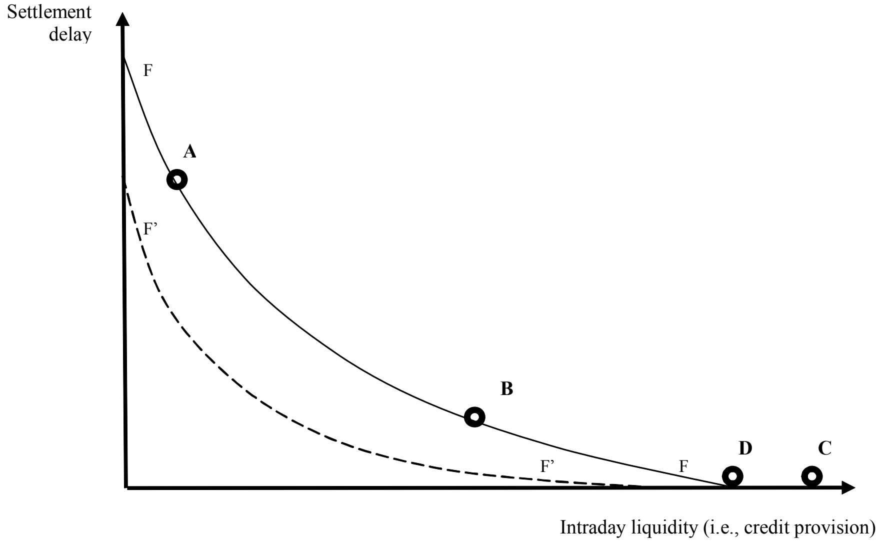
4.3.1 Description of the framework

The expected relationship between settlement delay and intraday liquidity in an LVPS is illustrated in Figure 4.2 below. Figure 4.2 is inspired by the concept of an 'efficient frontier' presented by Berger, Hancock and Marquardt (1996).[10] This framework will help in interpreting the empirical results later in the paper.

---

[10] In describing this framework, the terms 'intraday liquidity' and 'intraday credit' are used synonymously.

Figure 4.2 **The LVPS delay-liquidity efficient frontier**



The framework is presented in delay-liquidity space. All points in the space represent possible settlement delay-intraday liquidity combinations necessary to produce a given level of payment activity. The vertical axis measures the magnitude of overall settlement delay in the LVPS while the horizontal axis measures the provision of intraday credit. It is useful to think of the magnitude of settlement delay in an LVPS as reflecting both the number of payments entering the queue upon intended submission and also each payment's duration in the queue until being processed. The tradeoff is captured by the curve denoted FF, and this curve is generated based on the existing technology for processing payments (ie the existing LVPS design). Specifically, the curve shows how settlement delay and intraday credit provision can be traded off against each other for a given level of payment activity under current LVPS arrangements. The slope of FF captures the reduction in settlement delay that can be achieved by participants following a unit increase in the provision of intraday credit.

The decreasing convex shape of the tradeoff curve reflects the assumption of diminishing marginal returns to liquidity. An increase in intraday credit provision is anticipated to have a lesser impact in terms of reduced settlement delay when moving further along the frontier from left to right. This assumption is attributed to the positively skewed nature of the distribution of individual payment values in an LVPS.[11] At a very low level of liquidity (point A), a small increase in intraday credit provision will lead to a higher reduction in settlement delay since many smaller payments that would otherwise have been delayed can now be immediately processed upon intended submission. As intraday credit provision is continuously increased, it is expected that more payments will be processed upon intended submission and the delayed finality of these payments will be averted. However, even at higher levels of intraday credit provision (such as point B), it is expected that a few very large payments will still be delayed. Only a substantial injection of intraday credit would allow these payments to be processed immediately.

All combinations along the curve, and also above and to the right of the curve, represent feasible combinations of settlement delay and intraday liquidity for a given level of payment activity under the existing LVPS design. The tradeoff curve is the most technologically efficient of these feasible combinations and, therefore, an LVPS is considered to be technically efficient if it is processing payments anywhere along the curve. This notion of efficiency captures the idea that, when operating along the curve, reductions in settlement delay can only be achieved by an increase in intraday credit provision, and vice versa, for a given level of payment activity. Processing the same level of payment activity at a point above, or to the right, of the tradeoff curve represents inefficiency. For instance, producing at a point like C in Figure 4.2 means that intraday credit provision could be reduced and participants' liquidity costs lowered without causing any increase in settlement delay. In fact, intraday credit provision could be lowered from point C all the way to point D before any further reductions lead to increased settlement delay in the LVPS. Point D represents the familiar upper bound of liquidity as described in Leinonen and Soramäki (1999, 2003). Points below the efficient frontier are currently unattainable given the existing LVPS technology and can only be achieved through some form of innovation.

---

[11] For instance, in Canada's LVTS, the average payment value is around CAD 7.5 million while the median value is around CAD 50,000. Moreover, the value of some payments in the LVTS is well over CAD 100 million.

4.3.2 Innovation: a complex queue-release algorithm

As mentioned above, points below the tradeoff curve are not attainable given the existing LVPS technology. An improvement that allows lower settlement delay for any given level of intraday liquidity, or vice versa, is required to attain such an outcome. The impact of this improvement appears in Figure 4.2 as a shift of the tradeoff curve FF to its new position closer towards the origin at F'F'. Along the new curve, the same amount of payment activity can be produced with lower settlement delay for each level of intraday liquidity, and therefore at a lower overall cost to participants.

Such an improvement can be achieved through a technological innovation in LVPS design. Reductions in settlement delay can be achieved through either faster processing of queued payments or fewer payments entering the queue upon submission, where the latter may occur as a result of the former. Faster processing of queued payments means that intended receivers will obtain incoming funds more quickly, reducing the likelihood that their own subsequent outgoing payments will become queued upon submission. It is argued that the replacement of standard queuing arrangements with the introduction of central queuing with a complex queue-release algorithm represents such an innovation. The benefit of these types of algorithms, in terms of both reduced settlement delay and intraday liquidity needs in an LVPS, are frequently highlighted throughout the payments literature. For example, see McAndrews and Trundle (2001), BIS (2005), Leinonen and Soramäki (1999), Bech and Soramäki (2001), Güntzer, Jungnickel, and Leclerc (1998) and Koponen and Soramäki (1998).

These algorithms are designed to simultaneously search for and offset batches of queued payments, thus serving as an effective coordination device for participants' incoming and outgoing payments. Recall, under standard queuing procedures, payments are released from the queue *individually* when a participant's intraday liquidity is sufficient for them to be processed. In contrast, under central queuing with a complex queue-release algorithm, the simultaneous processing and release of a batch of queued payments is attempted at regular intraday intervals. In this latter case, LVPS participants no longer must wait to obtain sufficient intraday funds for their queued payments to be released individually, but rather they only need to hold the amount of intraday funds necessary to settle any net debit position resulting from the payment offset. The anticipated benefits to LVPS participants from this innovation include lower intraday liquidity needs and related costs for the release of queued payments, faster processing times for these queued payments, and a reduction in average intraday queue length, when compared to a standard queuing arrangement.

The addition of a complex queue-release algorithm will not necessarily represent a new development in all LVPS, since these algorithms have been used in some systems in the past as a gridlock resolution mechanism. However, over the last decade increases in computing power have led to the improved design and more frequent use of these algorithms within an LVPS central queue. The complexity of these algorithms has also risen considerably; the choice of full or partial optimisation is available and offsetting may take place on a bilateral and/or multilateral basis; BIS (2005).

To sum up, it is expected that the addition of a central queue with a complex queue-release algorithm will lead to an improvement in the tradeoff between settlement delay and intraday liquidity in an LVPS and will allow participants to complete the same level of payment activity at a lower overall cost, relative to a standard queuing arrangement.

4.4 Empirical study: estimating the tradeoff in Canada's LVTS

This empirical exercise considers the tradeoff between settlement delay and intraday liquidity in Canada's LVTS. Some questions that may arise are: What does the tradeoff curve look like for the LVTS? Does it have the same shape as outlined above? Are there possible LVTS design changes, relating to queuing arrangements or otherwise, that could potentially improve this tradeoff, where the same level of payment activity can be processed with either reduced settlement delay or lower intraday liquidity needs or both? The remainder of this paper is devoted to answering these questions using simulation analysis. Simulation analysis is a recent development in payment systems research. Simulation models are a valuable tool since they often can be calibrated to replicate a specific LVPS environment. These models can then be used to assess the impact of changes in the structural arrangements and decision parameters of an LVPS without causing any costly disruption to the operation of the actual system.

4.4.1 Background on the LVTS[12]

The LVTS is an RTGS-equivalent system, where individual payment messages are processed on a gross basis in real-time and settlement of the system occurs on a multilateral net basis at the end of the day. The LVTS's risk controls and collateral arrangements, coupled with a settlement guarantee provided by the Bank of Canada, provide certainty of settlement for the system.[13] Certainty of settlement facilitates intraday finality for all individual payments sent through the LVTS. Recipients of LVTS payments can make use of these funds immediately upon receipt without any possibility that a payment will become unwound. The LVTS consists of two payment streams – Tranche 1 (T1) and Tranche 2 (T2) – and participants may use either stream when sending payments through the system. Each stream has its own real-time risk controls and collateral arrangements. The focus of this analysis is on the T2 payment stream since, due to its more economical collateral requirements relative to T1, it is the dominant stream for LVTS activity.[14]

Intraday liquidity in T2 is facilitated by T2 payments previously received and also by drawing on a T2 intraday line of credit. This intraday line of credit is subject to both a (indirect) collateral requirement and a net debit cap. Specifically, LVTS participants grant bilateral credit limits (BCLs) to each other, where the value of a BCL represents the maximum bilateral T2 net debit position that a grantee (credit line recipient) may incur vis-à-vis the grantor (credit line provider) at any time during the payment cycle. A participant's T2 intraday credit limit, known as its T2 Net Debit Cap (T2NDC), is calculated as the sum of all BCLs granted to it by others in the system multiplied by a system-wide parameter (SWP), which is currently equal to 0.24.[15] The T2NDC represents the maximum multilateral T2

---

[12] Only LVTS background information relevant to the analysis is provided here. For more information on the LVTS, see Dingle (1998) and Arjani and McVanel (2006).

[13] In the extremely remote event of multiple participant defaults in the LVTS, and if collateral value pledged by participants to the Bank of Canada is not sufficient to cover the final net debit positions of all defaulters, the Bank stands ready to exercise its settlement guarantee by realising on available collateral and absorbing any residual loss.

[14] Approximately 87% of daily LVTS value and 98% of daily LVTS volume are sent through the T2 payment stream, on average. T1 consists of mostly time-sensitive payments between LVTS participants and the Bank of Canada.

[15] The SWP is an exogenous parameter established by the CPA. When the LVTS began operations in February 1999, the SWP was equal to 0.30. Since then, it has been gradually reduced and has been equal to 0.24 since March 2000. The choice of SWP value (SWP < 1) reflects the effect of multilateral netting; Engert (1993). See LVTS Rule No. 2 for information on the SWP.

net debit position that a participant can incur during the LVTS payment cycle. The T2NDC of hypothetical bank n (where n = 1,..,N) is calculated as follows

$$T2NDC^{n} = \sum_{j \neq n}^{N-1} BCL_{jn} \cdot SWP$$

It follows that two real-time risk controls are applied to payments submitted to the T2 payment stream. A payment will only be processed if it does not result in the sending participant exceeding either its BCL vis-à-vis the receiver or its T2NDC.

A survivors-pay collateral pool is used in T2 to facilitate LVTS settlement in the event of participant default. Eligible collateral consists mainly of government securities and also high-quality corporate debt. Participants are required to pledge T2 collateral equal to the value of the largest BCL that they grant to any other participant, multiplied by the SWP. The value of this T2 collateral obligation is referred to as a participant's Maximum Additional Settlement Obligation, or MaxASO. Essentially, a participant's MaxASO represents its maximum financial loss allocation as a result of another participant's default in the LVTS. Hypothetical bank n's MaxASO is calculated as follows

$$MaxASO^{n} = \max(BCL_{n, j \neq n}) \cdot SWP$$

The LVTS employs a central queue. Submitted payments to the LVTS failing the real-time risk controls are stored in this queue.[16] The queue is equipped with an offsetting algorithm that runs at frequent intervals (every 15 minutes) throughout the payment cycle. This complex queue-release algorithm, called the Jumbo algorithm, searches for and offsets full or partial batches of queued payments on a multilateral and/or bilateral basis.[17] Payments successfully released by this mechanism are processed by the LVTS as normal. However, current LVTS rules state that excessive use of the central queue is not

---

[16] Payments are stored on a First-In First-Out (FIFO) basis within each tranche type. Currently, only 'Jumbo' payments (> CAD 100 million) failing the real-time risk controls become centrally queued in the LVTS.

[17] For queued T2 payments, the Jumbo algorithm applies partial offsetting on both a bilateral and multilateral basis over two stages. See Arjani and McVanel (2006) for more information on this algorithm.

encouraged.[18] Instead, participants utilise internal queues to store payments that are unable to pass the real-time risk controls upon intended submission. Internally queued payments are typically re-submitted against the LVTS's risk controls (within a participant's internal LVTS workstation) *individually* on a by-pass FIFO basis each time that its intraday liquidity position is increased.[19] If this process reveals that an internally queued payment can pass the risk controls, it is automatically released to the LVTS for processing.

4.4.2 Settlement delay and intraday liquidity in T2: tradeoff and improvement

Deciding on how to hypothetically impose a reduction in participants' intraday liquidity represents a key aspect of the analysis. For the LVTS T2 payment stream, one way to accomplish this is to constrain the intraday credit available to participants by lowering the value of the SWP.[20] As in the earlier discussion, a reduction of the SWP will entail both a benefit and cost for LVTS participants, holding BCL values constant. The benefit is that a reduction in the value of the SWP will lower participants' T2 collateral requirement and related liquidity cost. However, assuming that no migration of payments from T2 to T1 occurs, reducing the SWP will likely also increase the level of settlement delay in the T2 payment stream. This is because participants' T2NDCs will decline, lowering T2 intraday liquidity in the system, and causing more payments to become queued upon their intended submission. Under current queuing arrangements, delayed payments will accumulate in participants' internal queues until the sending participants' T2 liquidity is sufficient for these payments to be processed by the LVTS.

The tradeoff curve between settlement delay and intraday liquidity in the LVTS is expected to have a decreasing convex shape as outlined in the earlier graphical framework. As the SWP is reduced further, overall settlement delay in the system is expected to rise at an

---

[18] LVTS Rule No. 7 states that participants are able to track their bilateral and multilateral positions in real-time through their internal LVTS workstations and are expected not to submit payments that will fail the risk controls.

[19] Under bypass-FIFO, a participant's first (earliest) queued payment will be re-tried against the risk-controls. If it does not pass, this payment will be by-passed and the participant's second queued payment will be re-tried, and so on.

[20] Alternatively, such reductions in intraday credit availability can also be achieved through reductions in the value of BCLs that participants grant to each other, while maintaining the current SWP value of 0.24.

increasing rate. Participants will become constrained by their T2NDC more quickly and frequently throughout the day when trying to send payments. In the extreme case, an SWP equal to zero will result in a state of payments deadlock where settlement delay reaches a maximum. No participant will have access to T2 intraday credit and therefore will not be able to incur a T2 net debit position. Consequently, no payments will be sent and all will remain unsettled in participants' internal queues until the end of the day.

It has been argued that an improvement in the tradeoff between settlement delay and intraday liquidity can be achieved with the introduction of a complex queue-release algorithm in the central queue. The LVTS already contains a central queue with a partial offsetting algorithm, but use of this queue is currently discouraged. It is anticipated that, by allowing increased use of the LVTS central queue (and this algorithm), overall settlement delay could be reduced for each hypothetical level of T2 intraday credit provision. Under this alternative scenario, participants would no longer need to manage an internal payments queue and instead would submit all payments to the LVTS at the time they are intended regardless of whether these payments could be immediately processed by the system. Release of these queued payments could then be attempted on a multilateral net basis rather than individually.[21] This proposed change in queuing regime is expected to increase the efficiency of the system since, even where the amount of T2 intraday credit available to participants (and related cost) is lowered, the processing time for queued payments can be faster, and average queue length could decrease, compared with current internal queuing arrangements.

In the next sections, a simulation approach will be utilised to shed light on the following questions:

- Under current internal queuing arrangements, what does the tradeoff between settlement delay and intraday liquidity in the LVTS look like? Is it consistent with the assumptions of the graphical framework presented above?
- Could increased use of the LVTS central queue improve this tradeoff? In other words, can the level of settlement delay associated with each amount of intraday credit be reduced for a given level of payment activity?

---

[21] The key benefit of central queuing compared to internal queuing is that multilateral offsetting of payments is only possible in the former case.

4.5 Data description and simulation methodology

4.5.1 Description of data

Three months of LVTS T2 transaction and credit limit data have been extracted over the period July-September 2004. Transaction data include the date and time that each transaction was submitted to the LVTS as well as the value of each payment and the counterparties involved in the transaction. It is assumed that the time stamp attached to each payment represents the intended submission time of the payment. Transactions data include only those payments processed by the LVTS and do not include rejected or unsettled payments. Data on credit limits include the value of the T2NDC available to each participant as well as the date and time that the value of the T2NDC is effective. These data represent 64 business days and approximately 1.05 million transactions and are believed to be representative of normal LVTS activity. Table 4.1 provides a summary of the transaction data.[22]

Table 4.1 **Summary of LVTS T2 transaction data**

	Jul 2004	Aug 2004	Sep 2004
Total value of T2 payments (CAD billion)	2,283.0	2,203.5	2,446.5
(% of LVTS total)	(87.8)	(87.9)	(86.3)
Total volume of T2 payments	349,948	344,357	356,676
(% of LVTS total)	(98.0)	(98.0)	(98.1)
Daily average value (CAD billion)	108.7	100.2[23]	116.5
Daily average volume	16,664	15,653	16,985
Average payment value (CAD million)	6.52	6.40	6.86
Median payment value (CAD)	42,436	40,377	45,719

---

[22] In addition, the Hirschman-Herfindahl Index (HHI) suggests that payment activity over the sample period is somewhat concentrated. The HHI will vary between 0.50 (concentration among only two banks) and 1/N (equal distribution of payment activity among all participants), where N represents the number of banks in the sample. In this case, 1/N = 0.08. The average HHI value for the sample is 0.1944 and 0.1813 for T2 payments value and volume, respectively. A value in this range is consistent with payment activity being distributed evenly across approximately 5–6 banks. Indeed, the largest five Canadian banks account for between 85–90% of daily LVTS value and volume.

[23] A lower average daily T2 payments value in August is expected given that the Canadian civic holiday occurs during this month. Total value reached only CAD 6.9 billion on this holiday in 2004.

4.5.2 Simulation description and methodology

The simulation analysis is conducted using a payment and settlement simulator developed by the Bank of Finland (the BoF-PSS2). This software application is currently being used by over thirty central banks. It should be noted that the version of the BoF-PSS2 used for this analysis does not contain BCL functionality, which is an important component of the LVTS.[24] As a result, the methodology in this paper includes the assumption that BCL values remain constant in light of proposed changes to LVTS rules on queue usage. Further, participants' payment-sending behaviour is also treated as exogenous and therefore the same transactions data are used throughout the analysis. Potential implications associated with these assumptions are addressed later in the paper.

Two batches of simulations will be run where each batch is intended to replicate a different LVPS design. In particular, batch one replicates the current internal queuing arrangement in the LVTS, while batch two replicates the alternative central queuing arrangement. Each batch consists of eight individual simulations (s = 1,2,..,8), where each simulation is distinguished by tighter constraints on participants' intraday liquidity. Changes in intraday liquidity are introduced by altering the value of each participant's T2NDC. Since it is assumed that BCLs remain constant, a reduction in each participant's T2NDC is achieved by hypothetically lowering the value of the SWP. Specifically, each individual participant n's T2NDC in simulation s is calculated as follows

$$\mathrm{T2NDC}_s^n = \mathrm{SWP}_s \cdot \sum_{j \neq n}^{N-1} \mathrm{BCL}_{jn}$$

where $\mathrm{SWP}_{1,\ldots,8}$ = 0.24, 0.21, 0.18, 0.15, 0.12, 0.09, 0.06, 0.03.[25]

In specifying the first batch of simulations, the objective is to mimic participants' decision to either submit a payment to the LVTS for processing or hold the payment internally when sufficient intraday funds are unavailable. Settlement delay occurring in this batch

---

[24] A version of the BoF-PSS2 was released in 2006 that includes both multilateral and bilateral credit limits functionality. Bank of Canada staff were involved in the development and testing of this new version.

[25] Transactions data include only processed payments under the current SWP value of 0.24. Thus, it is not possible to observe potential reductions in settlement delay from an SWP value greater than 0.24, due to a lack of readily available data on delayed or unsettled transactions for this SWP value.

represents payments being held internally by participants, ie the simulator's queue is replicating participants' internal queues. A bypass-FIFO queue-release algorithm is specified to imitate current internal queuing practices of LVTS participants. When this algorithm is applied, a participant's queued payments are re-submitted from the queue and re-tried against the risk controls on an individual bypass-FIFO basis whenever its intraday liquidity position improves. In the real LVTS, this occurs within the participant's internal workstation. Internally queued payments that can successfully pass the risk controls are assumed to be released from the participant's queue and submitted to the LVTS for processing. In interpreting the simulation results for this first batch, settled transactions are assumed to be those that participants were able to submit to the LVTS for processing, while unsettled transactions represent those remaining in participants' internal queues due to lack of intraday liquidity.

Specification of the second batch is intended to replicate a central queuing regime similar to that available in the LVTS. In these simulations, two queue-release algorithms are specified that closely match the LVTS's actual release mechanisms. The first of these algorithms is a FIFO (no by-pass) queue-release algorithm which re-submits a participant's centrally queued payments against the risk controls on an individual FIFO basis each time its intraday liquidity position improves. The second is a complex queue-release algorithm which employs partial offsetting on a multilateral basis and is scheduled to run every twenty minutes, similar to the LVTS's Jumbo algorithm.[26] Settlement delay captured in this second batch of simulations is meant to represent payments being held in the system's central queue, ie the simulator's queue is replicating the LVTS central queue. In the simulation results for this batch, all payments in the sample are assumed to have been submitted to the LVTS at their intended time of submission, and unsettled transactions are those remaining in the central queue which cannot be processed due to a sender's lack of intraday liquidity.

---

[26] At the time that the analysis was conducted, the frequency of the Jumbo algorithm was every 20 minutes. The frequency of this algorithm increased to every 15 minutes in December 2005. Since bilateral credit limit functionality is currently not incorporated in the simulation application, the partial offsetting algorithm used in the simulations does not exactly replicate the LVTS Jumbo algorithm for T2 payments. Despite this limitation, the results generated by the simulations are still expected to be useful and relevant. Further, in specifying this second batch of simulations, it is also assumed that the LVTS's queue expiry algorithm is no longer utilised and *all* payments failing the risk control check become centrally queued (not just 'Jumbo' payments).

Three alternative measures of settlement delay are calculated for each simulation within each batch. These measures are intended to capture the daily level of settlement delay associated with each amount of intraday credit provision under both the current and alternative queuing environments described above for the same level of payment activity. They are described as follows

1. Daily proportion of unsettled transaction value (PU):

$$PU_t^N = \left(\frac{\text{Value of unsettled transactions}_t^N}{\text{Value of submitted transactions}_t^N} \right)$$

   This indicator is calculated on an aggregate level (ie across all participants) for each day t in the sample, where t = (1,…,64). This measure represents the occurrence of the maximum settlement delay possible for a payment in this analysis. Unsettled transactions represent those that enter the queue upon intended submission and remain there until the end of the day.

2. Daily system-wide delay indicator (DI):

$$DI_t^N = \left(\sum_{n=1}^{N} \omega^n \rho^n \right)$$

$$\text{where } \rho^n = \left(\frac{\sum_{i=1}^{T} Q_i^n}{\sum_{i=1}^{T} V_i^n} \right) \text{ and } 0 \le \omega^n, \rho^n, DI^N \le 1$$

   Adapted from Leinonen and Soramäki (1999) and commonly used in payment simulation analyses, this indicator is calculated on an aggregate level and is based on a weighted average of each individual (n) participant's daily delay indicator (ρ). This indicator (and the ratio ρ) can take on any value between 0 and 1, where a value of 0 is achieved when all payments are successfully processed by the LVPS upon intended submission and no settlement delay occurs. A value of 1 is calculated where all payments become queued upon intended submission and remain unsettled at the end of the day. Weights (ω) are based on participants' average share of total transaction value over the 64-day sample period. Calculation of this measure requires dividing each LVTS business day into T=108 ten-minute intervals (i = 1,..,T). The numerator of ρ represents the sum of a participant's queued payment value (Q) over all T ten-minute intervals throughout the day. The denominator represents the sum of the cumulative value of a participant's submitted payments (V) over all T ten-minute intervals throughout the day. It follows that this indicator is influenced by both the value and delay duration of each payment in the queue calculated for each intraday interval.

3. Average intraday (interval) queue value (AQV):

$$AQV_t^N = \left(\frac{\sum_{i=1}^{T} Q_i^N}{T} \right)$$

This is an aggregate measure which calculates the average value of queued payments in an interval over day t. It is found by dividing the sum of total queued payment value (Q) over all T ten-minute intervals on each day by the number of intervals per day (T=108).

4.6 Simulation results and discussion

4.6.1 The delay-liquidity tradeoff in the T2 payment stream

Simulation results for each of the three delay measures are presented in Figures 4.3 through 4.5. Two curves are presented in each graph corresponding to each batch of simulations. The curve denoted 'internal queuing' portrays the simulation results estimated under current LVTS (internal) queuing arrangements. The curve denoted 'central queuing' depicts results estimated under the alternative LVTS (central) queuing environment.

Figure 4.3 **Average daily proportion of unsettled transaction value**

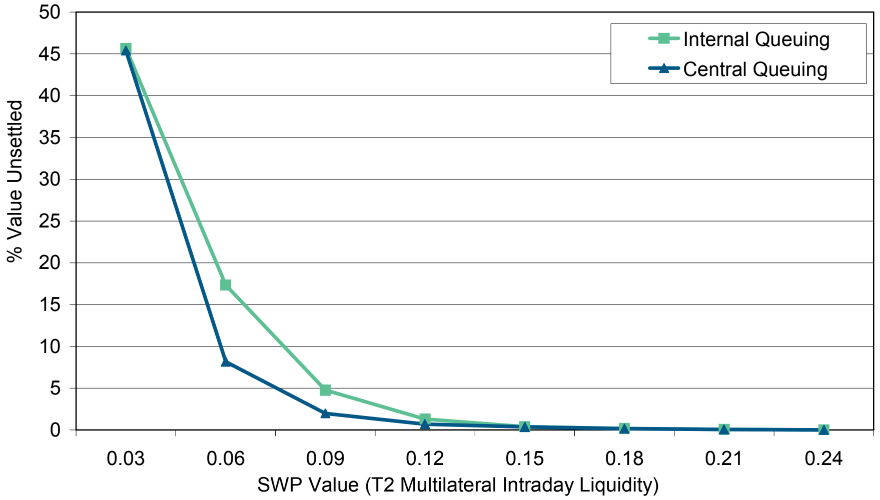


Figure 4.4 **Average daily system-wide payments delay**

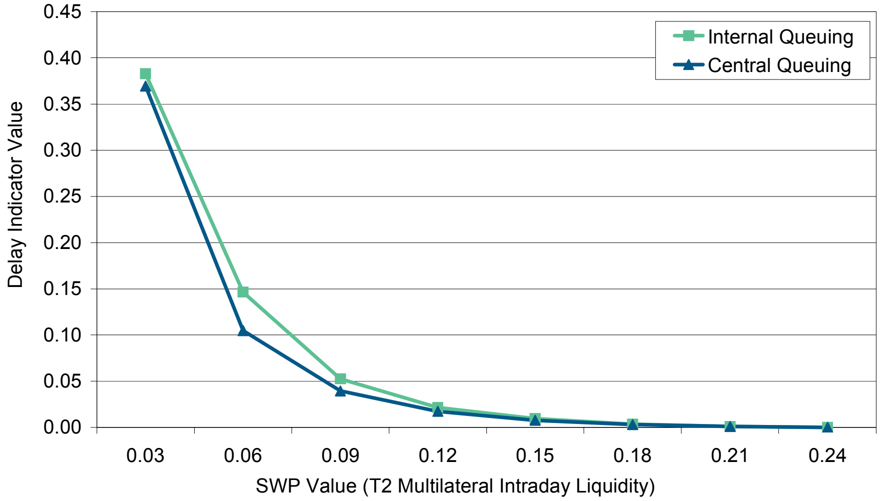
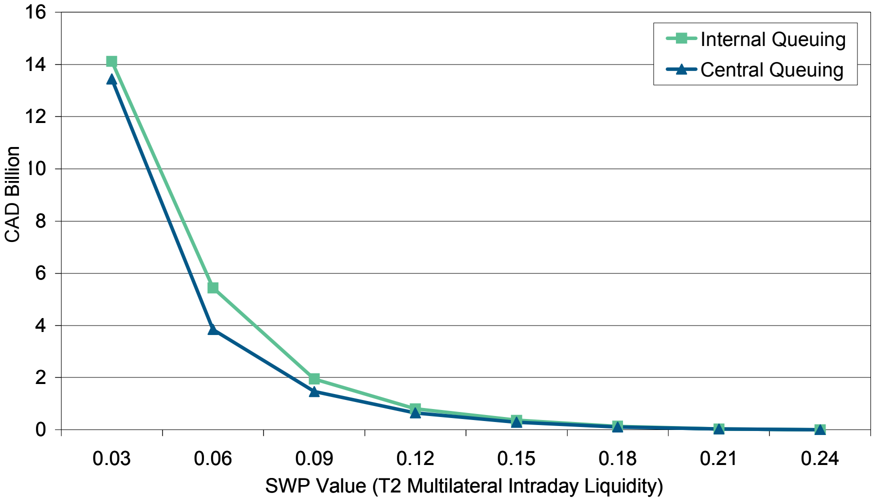


Figure 4.5 **Average intraday (interval) queue value**



Earlier hypotheses regarding the tradeoff between settlement delay and intraday liquidity are confirmed by the simulation results. Under current LVTS queuing arrangements, a tradeoff exists in the LVTS's T2 payment stream according to all three delay measures. Like the earlier graphical framework, the curve is convex; as intraday credit constraints are further tightened (by lowering the value of the SWP), participants' intraday liquidity becomes more scarce and settlement delay in the system rises at an increasing rate. The slope of this curve increases substantially at low amounts of intraday credit provision.

The introduction of a design innovation – allowing increased use of the LVTS central queue – results in an improvement to this tradeoff and the curve shifts closer towards the origin according to all three measures. Settlement delay associated with each level of intraday credit provision is reduced following the introduction of the partial offsetting algorithm. The relative benefit of partial offsetting (in terms of reduced delay) increases gradually as intraday liquidity is further constrained. At the SWP value of 0.06, the difference in settlement delay between the two queuing regimes is greatest. In this case, the average proportion of unsettled transactions value is reduced by 9 percentage points or about CAD 10 billion (Figure 4.3), the system-wide delay indicator is reduced by 28% (Figure 4.4) and average intraday queue value is reduced by 29% or about CAD 1.6 billion (Figure 4.5), relative to the first batch of simulations.

Gains from the alternative central queuing design begin to decline when the SWP is reduced beyond 0.06, as the system begins to approach a state of deadlock. When the SWP value is 0.03, settlement delay is only slightly reduced following the introduction of a partial offsetting algorithm, which could mean that participants' intraday liquidity levels are so low that only very small batches of queued payments can be processed each time this algorithm runs. At this level of SWP, close to half of all daily payment value remains unsettled on average under both queuing regimes (Figure 4.3).

The simulation results also reveal another finding that is closely related to the notion of technical efficiency described earlier. The above results suggest that, under current queuing arrangements, settlement delay in T2 increases when the SWP value is lowered from 0.24 to 0.21. However, it remains to be seen whether reductions in the SWP below 0.24 but still greater than 0.21 can be achieved without inducing any further settlement delay in the LVTS. In other words, can a lower amount of T2 intraday credit (and an associated reduction in T2 collateral requirements) be accommodated without increasing the level of settlement delay for payment activity during the three-month sample period, holding all other factors constant? If this were the case, it would be similar to operating at point C in the graphical framework. Indeed, the simulation results suggest that the current value of SWP (= 0.24) is needed to process payments in this sample and cannot be reduced further without increasing the level of settlement delay. This is not necessarily a surprising result since one might expect participants to conform to this value of SWP when sending payments through the system. A complete discussion of this analysis, including full details of the simulation methodology used, is provided in Appendix 1.

4.6.2 Discussion

Some other interesting discussion points emerge from these results, offering areas for future research. First, the simulation results suggest that, under both existing LVTS queuing arrangements and also under the alternative central queuing arrangement, settlement delay in T2 will increase only marginally as the SWP is initially reduced from its current value of 0.24, holding all other factors constant. For example, a reduction in the SWP from 0.24 to 0.18 is estimated to increase the average proportion of unsettled daily transaction value by only 0.15 per cent under the current queuing regime and 0.14 per cent under a central queuing arrangement (Figure 4.3). Similar results are also observed according to the other two delay measures. Reducing the SWP entails a benefit for LVTS participants in the form of lower T2 collateral requirements and related liquidity cost, as has already been mentioned. Specifically, a reduction in the SWP to 0.18 reduces the aggregate value of T2 collateral required by about CAD 750 million per day on average over the sample period, holding BCL values constant. On one particular day in the sample, the value of T2 collateral required is about CAD 1 billion less when the SWP is equal to 0.18.

This raises the question as to whether or not a lower-cost combination of intraday credit provision and settlement delay currently exists for LVTS participants in the T2 payment stream.[27] Put differently, is it the case that the marginal settlement delay cost incurred by moving to an SWP value of 0.18 equals the marginal cost of additional intraday credit provision (and collateral) associated with the current value of 0.24? If the former cost is less than the latter, then lowering the SWP to 0.18 could lead to overall cost-savings for participants. Of course, answering this question entails, among other things, the difficult task of quantifying the cost of the additional settlement delay associated with moving to a SWP value of 0.18.

Secondly, the analysis highlights the possible benefit of central queuing with a complex queue-release algorithm with respect to settlement delay and intraday credit provision. Nonetheless, participants face other types of risk and cost in the LVPS environment, and such a change in LVTS queuing arrangements could increase participants' other costs. For example, as outlined in BIS (1997), a possible implication of permitting unrestricted use of the central queue pertains to the issue of queue transparency and specifically whether the reduction in settlement delay could be replaced by an increase in credit risk taken on by participants. A participant, upon observing an incoming payment in the central queue, may choose to provisionally credit its client's account with these expected funds before the payment actually arrives, thus exposing itself to credit risk until the payment is successfully received. If these funds do not eventually arrive for some reason, the participant would seek to unwind this payment, which would be costly for both the participant and its client. This issue is pertinent to the LVTS because participants have the ability to track expected incoming and outgoing payments in the queue in real-time through their internal participant workstations. Although details regarding client recipients of incoming queued payments are not included in these workstation reports,

---

[27] Alternatively, the question could instead be posed as whether current values of BCLs granted by participants to each other are cost-minimising holding the current SWP value constant.

participants could informally access this information. However, it is not clear that LVTS participants would be willing to incur this credit risk in any case.[28]

4.7 Conclusions and caveats

The objective of this paper has been to gain a better understanding of the tradeoff between settlement delay and intraday liquidity in an LVPS, with a specific focus on the Canadian LVTS. Simulation analysis shows that a tradeoff exists in the LVTS between settlement delay and intraday liquidity, and that this tradeoff exhibits a decreasing convex shape. Further, allowing increased use of the LVTS central queue (and the Jumbo algorithm) is expected to improve this tradeoff, ie, reduce settlement delay in the system for all levels of intraday liquidity considered. Such an innovation improves the efficiency of the system, leading to overall cost-savings for participants.

At the same time, it was found that under both the current and proposed queuing regimes, a modest reduction in the SWP below its current value results in only a marginal increase in the level of settlement delay in the LVTS, while providing substantial T2 collateral cost-savings for system participants. Further research is necessary to quantify whether this collateral cost-saving benefit is worth the associated increase in settlement delay cost. It was also argued that, although increased use of the central queue is expected to reduce total settlement delay and liquidity costs for participants, this may result in a potential increase in credit risk taken on by participants. However, LVTS participants may not necessarily react to a change in LVTS queuing arrangements in this manner.

These results are preliminary, and certain caveats exist. These caveats are raised here with the intention of motivating further research. The first caveat relates to behavioural assumptions made throughout the analysis. Significant changes to LVTS queuing arrangements were proposed in the analysis. However, despite these changes, the current simulation methodology assumes that LVTS

---

[28] This credit risk issue may also be avoided in the LVTS since a client beneficiary of funds can always request a Payment Confirmation Reference Number (PCRN) from its participant bank. All payments processed by the LVTS are assigned a PCRN indicating that the payment has successfully passed all LVTS risk control tests and is thus considered final and irrevocable. Upon obtaining the PCRN, the beneficiary does not have to worry about the funds being revoked at a later time.

participants' payment sending and bilateral credit granting behaviour remains unchanged. One must question whether this is a realistic assumption. For example, following discussion in McAndrews and Trundle (2001), the availability of netting is likely to increase the incentive for participants to submit payments to the system earlier in the day, relative to these payments' current intended submission times, essentially increasing the scope for multilateral netting of payment messages. The benefit of netting is expected to increase with the number and value of payments in the queue at the time that it occurs. Anecdotal evidence suggests that LVTS participants typically receive information regarding outgoing payment requests well in advance of their intended submission time. Participants' collective submission of as many payments as early as possible to the system under a central queuing regime is anticipated to result in a greater turnover of intraday funds, a lesser need for costly intraday credit, and faster processing of these payments. This may result in a further downward shift of the tradeoff curve closer to the origin thus leading to further cost-savings for participants.

At the same time, it is argued that participants, in granting BCLs to each other, strive to minimise the value of their T2 collateral requirement subject to achieving an established level of throughput efficiency, ie an acceptable level of settlement delay. It is likely that payment activity under current internal queuing arrangements may already reflect participants' acceptable levels of settlement delay. Thus, participants may not perceive the benefit of central queuing to be a further reduction in settlement delay, but instead may treat this as an opportunity to realise lower T2 collateral requirements (and costs) while maintaining the same level of settlement delay in the system. This suggests that, under the central queuing arrangement, participants may collectively choose to reduce the BCLs they grant to each other in order to achieve these cost-savings. This reduction in BCLs is expected to continue to the extent that any decline in settlement delay resulting from increased use of the central queue is fully offset.[29]

A second caveat follows closely with a discussion found in Bedford, Millard and Yang (2005) and relates to the statistical robustness of the simulation findings. The simulation analysis is

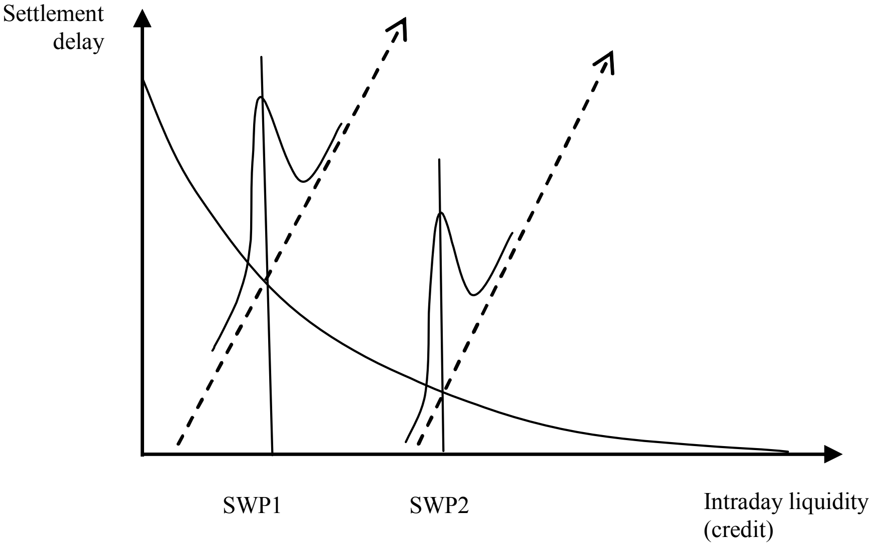
---

[29] Initially, participants are not likely to know exactly how much BCLs must be reduced to achieve the same level of settlement delay under the alternative central queuing regime. Instead, this will be an iterative process that eventually converges to the equilibrium of a perfect offset. In the interim, it may be the case that participants 'overshoot' this target level of BCL reduction, temporarily resulting in a higher level of settlement delay in the system relative to the existing level.

intended to estimate the increase in settlement delay brought on by a reduction in LVTS participants' intraday liquidity over a three-month sample period. Point-estimates of this impact for each amount of intraday liquidity are used to generate the tradeoff curves presented in Figures 4.3 through 4.5. Previous internal research conducted by the Bank of Canada shows that annual LVTS payment activity is affected by specific calendar events and also monthly trends. Consequently, the estimated impact on settlement delay following reductions in intraday liquidity is expected to take on different values based on the specific dataset used in the analysis. Although using a three-month sample helps to capture the effect of certain monthly and quarterly calendar effects occurring during this period, there is a desire to reduce the risk of small-sample bias and to obtain more statistically robust results. For example, it has been observed that the same calendar event may yield a different effect on LVTS payment activity depending on when it occurs throughout the year. Similarly, use of a single three-month sample may not capture the effect that semi-annual and/or annual calendar events may have on the simulation results. Nor will it capture the potential impact of monthly trends in LVTS T2 payment activity.

In order to achieve more statistically robust results, it is suggested that the same simulation methodology be repeated as many times as is feasible using real and/or artificially generated LVTS payment flow data over some fixed sample duration. Grouping the point-estimates of the impact on settlement delay for each amount of intraday liquidity from all of the samples will facilitate generation of an empirical distribution of this potential impact (Figure 4.6). It follows that the shape of the empirical distribution may be different for each amount of intraday liquidity. For example, the impact on settlement delay may be more volatile and will thus deviate from its mean value more often at lower amounts of intraday credit provision. The shape of the empirical distribution may also change over time.

Figure 4.6 **Plotting distribution of settlement delay outcomes**



A third and final caveat pertains to the absence of BCL functionality in the version of the BoF-PSS2 used in this analysis. This absence creates the possibility that the estimated tradeoff curves provided in Figures 4.3 through 4.5 represent a 'lower bound' of the impact on settlement delay resulting from reduced intraday liquidity. As the value of the SWP is reduced and payments become delayed upon failing the T2 multilateral risk-control test, intended receivers of these payments may consequently be prohibited from sending their own payments when due. All of this will result in added volatility in bilateral net positions, possibly to a point where some participants' bilateral net debit positions are greater than the BCLs granted to them. In the LVTS, this cannot occur due to a bilateral risk control test being applied to every payment which guarantees that participants do not exceed their BCL vis-à-vis a receiving participant. Payments failing the bilateral risk control test become queued until the sending participants' bilateral liquidity position improves. This added delay is not captured in the results generated by the current version of the simulator. This forces the assumption that all LVTS payments, when processed by the simulator, have passed not only the multilateral risk control test, but also the bilateral risk control test. Thus, it would be useful to repeat the analysis again with Version 2.0 of BoF-PSS2 to compare how much greater is potential settlement delay in the system when bilateral risk controls are also taken into account.

References

Arjani, N – McVanel, D (2006) **A Primer on Canada's Large-Value Transfer System.** Bank of Canada Web Document, www.bankofcanada.ca/en/financial/lvts_neville.pdf.

Bank for International Settlements (2005) **New developments in large-value payment systems.** May.

Bank of International Settlements (1997) **Real-Time Gross Settlement Systems.** March.

Bech, M – Soramäki, K (2001) **Gridlock Resolution in Interbank Payment Systems.** Bank of Finland Discussion Papers 9/2001.

Bedford, P – Millard, S – Yang, J (2005) **Analysing the impact of operational incidents in large-value payment systems: A simulation approach.** In Leinonen, H (ed.) Liquidity, risks and speed in payment and settlement systems – a simulation approach. Bank of Finland Studies, 247–274.

Berger, A N – Hancock, D – Marquardt, J (1996) **A Framework for Analyzing Efficiency, Risks, Costs, and Innovations in the Payment System.** Journal of Money, Credit and Banking 28(4), 696–732.

Dingle, J (1998) **The LVTS – Canada's large-value transfer system.** Bank of Canada Review (Autumn), 39–55.

Engert, W (1993) **Certainty of Settlement and Loss Allocation with a Minimum of Collateral.** Bank of Canada Working Papers No. 93–14.

Güntzer, M M – Jungnickel, D – Leclerc, M (1998) **Efficient algorithms for the clearing of interbank payments.** European Journal of Operational Research 106, 212–219.

Koponen, R – Soramäki, K (1998) **Intraday Liquidity Needs in a Modern Interbank Payment System: A Simulation Approach.** Bank of Finland Studies E:14.

Leinonen, H – Soramäki, K (1999) **Optimising Liquidity Usage and Settlement Speed in Payment Systems.** Bank of Finland Discussion Papers 16/99.

Martin, A (2005) **Recent Evolution of Large-Value Payment Systems: Balancing Liquidity and Risk.** Economic Review (First Quarter). Federal Reserve Bank of Kansas City, 33–57.

McAndrews, J J – Trundle, J (2001) **New payment system designs: causes and consequences.** Financial Stability Review (December). Bank of England, 127–136.

Appendix 1

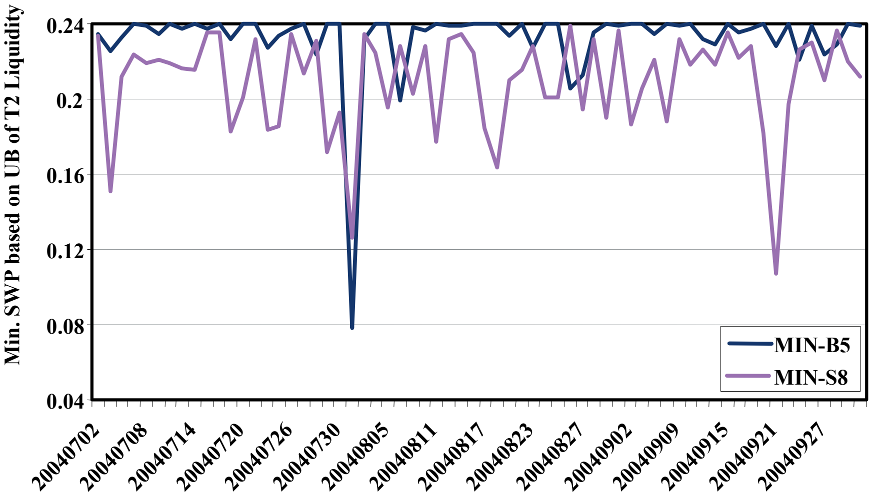
Is the T2 payment stream technically efficient?

The objective of this supplemental analysis is to find the minimum SWP (call this SWP*) necessary to process all payments in the sample without delay, holding all other factors constant. It may be the case that SWP* < 0.24, which means that existing levels of T2 intraday credit, and perhaps more importantly for participants, T2 collateral requirements could be lowered without inducing additional settlement delay during the three-month sample period (recall point C in Figure 4.2).

Simulation results produced by the BoF-PSS2 can provide insight into this issue. Treating participants' payment-sending behaviour as exogenous, a simulation is run using the same sample data but this time specifying unlimited intraday credit. Under this simulation scenario, all payments will pass the risk controls immediately upon submission and therefore no queuing algorithms need to be specified. The daily T2NDC each participant actually needs in order for its payments to be passed without delay can be derived from these simulation results, and is equal to the largest multilateral net debit (negative) position incurred by each participant during the day. This value is defined as a participant's upper bound (UB) of T2 liquidity. The daily UB of T2 liquidity for each participant can then be used to calculate a value of SWP* that, when multiplied by the sum of the actual BCLs granted to each participant, will produce this UB value. It follows that the highest value of SWP* calculated for any participant on any day is considered the minimum SWP* value necessary to send all payments in the sample through the system without delay. This SWP* can then be compared with the current value of 0.24.

The results from this simulation analysis reveal that on 45 of the 64 days, SWP* reached 0.24 for at least one LVTS participant. This means that the current value of SWP was necessary for the immediate processing of T2 payment activity during this three-month sample period. Hence, further T2 collateral cost-savings could not be realised without an increase in the level of settlement delay, holding payment activity constant. The results also indicate that the T2NDC constraint (when SWP=0.24) is binding more often for large LVTS participants (denoted 'B5' in Figure 4.7). Figure 4.7 below shows that on 42 days in the sample at least one of the major Canadian banks reached their T2NDC at some point in the day.

Figure 4.7 **Minimum SWP required – B5 vs. S8**



Focusing on the large LVTS participants, the simulation results show that, on these 42 days, four different institutions bumped up against their T2NDC at least once intraday. One of these participants reached its T2NDC at least once on 37 different days, while the three others reached this limit on 10, 2 and 1 day(s), respectively. The results also indicate that participants did not reach their T2NDC constraint at the same time each day. For example, regarding the first two large participants mentioned above, the LVTS day has been divided into four periods and the time that each of these participants reached its T2NDC has been located in the simulation results and tabulated. A summary of these findings is provided in Table 4.2.

Table 4.2 **Percentage of instances where T2NDC is binding by time of day**

Time of day	Bank 1 (37 instances)	Bank 2 (10 instances)
00:30–06:00	0	0
06:00–12:00	19	0
12:00–17:00	73	40
17:00–18:30	8	60

It also deserves mention that, where a high number of instances occur within a certain period (eg 27 instances for Bank 1 during the interval between 12:00 and 17:00 hours), these occurrences typically do not take place at the same time within the interval, but rather were scattered throughout the period.

It is not necessarily surprising that SWP* reaches 0.24 on most days in the sample period. The gradual reduction of the SWP from 0.30 to 0.24 between February 1999 and March 2000 was influenced by participants' preferences, and this value has held steady at 0.24 since that time. Given participants' perceived contentment with this SWP value, one might expect participants' to conform to it, meaning that they choose to structure their payment submission behaviour in a certain way so as to make full use of their available T2 intraday credit when sending payments through the system.

Some discussion is also warranted regarding results for the eight smaller LVTS participants (denoted 'S8' in Figure 4.7). On only 4 of the 45 days, SWP* reached 0.24 for one of these participants. Further, this occurred for a different participant in each of these four instances. There exist a variety of possible explanations for these results. It may be the case that larger LVTS participants, in sending a higher volume of payments earlier in the day, are 'subsidising' smaller participants' intraday liquidity in the system, to the extent that smaller participants need to rely less on intraday credit as a source of funding for their outgoing payments. Indeed, SWP* was equal to zero (ie no T2 intraday credit was drawn upon) for at least one small participant on 18 of 45 days in the sample. In contrast, this did not occur on any day for large LVTS participants. A second possible explanation could be that, for various reasons, small LVTS participants may tend to bump up against their BCLs far more frequently relative to their T2NDC. Of course, further research is necessary before either of these explanations can be confirmed.

Chapter 5

Funding levels for the new accounts in the BOJ-NET

Kei Imakubo – James J McAndrews

5 Funding levels for the new accounts in the BOJ-NET

Abstract

The Bank of Japan decided to implement the next-generation RTGS project of the BOJ-NET Funds Transfer System. Under the project, the new system will have liquidity-saving features and will incorporate large-value payments that are currently handled by two private-sector designated-time net settlement systems, the Foreign Exchange Yen Clearing System and the Zengin System. We analyse characteristics of the optimal funding levels under the new features using simulation analysis and find that the optimal funding levels can be described with the total balances in the system, the distribution of the total balances across participants, and the timing of funding.

5.1 Introduction

In February 2006, the Bank of Japan decided to implement the next-generation RTGS (RTGS-XG) project of the BOJ-NET Funds Transfer System (BOJ-NET), its primary large-value payment system.[1] Under the RTGS-XG project, BOJ-NET will introduce liquidity-saving features in a current real-time gross settlement (RTGS) mode. The new system will also incorporate payments from three different streams of the current payment activities, two of which now settle toward the end of the processing day in private-sector designated-time net settlement (DNS) systems. The project will be implemented in two phases, with the first phase scheduled for fiscal 2008 (April 2008 to March 2009) and the second for 2011. One of the primary motivations for the development of the new system is to quicken settlement of large-value payments relative to the current pattern and to reduce intraday settlement exposure of those payments by allowing for intraday settlement finality and liquidity-saving at the same time.

Much of the design work for the new system is already completed, while some decisions related to the implementation still remain. In the

---

[1] See Bank of Japan (2006b) for an overview of the RTGS-XG project.

paper, we focus on one aspect of the new system – the levels of funding for newly-developed accounts that will be drawn on to effect settlement throughout the day in a liquidity-saving mode.

The first issue that we explore is whether the plan to incorporate the payments that are currently settled on the two private-sector DNS systems and most payments on the current BOJ-NET into the new system will yield liquidity-saving under a certain level of funding. It is plausible to think that maintaining separate systems might require less liquidity or might result in speedier settlement for a given level of liquidity. If incorporating the payments in the three systems turns to be liquidity-saving, then it can be said that there are liquidity complementarities among the three systems to be combined. As demonstrated in the paper, strong complementarities do exist among the three systems.

Second, we simulate the performance of the new system using several levels of initial balances for the new accounts. In general, there is a clear trade-off between the rate of settlement of a group of payments and the level of funding devoted to those settlements. With a large level of funding, settlement can be made more quickly. Firstly, the total level of funding of initial balances is important in establishing how much value is settled prior to the end of the settlement period. Once the total level of funding is determined, participants can seek to optimise the distribution of initial balances across participants. The optimum distribution of balances across participants leads to the greatest value of settlement within the settlement period for that total level of funding used. A characteristic of the optimum distribution of balances across participants is that additional balances placed in any participant's account yield equal increases in amounts settled. This 'equalisation of marginal benefits' is a characteristic common to many allocation problems in economics.

We examine how changes in a level of initial balances affect the value of payments settled, the amounts left unsettled after a particular time, and the average time of settlement. This information can be useful to participants and planners in seeking the right balance between the value settled during the day and the liquidity-saving potential of the new system. In the context of Japan's payment activities, this is the first examination studying effects of liquidity on intraday settlement.

The paper is organised as follows. We begin in Section 5.2 by briefly describing the current large-value payment landscape in Japan, and how the design of the new system is expected to alter that landscape. We also provide a rough description of the planned new system and explain the purpose of the new account and its funding. In

Section 5.3 we examine changes in liquidity efficiency of combining the two new payment streams with the payments on the current BOJ-NET. In Section 5.4 we describe the problem of finding optimum funding levels, and in Section 5.5 we present the results of simulation analysis. In Section 5.6 we provide a short summary and conclusion.

5.2 Large-value payments in Japan

5.2.1 Current structure of large-value payment systems

BOJ-NET plans to incorporate payments currently made on BOJ-NET, the Foreign Exchange Yen Clearing System (FXYCS) and the large-value payments on the Zengin Data Telecommunication System (Zengin). We briefly describe some aspects of these three systems.[2]

BOJ-NET is a pure RTGS system for the Japanese yen, owned and operated by the Bank of Japan. The system is one of the core financial infrastructures supporting economic and financial activities in Japan. It settles almost JPY 100 trillion daily with annual turnover ranging 40 times as high as Japan's nominal GDP.

BOJ-NET handles both Japanese government Securities (JGSs) and funds transfers. The latter mainly consist of money-market transactions, but also include the settlement payments for various payment and securities settlement systems that use BOJ-NET to transfer the final settlement payments and the cash legs. In addition, money-market operations of the Bank of Japan are carried out using BOJ-NET. There are a limited number of third-party, or customer, payments settled on BOJ-NET, and those are very high-value payments, indicating that these are also money-market transactions conducted by market participants that do not have accounts with the Bank of Japan. Settlement amounts in 2005 indicated that on a daily average basis BOJ-NET settled 21,641 transfers with a total value of JPY 88.3 trillion. The average value per settlement was JPY 4.1 billion.

FXYCS is basically a DNS system that handles yen legs of foreign exchange trades. It conducts the final settlement at 14:30 using BOJ-NET. The volume and value of its daily average activities in 2005 indicated that it settled 28,022 transactions per day with a total value of JPY 16.4 trillion. The average value per transaction was JPY 586

---

[2] For an overview of payment systems in Japan, see the Japan section of BIS (2003).

million. The net amount transferred on BOJ-NET in 2005 averaged JPY 4.1 trillion. FXYCS has not only a DNS mode but also an RTGS mode, although its use is rather limited.

Finally, Zengin is a simple DNS system, whose final payment takes place at 16:15. In 2005, Zengin averaged 5.4 million transactions per day with a total daily average value of JPY 9.5 trillion. The average size of payments was JPY 1.8 million. It is mainly used for commercial payments. On average, the daily settlement amounts made through BOJ-NET were JPY 1.8 trillion per day in 2005. It is estimated that roughly two-thirds of the value transferred on Zengin, approximately JPY 6 trillion per day, is made up of payments that are larger than JPY 100 million.

5.2.2 Future structure of large-value payment systems

The new system plans to operate as a queue-augmented RTGS system.[3] The new liquidity-saving features will be provided on a new type of accounts as shown in Table 1. Participants will be able to designate payment instructions to be settled either via the new accounts, that will not offer intraday overdrafts capability, or via the standard accounts, on which collateralised overdrafts will remain available. The intent of both participants and the Bank of Japan is that most of the three payment streams described above will be settled via the new accounts. The standard accounts and the dedicated accounts for simultaneous processing of delivery-versus-payment and collateralisation, known as SPDC, will still operate and are intended to be used for the rest of settlements.[4]

The new system will operate the new accounts as follows. The new accounts will be funded by participants each morning at the start of the processing day (9:00) with an infusion of funding from the standard accounts. That establishes the participants' initial balances in the new accounts, because the new accounts will have a zero balance overnight. Participants will then submit payment instructions to the

---

[3] See BIS (1997), McAndrews and Trundle (2001), and BIS (2005) for basic ideas on a queue-augmented RTGS.

[4] The SPDC facility is another type of liquidity-saving facility used only for settlement of cash legs of JGSs transactions. It allows the receiver of JGSs to pledge the incoming securities as collateral for intraday overdrafts while using the overdrafts to pay for the incoming securities. Similarly, the deliverer of JGSs is able to withdraw the securities pledged with the Bank of Japan for delivery to the receiver while using the funds received to repay the overdrafts.

new accounts, and a bilateral offsetting algorithm will initiate a search for bilaterally offsetting payments on a FIFO basis. If a pair of bilaterally offsetting payments is found, and if funds are sufficient to settle the payments, settlement of the selected payments takes place simultaneously. At designated times, a multilateral offsetting algorithm will attempt to find the largest set of payments that can be settled using available balances.[5] See Appendix 1 for the details of bilateral and multilateral offsetting algorithms in the new system.

Table 5.1 **Account structure in the new system**

	Standard account	SPDC account	New account
Types of transactions settled	– interbank transfers (eg money market, foreign exchange) – third-party transfers – the cash legs of securities transactions – settlement obligations arising from clearing systems – transactions with BOJ/ government	– the cash legs of JGSs transactions using the SPDC facility	– interbank transfers (eg money market, foreign exchange) – third-party transfers (including large-value Zengin payments)
Liquidity supply	Intraday overdrafts	Intraday overdrafts, liquidity transfers from standard account	Liquidity transfers from standard account
Liquidity saving	Not applicable (pure RTGS)	SPDC facility	Queuing and offsetting mechanisms
Account management	Overnight	Intraday (zero balance at the end of the processing day)	Intraday (zero balance at the end of the processing day)
Opening and closing times	9:00–17:00*	9:00–16:30	9:00–16:30

* Closing time is 19:00 for participants that have applied for access to extended hours.

Participants will be able to transfer funds between their new accounts and their standard accounts freely throughout the day. Payment instructions remaining in the queue will be rejected if insufficient

---

[5] The algorithm will include all queued payments in the initial offsetting and successively drop the largest payment from the participant with the largest funding shortfall until a set of payments that have no funding shortfalls is found. Bech and Soramäki (2001) show that this algorithm finds the largest set of payments that can be settled using a multilateral offsetting given that one breaks a FIFO ordering rule.

funds are submitted to the new accounts by 16:30. The standard accounts will remain open until 17:00.

5.3 Liquidity effects of combining FXYCS, Zengin and BOJ-NET payments

As described above, the new system plans to incorporate payments currently made on BOJ-NET and FXYCS, and the large-value payments on Zengin. The question is whether the combination of these payment streams increases liquidity efficiency by aggregating the currently fragmented payment systems or reduce it by eliminating the DNS systems but with the obvious benefit of permitting intraday settlement of payments. We examine this question by first simulating operations of the new system with payments that are currently settled in BOJ-NET. Then we conduct simulations of the performance of FXYCS and the large-value Zengin, using the settlement method of the new system, while assuming (contrary to the planned design) that they were separately operated from BOJ-NET. Adding liquidity required in each of these two simulations provides an indication of liquidity that would be used if BOJ-NET, FXYCS, and Zengin remained separate systems, but all adopt an intraday finality capability. Finally, we simulate the performance of the new system when payment streams from all these systems are combined and settled in the same system. If liquidity required to settle the combined payment streams is lower than that required to settle the payments when the systems are operated separately (for a fixed level of delay), then it can be expected that there are liquidity complementarities, or scale economies in liquidity use, in combining the payment streams. If, on the other hand, liquidity use is lower with the systems operated separately, then there are diseconomies in liquidity use in combining the systems.

For each system, we conduct three treatments on each day's data (the ten days of historical data in September 2003 are used in the simulations that we report on here).[6] The first treatment is to endow participants with sufficient liquidity to settle the day's payments without delay. The second is to endow them with sufficient liquidity only to settle their multilateral net debit, with which the payments will be settled as quickly as possible (using the new settlement method).

---

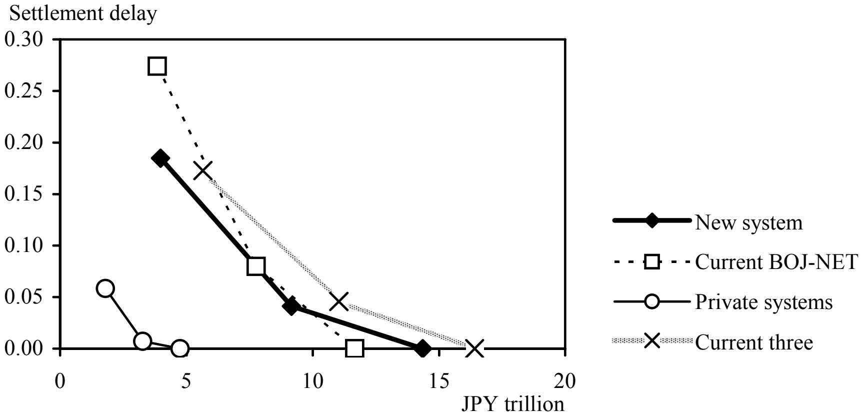
[6] See Appendix 1 for the summary statistics of simulation data.

Finally, in the third treatment, participants are endowed with the average of the two other levels of liquidity – in other words, they are endowed with liquidity that is halfway between the level sufficient to settle payments without delay and the level of multilateral net debits.

We examine a trade-off between liquidity necessary to settle the payments and delay with which the payments are settled. If the locus of points that describes this trade-off shifts inward or outward as the different payment streams are added, it can be said that there are liquidity efficiencies or costs respectively in combining the different payment streams.

The results of these simulations, using the ten days of historical data and the settlement method of the new system, are shown in Figure 5.1. On average it is found that there are significant liquidity complementarities in combining the payment streams. This can be seen clearly in the inward shift of the black line (new system), which illustrates the performance of the new system, relative to the grey line (current three), which illustrates the total liquidity requirements of the three systems when operated separately. The inward shifts show that at all the three levels of delay simulated the new system requires less liquidity to settle the payments.

Figure 5.1 **Delay indicator and liquidity for the separate systems, the sum of the separate systems operating in isolation, and for the new system**



Source: Authors' calculation.

Table 5.2 provides more details on each of the ten days of simulated data and presents both the delay indicator measure and the value-weighted average time of settlement.[7] In every simulation, and for any average time of settlement or any indicator of delay of settlement, the new system requires less liquidity to settle the payments. The results therefore suggest that there are significant liquidity complementarities, or economies of scale in liquidity use associated with the combination of the payment streams from the three systems. On average, across the treatments and the days, combining the payment streams results in 20% reduction in liquidity use.

Table 5.2 **Liquidity use, delay indicator and value-weighted average time of settlement for the separate systems and for the new system**

JPY billion; hh:mm

	Level (1)	Level (2)	Level (3)
New system			
Liquidity	3,975	9,159	14,344
Delay	0.185	0.041	0.000
Average time	12:22	11:38	11:26
Current three systems			
Liquidity	5,649	11,032	16,415
Delay	0.173	0.042	0.000
Average time	12:17	11:39	11:26
Current BOJ-NET			
Liquidity	3,850	7,760	11,670
Delay	0.274	0.042	0.000
Average time	12:56	11:39	11:34
Two private systems			
Liquidity	1,799	3,272	4,745
Delay	0.058	0.007	0.000
Average time	11:34	11:18	11:16

Source: Authors' calculation.

Note: Level (1) endows participants with sufficient liquidity only to settle their multilateral net debit, Level (2) with liquidity that is halfway between the level sufficient to settle payments without delay and the level of the multilateral net debits, and Level (3) with sufficient liquidity to settle payments without delay.

---

[7] Specific definitions of these indicators are described in Appendix 1.

It is an interesting feature of the system that the current BOJ-NET requires less liquidity than the new system to process its payments without delay, but requires almost the same level of liquidity as the new system to settle its payments on a multilateral net basis. This suggests that as some of FXYCS and large-value Zengin payments arrive later in the day, they offset with some current BOJ-NET payments that arrive earlier in the day but still remain in the queue. As the current BOJ-NET payments are settled with a slight delay, they settle with less liquidity when combined with payment streams from the other two systems. Again, this indicates particularly strong liquidity complementarities among the systems. It should also be noted that while the combined payments settle without delay using more liquidity, a close examination of Table 5.2 shows that the new system settles at an earlier hour of the day than the current BOJ-NET where participants are endowed with sufficient liquidity to settle payments without delay.

5.4 Optimising funding levels

The funding levels in the new accounts will be determined by a choice of participants. In general, the higher the funding levels, the greater a proportion of those payments that are submitted to the new accounts can be settled. In addition, the higher the funding levels, the more quickly settlements will occur.

A feature of the new system is that funding for the new accounts can be supplied from the standard accounts at any time of the day. To some degree, this option simplifies the problem for participants regarding the amount of funding to transfer to the new accounts at the start of the processing day as any shortfalls or overages in funding can be corrected during the day.

When designing a payment system that uses a liquidity-saving mode of operations as well as a pure RTGS mode of operations, one question designers face is whether to create another account, as in the BOJ-NET's new accounts. One choice is simply to rely on a single account and have participants decide on the priority of the payment, in other words, decide whether to send the payment instruction in a pure RTGS or in a liquidity-saving mode. The liquidity-saving mode then relies on incoming funds over a period of time as well as offsetting. Such a choice is described by Johnson, McAndrews, and Soramäki (2004). In the case of the new system, the computational requirements of BOJ-NET are reduced considerably with the introduction of the new accounts.

The efficiency of the new system could potentially be negatively affected if participants were to transfer funds into and out of their new accounts often during the day. The multilateral offsetting algorithm, for example, might not find many payments that can be settled if some participants had withdrawn funds immediately prior to operations of the algorithm. Because of this potential negative effect of rapid changes in funding levels, it may be useful to conduct the following thought experiment. Suppose, contrary to the design of the new system, that participants could only fund their new accounts twice during the day, at the opening of the processing day and for settlement of their unsettled queued payment instructions at 16:00. Under that counterfactual assumption, what would be efficient levels of initial funding?

Higher levels of initial funding will be associated with a faster rate of intraday settlement and a higher proportion of payments settled prior to 16:00. There is, however, no clear answer to the question of how to value an increased rate of intraday settlement as there is no easily observable intraday rate of interest that would provide a benchmark level of benefits from a faster rate of intraday settlement and a benchmark level of costs of intraday funds. Similarly, there is no clear measure of increases in credit and liquidity risks caused by leaving more payments unsettled until 16:00.

In the following exercises we investigate levels of initial funding that are sufficiently high so as to quicken the overall settlement of large-value payments in Japan. In addition, we investigate funding levels high enough to assure that a level of unsettled payments at 16:00 is no greater than it is in today's large-value payment systems.

Consider the following problem.

$$\min \sum_{i} b_i \text{, subject to } \quad \{P_{ij}\},\ \forall i, j\,;\ i \neq j$$

$$b_i \geq 0$$

$$\sum_{t=t_k}^{t_k+h} \sum_{i} \sum_{j} s_{ij}^{t} \geq S,\ \forall 0 \leq k \leq \bar{k},\ \bar{h} > h > 0.$$

It seeks to minimise the sum of initial balances of each participant i in the new account (b_i), under the constraints that a set of payments that day is fixed and given by P_{ij}, that the balances are non-negative and that settlement (in a value term) under the new system procedures over a given time interval during processing is at least as high as a rate of settlement S, where S is some yen-rate of settlement per h minutes of the day.

By examining the structure of the problem, we can infer that the optimal levels of initial balances satisfy the following 'equalisation of marginal benefit condition'. An extra yen added to any participant's initial balance has the same incremental effect on the total settlement as an extra yen added to any other participant's initial balance. We can infer that because the variables of initial balances enter the objective function in an additively separable way, there cannot be any way, at the optimal level of balances, to shift balances among accounts (holding fixed the sum of balances) and increase a rate of settlement. Otherwise we could reduce the sum of balances from the minimum level, which contradicts that the level is at a minimum. From that, it must then be the case that an extra yen of initial balances increases a rate of settlement by the same amount regardless of into whose account that yen is added.

The problem outlined above is not fully specified as it does not contain full richness and complexity of the settlement algorithms used by the new system. Nonetheless, an examination of the problem clarifies the heuristic strategy we employ in seeking the efficient levels of initial funding for the new accounts. First, notice that a rate of settlement is specified as the sum of all payments settled. The goal is therefore not to increase a particular participant's rate of settlement but to increase a rate of settlement for the whole system. Second, the problem seeks to minimise the sum of initial balances, not any participant's initial balance. Thus the efficient levels of funding we discuss are characterised by the following three factors: the total level of funding, the distribution of balances across participants, and the timing of funding.

5.5 Simulations and results

To find a locally optimum distribution of balances using simulations on historical data would require a large number of simulations. It is rational that we rely on that feature of the optimum levels of initial balances to guide the following heuristic strategy to characterise the efficient levels of balances. We first simulate the working of the new system starting with various levels of initial balances. After each simulation we examine the performance of the system in terms of the value of payments settled prior to 16:00, the value of the remaining unsettled payments at that time, the value of additional amounts that need to be paid in to settle all the remaining unsettled payments, and the value-weighted average time of settlement. We also examine the effects of alternative levels of balances on the system as a whole, and on a separate basis, for the five largest banks and all the other participants. We then investigate the intertemporal distribution of balances as we seek a local optimum distribution of balances.

The results of these simulations give participants and planners a sense of how the alternative levels of balances would affect the system's performance.

5.5.1 Four baseline simulations

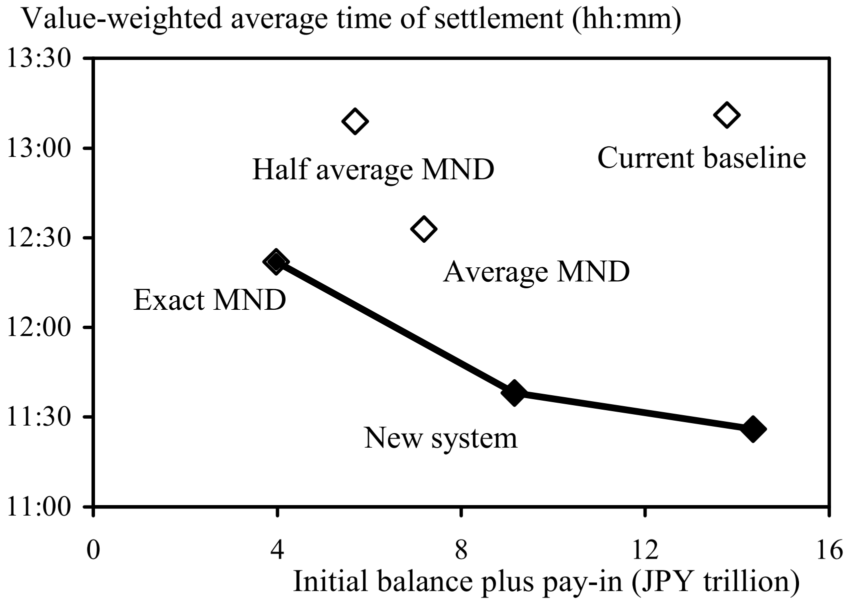
We perform simulations using the ten days of historical data in September 2003. We conduct four sets of baseline simulations. The first scenario is to simulate the performance of the current situation in which BOJ-NET, FXYCS and the large-value Zengin independently operate as they operate now. The scenario endows participants with sufficient liquidity to settle their payments without delay (although it treats FXYCS and Zengin as simple DNS systems) and uses the time of entry of payments. As a result, these baseline simulations provide a measure of current liquidity usage in the systems. These simulations are referred to as *current baseline* simulations.

Another baseline simulation is to endow participants with the exact amount of funds (in the new accounts) equal to that day's multilateral net debit of each participant, given that day's payments history. A participant's multilateral net debit is the amount it would owe to settle its payments if the system were a DNS system. In general, participants do not necessarily know their own multilateral net debits in advance. This scenario can be thought of approximating the case in which participants make pay-ins throughout the day as they gradually learn the exact size of their multilateral net debit. The multilateral offsetting operations may be one way participants do learn the amount of their multilateral net debits, and this scenario approximates the learning process by assuming that they know the amounts with certainty in advance. These simulations are referred to as *exact multilateral net debit (MND) funding* simulations or *progress-payment approximation* simulations.

The third baseline simulation endows participants with their average multilateral net debit funding, where the average is taken over the ten days of the sample period. This scenario is first to assume that participants fund their new accounts in the morning and then make another pay-ins to the new accounts after 16:00 to settle the payments that remain unsettled at that time. The average multilateral net debit is, of course, quite close in size to the exact multilateral net debit amount used in the *exact MND funding* simulations. However, because it is an average, some payments on some days will remain unsettled at 16:00. These simulations are referred as *average multilateral net debit (MND) funding* simulations.

The fourth baseline simulation endows participants with half the amount of funding as in the *average MND funding* simulations. These simulations are referred as *half average multilateral net debit (MND) funding* simulations.

Figure 5.2 **Overview of the performance of the new system**



Source: Authors' calculation.

Figure 5.2 summarises the performance of the new system described in Section 5.3 and of these four baseline simulations. Points in the lower-left corner of the figure are more desirable combinations of total balances and settlement time. It can be found that conducting these baseline simulations attempts to search the local optimum level around the point at which participants are endowed with sufficient liquidity only to settle their multilateral net debits.

Table 5.3 shows the performance of these four baseline simulations on average across the ten days of the sample period with regard to the amounts of initial balances used in the simulations, the additional amounts of pay-ins to the new accounts that would be required after 16:00 to settle those payments that still remain unsettled at that time, the cumulative amounts settled by 16:00, the gross amounts unsettled at 16:00, and the value-weighted average time of settlement. Because the analysis of only ten days yields a small sample, we simply examine averages without considering the statistical significance.

Table 5.3 **Averages from the baseline simulations**

JPY billion; hh:mm

	Initial balances	Five LBs' balances	End-of-day pay-ins	Cumulative value settled at 16:00	Gross value unsettled at 16:00	Average time of settlement
Current baseline	13,780 (-)	3,460 (-)	0 (-)	56,673 (-)	12,625 (-)	13:11
Exact MND	3,975 (0.288)	492 (0.142)	0 (-)	61,106 (1,078)	8,192 (0.649)	12:22
Average MND	3,964 (0.288)	492 (0.142)	3,224 (-)	55,954 (0.987)	13,344 (1.057)	12:33
Half average MND	1,982 (0.144)	246 (0.071)	3,712 (-)	48,119 (0.849)	21,180 (1.678)	13:09

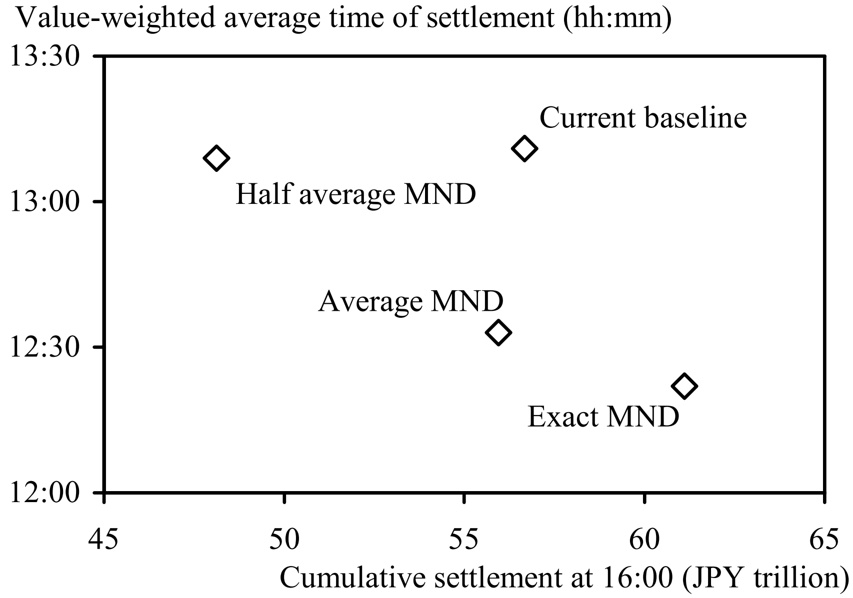
Source: Authors' calculation.

Note: Figures in brackets are ratios of each item to that of the *current baseline* simulations. 'Five LBs' stands for five largest banks.

The exact *MND funding* simulation clearly settles more payments by 16:00 with the initial balances as small as one-third of those the *current baseline* simulation requires. The *average MND funding* simulation also has the same qualitative results relative to the *current baseline* simulation, using fewer initial balances than the *current baseline* simulation. The *average MND funding* simulation results that payments unsettled at 16:00 reach up about 20% of that day's total payments. These payments would be settled with an additional pay-in of JPY 3.2 trillion, so that the total liquidity used in these simulations is about twice as high as in the *exact MND funding* simulation. The amounts settled by 16:00 in the *half average MND funding* simulation are far below those in the three other scenarios, though economising too much of initial balances. The *half average MND funding* simulation settles on average only slightly more quickly than the *current baseline* simulation, using much less liquidity than the *current baseline* simulation. Because of its larger pay-in after 16:00, the *half* *average MND funding* simulation uses almost as much liquidity in total as the *average MND funding* simulation.

Figure 5.3 shows the value-weighted average time of settlement and the cumulative settlement by 16:00 for the various cases. The settlement performance improves as the outcome plotted on the figure moves toward the bottom right, meaning a larger value settled in a quicker manner and vice versa. The four scenarios can be roughly arranged in the desirable order as the *exact MND funding* simulation, the *average MND funding* simulation, the *current baseline* simulation and the *half average MND funding* simulation.[8]

Figure 5.3 **Value-weighted average time of settlement and total value settled by 16:00**



Source: Authors' calculation.

Overall, the *exact MND funding* simulation settles payments most quickly and extensively and uses less liquidity than the *average MND funding* simulation. This suggests that if participants were to make pay-ins during the day in line with their multilateral net debit

---

[8] The *current baseline* simulation may be better than the *average MND funding* simulation, depending on the shape of indifference curves assumed. For example, the former improves if a high preference is given to settlement completion by 16:00.

positions, they might be able to have fewer payments unsettled after 16:00. In comparing the performance of the *average MND funding* simulation and the *half average MND funding* simulation, the latter settles fewer payments by 16:00 and has a later average time of settlement (although it also settles payments more quickly than the *current baseline* simulation on average). It has approximately 25% of the payments unsettled at 16:00. Settlement of these payments requires an additional pay-in of JPY 3.7 trillion. The *half average MND funding* simulation, after all, uses about 80% of liquidity used in the *average MND funding* simulation, after taking into account the large pay-ins at the end of the day. This result reminds one that as one limits the initial amount of liquidity available to the system, larger pay-ins will be required later in the day.

The results of these four baseline simulations suggest that the new system may perform quite satisfactorily with levels of liquidity that are significantly lower than those currently used in settlement of the three systems. In addition, the behaviour of a rough approximation to the progress payments suggests that participants may be better able to conserve funding by making pay-ins to the system during the day as they learn the multilateral net debit resulting from that day's payments.

5.5.2 Distributional funding simulations

As the results of the *exact MND funding* and *average MND funding* simulations have suggested, the different distribution of initial balances across participants leads to the different performance of intraday settlement even when the total balances in the system are the same.

It is well known that there are a few hub-like participants in Japan's interbank payment network.[9] They play a significant role in the redistribution of liquidity in the system by making outgoing payments and receiving incoming payments continuously during the day. Therefore the malfunctioning of these hub-like participants potentially has negative effects on the performance of the system as a whole.

In this section, in addition to the baseline simulations, we perform some additional simulations that show the effects of small changes in

---

[9] For the structure of Japan's payment network, see Inaoka et al (2004) and Bank of Japan (2006a).

the funding provided by the five largest banks, which are known to work as hub-like participants in BOJ-NET. These simulations are conducted with the other participants in the system being endowed first with the exact multilateral net debit funding and, for the second set of these simulations, with half that level of funding. Because those participants are endowed with the exact amount of their multilateral net debit, these simulations are probably best compared with the *exact MND funding* simulation. The amounts that the five largest banks are endowed with are quite small amounts equal to the 90th percentile of the size of the payments they each send and receive on the current BOJ-NET alone. So these simulations are indicative of a situation in which all but the five largest banks make regular progress payments in the amounts of their multilateral net debits, and the five largest banks supply very little in the initial funding amounts. These simulations are not meant to model the actual behaviour of participants but rather to investigate the possible behaviour of the new system as we vary the funding of some particular participants in different ways.

These simulations are quite illustrative of the effects of small changes in particular participants' funding levels. To investigate these effects for individual participants would be quite time-consuming and require many simulations. Because of those resource requirements, we forego such an investigation in the paper.

The first set of simulations shows that reducing the five largest banks' total funding from JPY 492 billion, as in the *exact MND funding* simulation, to JPY 18 billion does not substantially reduce the speed of settlement in the system (see Table 5.4). The value-weighted average time of settlement changes from 12:22 to 12:34. Nor is the total amount settled by 16:00 reduced appreciably, even though the largest five banks had multilateral net debits of approximately JPY 500 billion on the sample days. These results show that individual participants, or even groups of participants, may significantly reduce their initial level of funding without necessarily causing proportional changes in the amounts settled. Note that these results come at the cost of large amount of end-of-day pay-ins. Further research could determine the local optimum in the initial funding amounts.

Table 5.4 **Averages from the simulations with the 90th percentile funding**

JPY billion; hh:mm

	Initial balances	Five LBs' balances	End-of-day pay-ins	Cumulative value settled at 16:00	Gross value at 16:00	Average time of settlement
(1) Exact MND	3,975	492	0	61,106	8,192	12:22
	(-)	(-)	(-)	(-)	(-)	(-)
+90 percentile	3,500	18	1,527	58,170	11,129	12:34
	(-475)	(-474)	(+1,527)	(-2,936)	(+2,937)	(+0:12)
+90 percentile*2	3,518	35	1,452	58,495	10,803	12:34
	(-457)	(-457)	(+1,452)	(-2,611)	(+2,611)	(+0:12)
+90 percentile*3	3,535	53	1,405	59,025	10,274	12:33
	(-440)	(-439)	(+1,405)	(-2,081)	(+2,082)	(+0:11)
(2) Average MND	3,964	492	3,224	55,954	13,344	12:33
	(-)	(-)	(-)	(-)	(-)	(-)
+90 percentile	3,490	18	3,398	54,172	15,128	12:43
	(-474)	(-474)	(+174)	(-1,782)	(+1,784)	(+0:10)
+90 percentile*2	3,507	35	3,371	54,056	15,243	12:42
	(-457)	(-457)	(+147)	(-1,898)	(+1,899)	(+0:09)
+90 percentile*3	3,525	53	3,366	54,621	14,678	12:41
	(-439)	(-439)	(+142)	(-1,333)	(+1,334)	(+0:08)
(3) Half average MND	1,982	246	3,712	48,119	21,180	13:09
	(-)	(-)	(-)	(-)	(-)	(-)
+90 percentile	1,754	18	3,756	46,017	23,282	13:19
	(-228)	(-228)	(+44)	(-2,102)	(+2,102)	(+0:10)
+90 percentile*2	1,772	35	3,724	46,350	22,948	13:18
	(-210)	(-211)	(+12)	(-1,769)	(+1,768)	(+0:09)
+90 percentile*3	1,789	53	3,720	46,494	22,804	13:17
	(-193)	(-193)	(+8)	(-1,625)	(+1,624)	(+0:08)

Source: Authors' calculation.

Note: 'Five LBs' stands for five largest banks. Figures in brackets are differences from the benchmark level of each sub-scenario.

The second set of simulations endows all but the largest five banks with their average multilateral net debit amounts, as in the *average MND funding* simulations (see Table 5.4). The largest five banks are again endowed with an amount that is equal to the size of the payment that is at the 90th percentile of their payment size distribution on the current BOJ-NET alone. In this simulation, which is best compared with the *average MND funding* simulations, we see that the performance of the system remains quite good even though the largest five banks' funding levels are reduced substantially. The amounts settled by 16:00 falls by only 3%, and the value-weighted average time of settlement occurs 10 minutes later.

A final set of these simulations, in which participants other than the largest five banks have their initial funding levels set at half of the day's multilateral net debit, confirms the result that dramatically reducing the funding levels of the largest five banks does not reduce settlement by that proportion (see Table 5.4).

In each set of the simulations just discussed, we vary the funding levels of the five largest banks by endowing them with multiples of JPY 18 billion, namely 35 (doubled) and 53 (tripled) for their initial balances. These increases in the levels of initial balances do not appreciably change the outcome. One reason is that liquidity-saving features effectively reduce some distortions from optimal balances by running offsetting mechanisms continuously during the course of the day. Offsetting mechanisms can relax conditions for gross settlement in comparison with a pure RTGS mode and then achieve relatively smoother flow of payments despite the distortions of initial distribution of balances.

In general, there tends to be a greater amount settled as the initial funding levels of the largest five banks increases, but this is not always true. For example, raising the largest five banks' initial funding from JPY 18 billion to 35 slightly reduces the amounts settled by 16:00 in the second set of simulations. This result implies that the amount settled by 16:00 is not a monotone increasing function of some particular participants' initial balances.

5.5.3 Progress-payment simulations

The *exact MND funding* simulation has endowed participants with the exact amounts of the multilateral net debit at the beginning of the processing day. This simulation can also approximate the case in which participants make pay-ins continuously during the day as they learn the size of their multilateral net debit in that day. The question is how the performance in the system can be affected if the timing of intraday pay-ins is changed.

It has been already described that the *half average MND funding* simulation substantially underperforms the *exact MND funding* simulation because of the severe liquidity constraints in the system. In the progress-payment simulations, starting with the half average multilateral net debits and then making intraday pay-ins at 10:00 or 12:00, both the value settled by 16:00 and average time of settlement can approach those of the *exact MND funding* simulation (see Table 5.5). The high performance of the progress-payment simulations with intraday pay-ins comes at the cost of twice as large amount of the total liquidity in the *exact MND funding* simulation.

Table 5.5 **Averages from the *progress-payment approximation* simulations**

JPY billion; hh:mm

	Initial balances	Intraday pay-ins	End-of-day pay-ins	Cumulative value settled at 16:00	Gross value unsettled at 16:00	Average time of settlement
(1) Half average MND	1,982	0	3,712	48,119	21,180	13:09
	(-)	(-)	(-)	(-)	(-)	(-)
+ Exact MND at 10:00	1,982	6,095	2,780	61,621	7,678	11:51
	(0)	(+6,095)	(-932)	(+13,502)	(-13,502)	(-1:18)
+ Exact MND at 12:00	1,982	5,571	2,302	62,681	6,617	12:10
	(0)	(+5,571)	(-1.410)	(+14,562)	(-14,563)	(-0:59)
(2) Half average MND	1,982	0	3,712	48,119	21,180	13:09
	(-)	(-)	(-)	(-)	(-)	(-)
+ Half exact MND at 10:00	1,982	3,047	3,202	59,152	10,146	12:15
	(0)	(+3,047)	(-510)	(+11,033)	(-11,034)	(-0:54)
+ Half exact MND at 12:00	1,982	2,785	3,094	59,076	10,223	12:30
	(0)	(+2,785)	(-618)	(+10,957)	(-10,957)	(-0:39)

Source: Authors' calculation.

Note: Figures in brackets are differences from the benchmark level of each sub-scenario.

In comparing the scenarios with additional pay-ins at 10:00 and at 12:00, both of them can achieve almost the same level of the value settled by 16:00. The average time of settlement, however, is further improved with the additional pay-ins at 10:00 rather than with the additional pay-ins at 12:00. The earlier arrangement reduces the duration of the payments unsettled and then leads to the earlier average time of settlement. In comparing the performance of the intraday pay-ins with the exact multilateral net debit and the half of that, the latter settles fewer payments by 16:00 and has a later average time of settlement.

Participants are required to add intraday pay-ins at the appropriate timing to secure the sufficient funding. With such a careful management of liquidity and payment flows, smoother flow of payments can be achieved in the system. However, participants can learn the optimum timing of funding only ex post. The second-best solution to the optimum funding problem subject to a certain rate of settlement is, therefore, to endow participants with the exact amount of the multilateral net debit at the beginning of the processing day.

5.6 Concluding remarks

The new system with liquidity-saving features will require a level of liquidity below that necessary for intraday settlement in the current BOJ-NET. In the paper we have explored characteristics of the optimum funding level in the new system using simulation analysis. More specifically, we have analysed how quickly intraday settlement could occur if the level of initial funding were subject to some liquidity constraints. Our findings are summarised as follows.

(1) To minimise the total balances in the system subject to a certain level of the progress rate of intraday settlement, participants need to secure timely funding and to appropriately distribute the total balances across them in the system. In fact, the timing of funding may actually be less controllable because participants could hardly learn the optimum timing of funding ex ante. The simulation results suggest that it is one of the second-best arrangements for the local optimum of balances to endow participants with the multilateral net debit amounts at the beginning of the processing day.

(2) Offsetting mechanisms search for a set of payment instructions that can be settled when taking into account incoming payments as sources of liquidity as well as actual balances in accounts at that point. These mechanisms have side effects on the cross-sectional and intertemporal distribution problem of balances in the system. Through relaxing conditions for gross settlement, these mechanisms are expected to conduct some fine-tuning during the course of the day to reduce a certain level of distortion from optimum balances.

(3) The simulation analysis also indicates strong economies of scale in liquidity use in Japan's large-value payments. It suggests that participants enjoy liquidity efficiencies in combining the different payment streams rather than in operating individual payment systems separately.

Solving the optimisation problem for funding by using simulation analysis would require a large number of simulations. Although this work is supposed to be quite time-consuming, it gives participants and planners a sense of how alternative levels of funding would affect the system's performance.

References

Bank for International Settlements (1997) **Real-Time Gross Settlement Systems.** Prepared by the Committee on Payment and Settlement Systems.

Bank of International Settlements (2003) **Payment and Settlement Systems in Selected Countries.** Prepared by the Committee on Payment and Settlement Systems.

Bank of International Settlements (2005) **New Developments in Large-Value Payment Systems.** Prepared by the Committee on Payment and Settlement Systems.

Bank of Japan (2006a) **Financial System Report.** Prepared by the Financial System and Bank Examination Department.

Bank of Japan (2006b) **Japan's Next-Generation RTGS.** Prepared by the Payment and Settlement Systems Department. Bank of Japan Research Papers.

Bech, M – Soramäki, K (2001) **Gridlock resolution in interbank payment systems.** Bank of Finland Discussion Papers, No. 9.

Inaoka, H – Ninomiya, T – Taniguchi, K – Shimizu, T – Takayasu, H (2004) **Fractal Network Derived from Banking Transaction – An Analysis of Network Structures Formed by Financial Institutions.** Bank of Japan Working Paper Series, No. 04-E-04.

Johnson, K – McAndrews, J – Soramäki, K (2004) **Economising on Liquidity with Deferred Settlement Mechanisms.** Federal Reserve Bank of New York Economic Policy Review.

Leinonen, H – Soramäki, M (1999) **Optimising Liquidity Usage and Settlement Speed in Payment Systems.** Bank of Finland Discussion Papers, No. 16.

McAndrews, J – Trundle, J (2001) **New Payment System Designs: Causes and Consequences.** Bank of England Financial Stability Report.

Appendix 1

Offsetting algorithms in the new system

Offsetting mechanisms search for a set of payment instructions that can be settled when taking into account incoming payments as sources of liquidity and settle the selected instructions simultaneously. In the new system, a bilateral offsetting algorithm will run continuously throughout the day, with a multilateral offsetting algorithm running a few times a day to complement the bilateral offsetting algorithm.

The bilateral offsetting algorithm will search for a pair of bilaterally offsetting payment instructions or a single instruction that can be settled on a gross basis. It will run when one of the following events occurs: (i) a new payment instruction entering the system; (ii) an increase in balances of the new account; (iii) a change in the payment instruction at the top of the queue due to settlement, reordering, or cancellation. The target payment instruction for bilateral offsetting is the newly submitted payment instruction when (i) occurs and the top-queued payment instruction when (ii) or (iii) occurs.

For example, where the target payment instruction is a newly-submitted payment from Bank A to Bank B, the system searches from the top of the queue for a payment instruction from Bank B to Bank A that can be settled simultaneously using available balances.

The multilateral offsetting algorithm will run at fixed times. It will attempt to find the largest set of queued payment instructions that can be settled using available balances by first testing to settle all queued payment instructions at once and successively removing the largest queued payment instruction from the participant with the largest funding shortfall until a set of payment instructions that causes no funding shortfalls can be found.

Profile of the simulator

We use the BOJ-NET simulator developed by the Yajima Laboratory of the Tokyo Institute of Technology, whose research interests are focused on mathematical programming and operations research. Its basic functions are almost the same as those of the Bank of Finland Payment and Settlement Simulator.[10] Highly complicated offsetting algorithms with a settlement-value maximisation or time-weighted

---

[10] See Leinonen and Soramäki (1999) for the Bank of Finland Simulator.

average settlement-value maximisation mode are available on the BOJ-NET simulator as well as standard offsetting algorithms based on a FIFO ordering rule, which are described above.

Simulation data

The simulations were performed using Japan's actual data of ten consecutive business days in September 2003. The data includes the following transactions: money-market transactions (excluding those with the Bank of Japan); foreign exchange yen transactions (excluding CLS related transactions), which are handled either on a DNS mode or an RTGS mode in FXYCS; and large-value retail credit transfers, which are JPY 100 million and over per transaction. See Table 5.6 for a summary of the basic statistics.

Table 5.6 **Basic statistics on the simulation data**

JPY billion

	Daily average volume	Daily average value	Average value	SD of value
Total transaction	61,709	69,979	1.134	7.851
MM transactions	7,558	37,487	4.960	20.134
FXY transactions	40,368	23,010	0.570	3.801
LV Zengin transactions	13,783	9,483	0.688	1.483

Source: Authors' calculation based on data from the Japanese Bankers Association and the Bank of Japan.

Measurement of simulation results

A settlement delay of a payment instruction can be calculated as the time difference between the payment submission to the system and the completion of the payment. We use two types of statistics to measure a settlement delay in the system: the value-weighted average time of settlement and the indicator of settlement delay.

The value-weighted average time of settlement (VWATS), which is the average time (measured from the opening of the processing day) weighted by the value of payments settled, is defined as follows:

$$\mathrm{VWATS} = \sum_i t_i \cdot v_i \Big/ \sum_i v_i$$

where t_i and v_i represent respectively the settlement time (minutes) and the value of a payment i. If all payments are settled at the opening of the processing day (9:00), then VWATS has a minimum value of zero minutes because $t_i = 0$ for all i. If no payments are settled during the day, and if all the payments are settled at the end of the processing day (16:30), then VWATS takes a maximum value of 450 minutes because $t_i = 450$ for all i.

In the meanwhile, the indicator of settlement delay (ISD) is defined as follows

$$\mathrm{ISD} = \sum_i \left(t_{2,i} - t_{1,i}\right) v_i \Big/ \sum_i \left(t_{\mathrm{end}} - t_{1,i}\right) v_i$$

where $t_{1,i}$ and $t_{2,i}$ are respectively the submission time and the settlement time of a payment i, and t_{end} is the time for the end of the processing day (16:30). ISD runs from zero, which means no delay in the system, through one, which means no settlement during the day. See Bech and Soramäki (2001) for further discussions of ISD.

Chapter 6

Risks and efficiency gains of a tiered structure in large-value payments: a simulation approach

Ana Lasaosa – Merxe Tudela

6 Risks and efficiency gains of a tiered structure in large-value payments: a simulation approach

Abstract

The large-value payment system in the United Kingdom (CHAPS) is highly tiered: a few settlement banks make payments on behalf of many customer banks. This paper makes use of a simulation approach to quantify by how much tiering affects, on the one hand, node and credit risk and, on the other, the liquidity needs of CHAPS. We do so by creating scenarios where current settlement banks become customer banks and thus we increase the degree of tiering. The results show that node risk would rise substantially in what is already a highly concentrated system. As for credit risk, the size of intraday exposures compared with settlement banks' capital is very small and therefore the likelihood of contagion remote. More importantly, the increase in credit risk brought to the system by settlement banks leaving CHAPS bears little relationship to the values settled by each individual bank. We find that increasing the degree of tiering in CHAPS leads to substantial liquidity savings – although the liquidity saved is only a fraction of the spare liquidity currently posted in the system. Most of the savings are due to liquidity pooling rather than to internalisation of payments. There is a strong relationship between changes in values settled and liquidity needs. This relationship can be used to forecast the impact on liquidity needs if more banks were to join CHAPS. The quantification of the trade-off between risk and efficiency in different scenarios provides policy makers with a useful analytical framework for analysing the effects of tiering.

Summary

Only a few banks are direct members of the UK large-value payment system. The vast majority of banks access the system indirectly, through any of the few direct members. We describe a system in which a very small proportion of banks are direct members as a highly tiered system. The degree of tiering affects both how risky and how efficient the UK system is. Recent research has classified the various risks and benefits of tiering in large-value payments, but much less progress has been made in quantifying these risks and benefits. This paper seeks to fill this gap.

In order to gauge how the degree of tiering in CHAPS affects risks and benefits, we need to be able to vary it while holding other factors constant. A simulation approach allows us to do so. We create artificial versions of CHAPS where we increase the degree of tiering by reducing the number of direct members even further. We then use the simulation results to quantify the impact of tiering on concentration and credit risk and on how much liquidity the system needs to operate.

The results show that, in a more tiered system, concentration risk would rise substantially. As for credit risk, our figures have confirmed previous analysis at the Bank that, under normal circumstances, the likelihood of direct members failing due to the credit they extend to their customer banks for their payments business is very small and therefore the likelihood of contagion remote. But, more importantly, our analysis has shown that the increase in credit risk brought to the system by settlement banks leaving CHAPS bears little relationship to the values settled by each individual bank. The key determining factor of the size of intraday credit exposures is the pattern of intraday flows of second-tier banks – a variable that central banks do not observe.

Increasing the degree of tiering in CHAPS leads to substantial liquidity savings. The vast majority of the savings are due to liquidity pooling rather than to internalisation of payments. Moreover, the clear relationship between changes in values settled and liquidity needs shown by our simulations make it possible to project what would happen if current customer banks joined CHAPS as direct members. We estimate that the liquidity needs could increase by GBP 8 billion in aggregate if as many as five large banks (in terms of values of payments processed) joined CHAPS. While this figure is significant, it is only a fraction of the GBP 17 billion spare liquidity posted on average in the system as a whole during the same time period. On an individual bank basis however, given the different cost liquidity has for different banks, this might not be true. Some banks would not be ready to trade-off credit risk for liquidity costs.

6.1 Introduction

Only a few banks are direct members of the UK large-value payment system. The vast majority of banks access the system indirectly, through any of the few direct members. We describe a system in which a very small proportion of banks are direct members as a highly tiered system. Tiering is therefore defined as the proportion of banks that are direct members of the payment system. In this context, higher tiering means fewer direct members. The degree of tiering affects both how risky and how efficient the UK system is. Whilst recent theoretical research and central bank analysis[1] has classified the various risks and benefits of tiering in large-value payments, much less progress has been made in quantifying these risks and benefits. This paper seeks to fill this gap.

CHAPS[2] is the large-value payment system in the United Kingdom. It is a real-time gross settlement (RTGS) system employed since April 1996, processing an average of GBP 200 billion and over 100,000 payments each day. Member banks hold settlement accounts with the Bank of England and they can run fully-collateralised intraday overdrafts. Only 14 commercial settlement banks and the Bank of England are direct CHAPS members.[3] We refer to these banks are direct members, first-tier banks or settlement banks. All other banks (the second-tier banks) have to access CHAPS through these direct or first-tier members. They are called indirect members, second-tier banks or customer banks. A similar degree of tiering is found in the embedded payment system of CREST, the UK securities settlement system.

In order to gauge how the degree of tiering in CHAPS Sterling affects measures of risks, we need to be able to vary it while holding other factors constant. This is not possible with cross-country comparisons, where many factors other than the degree of tiering vary. A simulation approach, by contrast, is particularly useful in creating artificial versions of one payment system with different degrees of tiering. This paper makes use of a simulation approach to quantify by how much tiering affects, on the one hand, node and credit risk and, on the other, the liquidity needs of the UK large-value payment

---

[1] See Kahn and Roberds (2005), Jackson and Manning (2006), Harrison, Lasaosa and Tudela (2005) and Bank of England Payment Systems Oversight Report (2005) (http://www.bankofengland.co.uk/publications/psor/psor2005).

[2] There are in fact two different large-value payment systems, CHAPS Sterling and CHAPS Euro. Since the values settled in CHAPS Sterling are twenty times higher than those settled in CHAPS Euro, our analysis will be restricted to CHAPS Sterling. All references to CHAPS in the paper refer to CHAPS Sterling.

[3] CLS Bank and Abbey joined at the end 2005. We count RBS and Natwest as a single member even though they actually have two separate accounts. A complete membership list can be found in
http://www.apacs.org.uk/uk_payment_schemes/chaps_clearing_1.html.

system. Quantifying the trade-off between risks and efficiency enables us to provide policy-makers with a framework in which to discuss the effects of alternative models. Liquidity cost/risk and efficiency benefits of not in and of themselves make a tiered system optional from a welfare perspective. The paper then goes on to analyse how much of the liquidity gains are due to pooling and how much to internalisation of payments.

The rest of the paper is structured as follows. The next section outlines how tiering may affect risks and liquidity efficiency in the UK large-value payment system. Section 6.3 introduces the methodology used in our analysis. Section 6.4 goes on to describe the main results. Finally, our concluding comments are presented in Section 6.5.

6.2 Tiering, risks and efficiency

In their 2003 *Financial System Stability Assessment of the United Kingdom*,[4] the International Monetary Fund (IMF) highlighted the potential risks arising from the highly-tiered structure of the UK large-value payment systems. Table 6.1 shows how the degree of tiering in CHAPS – defined as the proportion of banks accessing the system directly – is high compared with other countries.[5] When looking at Table A we should bear in mind that systems with a large proportion of eligible banks as direct members can in reality be more tiered if some of those direct members still choose to send their payments through other settlement banks.[6] We know that in the United Kingdom more than half of the value of payments sent to CHAPS by settlement banks are on behalf of second-tier banks. Unfortunately, we do not have comparable data at our disposal for the rest of the countries in our table.

The IMF drew attention to the exposures arising between the first-tier and second-tier institutions and the potential for contagion risk, that is, the risk that credit problems in a second-tier bank might spill over to first-tier banks. Since the IMF assessment was published, the

---

[4] Available at http://www.imf.org/external/pubs/ft/scr/2003/cr0346.pdf.

[5] Tiering defined in this way is obviously correlated with the general degree of concentration in the banking system: if only a few banks account for a large share of the banking market (the UK case), access to payment systems is likely to be tiered due to economies of scale.

[6] This appears to be the case in Fedwire.

Bank has analysed the risks arising from tiering in UK payment systems.[7]

Table 6.1 **Settlement banks in large-value payment systems**

Country	System name	No. of settlement banks[a]	No. of credit institutions	Share of settlement banks
United Kingdom	CHAPS Sterling	13	420	3%
	CHAPS Euro	19		5%
Belgium	ELLIPS	16	109	15%
Canada	LVTS	14	45	31%
France	TBF	156	1,067	15%
	PNS	21		2%
Germany	RTGS Plus	93	2,370	4%
Italy	BIREL	204	821	25%
Japan	BOJ-NET	371	506	73%
Netherlands	TOP	106	95	112%
Sweden	E-RIX	13	125	10%
	K-RIX	19		15%
Switzerland	SIC	307	327	94%
United States	Fedwire	7,736	8,130	95%
	CHIPS		Not available	
European Union	TARGET	1,579	Not applicable	
	Euro1			

Source: Committee on Payment and Settlement Systems (CPSS) Statistics on payment and settlement systems in selected countries (2005) and OECD's Bank Profitability, 2003.
[a] Includes central banks. Data for 2003. As for 2006, CHAPS Sterling has 15 members, including the Bank of England.

A tiered payment system can give rise to several types of risk relative to a system in which all banks are direct participants:

Credit risk: Credit exposures arise when settlement banks offer their customer banks overdraft facilities when making outward payments on their behalf. Conversely, when customer banks hold positive intraday balances at their settlement bank, the customer bank is exposed to the settlement bank. As Flannery (1996) points out, a tiered system that relies heavily on private credit may not function so well in times of crises.

---

[7] 'Strengthening financial infrastructure' article of the December 2004 Financial Stability Review, available at http://www.bankofengland.co.uk/publications/fsr/2004/fsr17art4; 2004 and 2005 payment system oversight reports available at
http://www.bankofengland.co.uk/publications/psor/psor2004 and
http://www.bankofengland.co.uk/publications/psor/psor2005.

Node risk (also referred to as 'concentration risk'): Tiering increases the concentration of payments in each settlement bank. The system is therefore more sensitive to temporary outages experienced by individual settlement banks. The consequences of a temporary disruption can be either merely operational (other banks' inability to make and receive payments) or also liquidity-related. A stricken bank able to receive but unable to send payments may become a 'liquidity sink', draining the rest of the system of the liquidity needed to continue making payments. The larger the proportion of the liquidity in the system controlled at any one time by an individual bank, the greater the risk that operational problems lead to liquidity difficulties.

Legal risk: The finality of payments made or received on behalf of indirect participants is not as well-defined as in systems designated under the European Union's Settlement Finality Directive such as CHAPS. In addition, a tiered system makes internalisation of payments possible. Internalised payments are payments made between customer banks of the same settlement bank and settled internally across the settlement bank's books without being forwarded to the payment system. For a given degree of tiering, a greater degree of internalisation may increase the legal risk of payments being unwound.

Liquidity dependency: Second-tier banks may view their settlement banks as lenders of last resort. First-tier banks, conversely, may depend on the incoming transactions of their customer banks for their own liquidity needs.

So far, the Bank's analysis of the risks involved in tiering has focused on credit risk to settlement banks. Based on evidence that first-tier banks extend unsecured intraday credit to second-tier banks, Harrison, Lasaosa and Tudela (2005) analysed the credit risk exposure of settlement banks using a standard credit risk model. They examined the change in the distribution of credit losses incurred by a bank that moves from only processing payments on behalf of its own customers to carrying out correspondent business on behalf of second-tier banks. The model was calibrated to UK financial infrastructures. It concluded that, in normal market conditions, the credit risk to first-tier banks appeared to be low. Even under stressed circumstances, the assumptions had to be extreme to lead to a significant increase in the credit risk faced by the settlement bank.

As for node risk, research carried out by the Bank has found that operational problems at individual CHAPS settlement members would not in general prevent the remaining banks from making payments to each other due to liquidity shortages – providing that banks quickly stopped making payments to the stricken bank.[8] Bedford, Millard and Yang (2004) use a simulation approach and find that the system exhibits a high level of resilience. Nonetheless, the operational failure of a key node bank would still disrupt all of its own customers and those to whom they were making payments.

Jackson and Manning (2006) construct a model that examines the key factors affecting banks' decisions whether to become direct members of a particular system or not, and the central bank's decision to require collateralisation of intraday credit or not. Their findings suggest the existence of economies of scale in correspondent banking, which make concentration likely.

A tiered payment system has benefits too. It can reduce systemic risk[9] in two different ways. First, a tiered payment system depends less on the central infrastructure because some payments are 'internalised'. In the case of an operational failure of the central system, payments across the books of settlement banks can still take place. A second, potential, risk-related benefit from tiering is the increase in monitoring by first-tier banks of the financial position of second-tier banks. Kahn and Roberds (2005)[10] argue that tiering increases the level of monitoring by first-tier banks and reduces the incentive to default by second-tier banks. If a second-tier bank proves itself to be unreliable, it will be required to collateralise fully its payment activity at an additional cost. If it is reliable, it only needs to be monitored. Hence, the first-tier bank has an incentive to monitor efficiently and the second-tier bank has an incentive to behave reliably.

So far we have compared the systemic risks involved in a tiered and non-tiered payment system. But a tiered system can also be more efficient than one where all banks are direct members of the payment system. In a competitive market, the banks that have a competitive advantage when offering correspondent banking services become settlement banks. Larger banks are normally better placed to do so due to existing economies of scale in several areas:

---

[8] Whether banks do so in practice is an interesting subject for further research. If they do not, the results of Bedford et al (2004) may no longer hold.

[9] We define systemic risk as a risk to the financial system that is not adequately internalised by system participants and that imposes material costs to the banking system should it materialise.

[10] Available at http://www.bankoengland.co.uk/financialstability/futureofpayments/kahnroberdsBOE.pdf

- Infrastructure: IT, contingency arrangements, administration.
- Fee structure: this tends to fall more heavily on banks with small volumes of payments, due to the existence of both fixed fees and a sliding scale for per-transaction (volume-based) fees.
- Dedicated staff: a well-staffed liquidity management team.
- Liquidity requirements of RTGS systems: RTGS systems, such as CHAPS, are liquidity-intensive because they require payments to be made on a gross basis and to be fully pre-funded. Liquidity needs might decrease with the degree of tiering due to two effects, liquidity pooling and internalisation of payments – Section 6.4.5 explores which of the two effects is more important in CHAPS.

This paper focuses on the risks to and efficiency gains from tiering that we can quantify with the data at our disposal: node risk, credit risk, and liquidity savings.

6.3 Methodology

We do not have access to actual data of CHAPS under different degrees of tiering that would allow us to study the implications of a more, or less, tiered structure. We can turn, though, to simulation techniques to analyse different tiering scenarios and their implications for liquidity needs and system risks. Specifically, we use the Bank of Finland payment and settlement system simulator (BoF-PSS2) initially developed in the mid-1990s by the Bank of Finland to study the effects of the introduction of European Monetary Union (EMU) on the Finnish payment systems. Since then the simulator has been developed for and used by several central banks.[11]

The BoF-PSS2 models settlement processes according to a set of rules defined for a payment system environment, giving as outputs account balances and payments settled and received that can easily be analysed within the simulator or exported to other programs. These outputs allow us to draw conclusions about how system characteristics such as credit risk, liquidity consumption, settlement speed or gridlock resolution vary under the different scenarios.

The first step in a simulation process is to establish a ‘benchmark’ against which other simulations (or scenarios) are compared. Our benchmark involves a simple replication of real-life CHAPS Sterling using actual transactions for June 2005. This means replicating a total

---

[11] See Leinonen (2005) for a more technical and precise description of the BoF-PSS2.

of 2.5 million payments accounting for GBP 4.2 trillion of total value transferred. We then construct different scenarios in which we vary the level of tiering in CHAPS Sterling.

For each transaction we have information on the settlement banks sending and receiving the payment, the payment amount and the exact time at which it was sent. We do not have information on end payer or payee, or an indication of whether the payment is sent on behalf of the settlement bank itself or one of its customer banks. And, obviously, internalised transactions are not included in the data since they do not go through CHAPS.[12]

Our simulations involve an *increase* in the degree of tiering and, therefore, a reduction in the number of direct CHAPS members. Decreasing rather than increasing the number of direct members allows us to observe the payments sent and received by the banks that we will turn into customer banks in the data. We can then assign them to a settlement bank once they become customer banks and conduct the simulations. Since our data does not identify the transactions in which customer banks are involved, we could not construct a scenario where customer banks become direct CHAPS members.

If we assume that the effects of increasing and decreasing the number of CHAPS direct members are similar, it is easy to extend the conclusions of this experiment to a situation where tiering decreased. We use the estimated relationship between changes in values settled and the changes in liquidity needs to predict what would happen if tiering decreased.

It is hard to predict in which direction the degree of tiering in CHAPS will move in the near future. Market intelligence suggests that involvement in new systems requires large expenditures on IT, implying that economies of scale will become even more important. If that turns out to be the case, there will be a movement towards specialisation in the provision of customer banking services and CHAPS will become more tiered. On the other hand, the new way in which the Bank is now carrying out is open market operations (introduced in May 2006) gives CHAPS settlement banks the opportunity to use reserves as an additional source of intraday liquidity in addition to collateralised intraday overdrafts. This essentially makes it cheaper for banks to become direct CHAPS members.

We increase the degree of tiering in the system by turning the seven smallest banks by value into customer banks of the three major

---

[12] See Section 6.4.3 below for a detailed explanation of internalisation.

settlement banks in turn. All together the seven smallest banks account for 17% of CHAPS transactions by value and 19% by volume. We start off by assigning (or converting into customer) the smallest bank (accounting for less than 1% of CHAPS values) to one of the three major settlement banks (let us call it major settlement bank 1); we then assign this same bank to another of the three major settlement banks (major settlement bank 2); in another step we assign it to a third major settlement bank (major settlement bank 3). We continue assigning the smallest and second smallest banks to major settlement bank 1, then to major settlement bank 2 and finally to major settlement bank 3. We go on with the three smallest banks in a similar way and continue this process up to the seven smallest banks. This gives us 21 different tiering scenarios.

We construct two additional scenarios in the following way. We take the seven smallest banks as before and we assign some of them to settlement bank 1, some of them to settlement bank 2 and the rest to settlement bank 3. The choice of which of the smallest banks to assign to which of the major settlement banks is based on the relative proportion of bilateral payments between the two types of banks and involves two ways, first considering payments by volume and second by value. New customer banks are likely to choose as their settlement banks those banks with whom they have a large proportion of payments by virtue of their business, since internalised payments are usually charged at a lower rate than those going through CHAPS. We call the first scenario 'assign by volume', and the second one 'assign by value'. We have 23 different scenarios in total.

There are two main assumptions embedded in this experimental design. First, we assume that the timing of payments does not change when a settlement bank becomes a customer bank. Settlement banks have more discretion than customer banks about when to send payments to CHAPS –discretion constrained by intraday deadlines prevalent in financial markets and by throughput guidelines. But, with the information at our disposal, any changes to the timing of payments in our simulations would be completely arbitrary. The second assumption is that settlement banks that become customer banks take their own customer banks with them because there is no immediate reason to believe otherwise. Any other alternative would be arbitrary as well.

6.4 Results

In what follows we analyse how our measures of node risk, credit risk and liquidity efficiency change in the different scenarios.

6.4.1 Node risk

Following James (2003), we gauge the increase in node risk by looking at individual banks' share of all payments going through CHAPS. This measure shows the proportion of payments in the system that would be affected by an operational outage in that particular bank. This measure of node risk does not take into account the interconnectedness of the node. Node risk defined in this way does not encompass the impact of payments going through other settlement banks that could be affected through liquidity sink channels in such an operational crisis. We calculate the share of payments based on the value of all incoming and outgoing payments for each bank. Node risk for bank j is defined as value settled in CHAPS by bank j plus value received through CHAPS by bank j over the totals sent and received in CHAPS by all settlement banks.

$$\text{Node Risk}_j = \frac{\text{Value Sent}_j + \text{Value Received}_j}{\text{CHAPS Totals Sent and Received}}$$

Current concentration figures are 26% for the largest settlement bank, 25% for the second largest and 17% for the third largest.

Table 6.2 shows how concentration in CHAPS increases when the same seven customer banks are assigned to a combination of the three major settlement banks. As described in the previous section, we decide which customer bank to assign to each settlement bank according to the most common interbank payment flows that we observe in our data.

Table 6.2 **Increase in CHAPS Sterling payments share when new customer banks are assigned to several settlement banks**

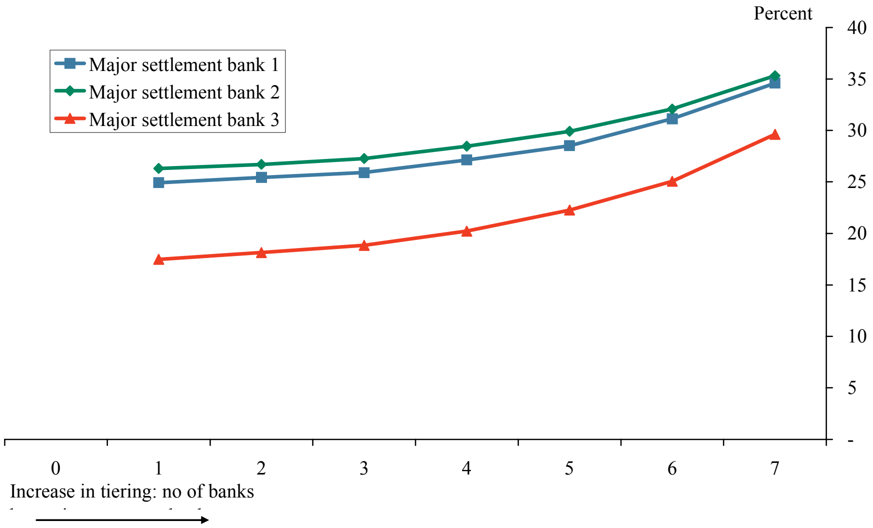
	Major settlement bank 1	Major settlement bank 2	Major settlement bank 3
Current share	26.0	24.5	17.1
Assign by value[(a)]	31.7	23.3	29.3
Assign by volume[(a)]	30.4	23.9	30.4

Source: Payments database and Bank calculations, average over business days, June 2005.

[(a)] For a definition of these terms or scenarios see page 11.

Figure 6.1 shows the increase in the share of payments of each major settlement bank when all the new customer banks are assigned to it. Each line shows the results of a different simulation: when all new customer banks are assigned to major settlement banks number one, two and three. Each dot represents the average over the business days in June 2005. The unfilled shapes correspond to our benchmark case of current node risk figures. The degree of tiering moves up as we move towards the right of the firuge, with each point on the horizontal axis representing one more small settlement bank that becomes a customer bank. Each differently coloured line shows how the node risk increases when the new customer banks are assigned to each of the three major settlement banks. Finally, the two single brown dots show the share of payments for major settlement bank 1 in scenarios 'assign by value' and 'assign by volume'.

Figure 6.1 **Node risk (average June 2005)**



Source: Payments database and Bank calculations.

We can see that the increase in node risk is greatest when the new customer banks are assigned to the major settlement bank that currently has the smallest payment share of the three. This is not surprising. Ceteris paribus, the bigger a settlement bank is, the larger the proportion of its payments that can be internalised. Internalised payments are not sent through CHAPS, so internalisation decreases both the numerator and the denominator in our measure of node risk – but the decrease is stronger in the denominator. As a result, the more internalised payments a bank has, the smaller the impact of making other settlement banks their customer banks on node risk. The increase when either major settlement bank 1 or 2 capture the new customer banks is practically identical.[13] There is little day-to-day variation in concentration risk within our sample. The coefficient of variation is low, ranging between 0.02 and 0.03 in all our scenarios.

---

[13] The change in our node risk measure as settlement banks become customer banks is made up of an increase in the numerator and a decrease in the denominator (total values settled) due to internalised transactions. This change in the denominator makes potential statistical relationships between overall values settled in CHAPS and node risk hard to interpret, despite obvious correlation coefficients. For this reason, we do not figure the relationship between values settled and node risk.

6.4.2 Credit risk

We measure the increase in credit risk in the different scenarios by taking the maximum intraday liquidity that banks currently need as direct CHAPS members to be able to settle all transactions on a gross basis in real time. CHAPS settlement banks have to post collateral to access intraday credit from the Bank of England. We know that it is common practice in the United Kingdom to grant customer banks unsecured intraday overdraft facilities – especially to large customers. Thus, the maximum intraday liquidity currently used by settlement banks is likely to be a reasonable measure of the maximum unsecured intraday credit that each individual bank would need to obtain from their settlement banks if they became indirect CHAPS members. It is worth stressing that this would be an upper bound measure: it is reasonable to assume that settlement banks would have some degree of discretion of when to process their customers' payments and try to avoid building unnecessary credit exposures. Table 6.3 shows the monthly average of maximum intraday liquidity for each settlement bank that becomes a customer bank.

The first point to emerge from this table is that the magnitude of the risk is insignificant with respect to the amounts of Tier 1 capital held by the three large settlement banks: between GBP 19 and GBP 43 billion. This finding is consistent with Harrison et al (2005) who conclude that in normal market conditions the risk to settlement banks from their intraday credit exposures to second-tier banks appears to be low. Interpreting these results, one should bear in mind that these figures are monthly averages of intraday maxima – the exposures on which the averages are calculated may have lasted for only a few seconds each day. And the fact that the mean is in all cases higher than the median implies that the distribution is asymmetric, skewed towards a relatively small number of high exposures. The coefficients of variation displayed in Table 6.3 range between 0.46 and 0.78, reflecting a very disperse distribution of intraday maxima over the month.

Table 6.3 **Maximum intraday credit that would have to be extended to each individual bank**

Customer bank number	Maximum intraday credit (GBP million)			Share of CHAPS payments (per cent)		Payments made on behalf of customers (%)		(a)/ (mean value settled)
	Mean (std. dev.) (a)	Coeff. variation	Median	By value	By volume	By value	By volume	
1	253 (135)	0.53	248	0.5%	1.2%	9.7	31.7	0.23
2	385 (177)	0.46	367	0.8%	0.9%	9.0	9.0	0.23
3	770 (421)	0.55	615	0.8%	2.2%	0.0	0.3	0.45
4	718 (557)	0.78	653	1.9%	7.5%	2.2	1.5	0.18
5	438 (320)	0.73	341	2.6%	0.7%	0.0	0.0	0.08
6	1,042 (673)	0.65	860	3.9%	1.8%	12.0	41.3	0.13
7	735 (486)	0.66	587	6.6%	5.1%	59.9	54.3	0.05

Source: Payments database (June 2005), 2005 Correspondent Banking Survey and Bank calculations.

It is also apparent from Table 6.3 that the size of the exposures is not proportional to the share of total CHAPS payment values made up by each bank.[14] Bank 3, with 0.8% of CHAPS payments, has similar mean peak liquidity usage as bank 7, with 6.6% of CHAPS payments. The correlation coefficient between the two series is only 0.52, with the ratio varying from 0.05 for bank 7 to 0.45 for bank 3. Economies of scale in liquidity usage (described in the next section) could explain why credit risk increases less than proportionally with the value of CHAPS payments. But they do not explain why two banks with a similar share of CHAPS values have different maximum liquidity needs. The relative size of each bank's correspondent business (column 5 of Table 6.3) does not emerge as a factor, either. Banks 3 and 5 have virtually no correspondent business and very different peak liquidity usage. The most plausible explanation is that these observed variations in liquidity usage may stem from differences in their customer base affecting the timing of incoming and outgoing payments – if a settlement bank needs to make a large number of payments before receiving many payments it has higher peak intraday funding needs.

The implication of the above analysis is that, unlike in the case of node risk, policy-makers interested in reducing credit risk by encouraging more second-tier banks to join the payment systems cannot rely on the value of sterling payments processed as a reliable

---

[14] The correlation coefficient between the median maximum liquidity needed and the share of CHAPS payment values is 0.5.

measure of potential intraday credit exposure. The key factor determining the size of intraday exposures is the intraday pattern of payments. Unfortunately, precise data on the intraday pattern of payments of customer banks is not usually available to policy-makers, making their task hard.

6.4.3 Efficiency gains

Compared with Deferred Net Settlement (DNS) systems, RTGS systems reduce credit risk at the expense of increasing liquidity costs.[15] Not being able to net off payments increases the liquidity needed by each settlement bank and, therefore, the overall liquidity needs of the system. In the case of CHAPS, banks obtain liquidity in central bank money by posting eligible collateral with the Bank of England.

As discussed above, tiered payment systems are more concentrated. Concentration in payment systems leads to liquidity savings due to two effects that go in the same direction: liquidity pooling and internalisation of payments.

- Liquidity pooling: despite the fact that payments cannot be netted off in an RTGS system, the larger the number of payments received and sent by a given settlement bank, the higher the probability that incoming payments fund (totally or partially) outgoing payments.[16] This results, on average, in smaller peaks of both liquidity needs (when the balance is negative) and liquidity surpluses (when the balance is positive). Since we are concerned with the average of *maximum* intraday liquidity needed, we expect our scenarios involving fewer but larger settlement banks to show a fall in liquidity needed with respect to current CHAPS figures. This saving involves no change in the payment values settled through CHAPS.
- Internalisation: when a settlement bank becomes a customer bank of another settlement bank, all transactions between the two (either on their own behalf or on that of their customers) that used to go through CHAPS are internalised. Since they are no longer sent to

---

[15] Selgin (2004) argues that the drive to replace DNS with RTGS systems on credit risk grounds in misguided. He claims that the only credit risks arising in DNS systems are those granted by the receiving banks to their customers. He also argues that regulators' guarantees distort the market.

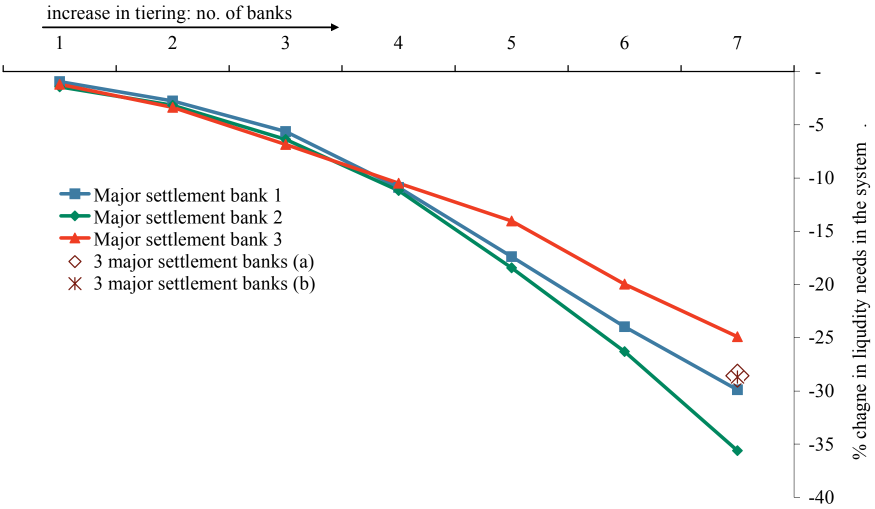
[16] Except in very asymmetrical distributions for incoming and outgoing payments.

the payment system, the settlement bank does not need to obtain intraday credit from the Bank of England to fund them. The internalisation effect leads to a decrease in liquidity needed as a result of the decrease in values settled through CHAPS.

An estimation of the relative size of each effect is presented in Section 6.4.5.

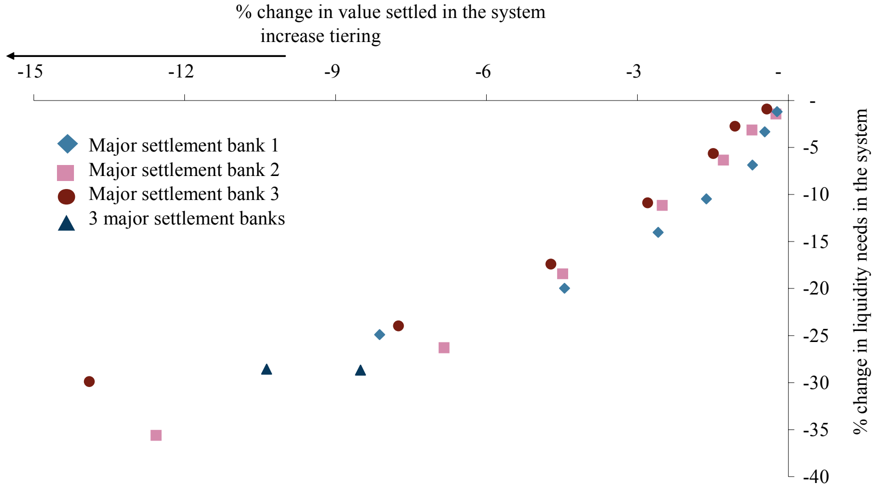
Figure 6.2 shows the savings in liquidity needs as tiering increases; similarly Figure 6.3 shows the savings in liquidity needs relative to the decrease in overall values settled in CHAPS for each of our simulations. The figures are percentage change relative to current values. Different colours represent different scenarios. The three scenarios where all new customer banks are assigned to one large settlement bank have seven points each, with two extra data points (brown in the figure) for the simulations where all seven customer banks are assigned to a combination of settlement banks. The data points for the 'assign by value' and 'assign by volume' scenarios in Figure 6.2 overlap and therefore only one is visible. The liquidity needs of the system are defined as the sum of each individual settlement bank's intraday maximum liquidity requirements. As tiering increases and the number of CHAPS settlement banks falls, the values settled in the system go down.

Figure 6.2 **Increase in tiering and liquidity needs**



Source: Payments database and Bank calculations.

Figure 6.3 **Percentage changes in liquidity needs in CHAPS and values settled in CHAPS**



Source: Payments database and Bank calculations.

The figure shows that there are substantial liquidity savings associated with a reduction in the number of settlement banks (an increase in tiering). Moreover, the savings are similar across the different scenarios: dots of different colours in the figure are relatively close together. The maximum saving is 36% of current liquidity needs when all new customer banks are assigned to settlement bank 2. This is equivalent to GBP 5.9 billion liquidity. It is apparent from figure that there is a close relationship between changes in overall values settled in CHAPS and changes in liquidity needs. For a given percentage reduction in values settled, the reduction in liquidity needs is approximately three times as big.

The variation in liquidity savings across the month is higher than in the case of node risk but much lower than for credit risk. The coefficients of variation in all 23 scenarios range between 0.12 and 0.15 – compared to 0.02–0.03 for node risk and 0.46–0.78 for credit risk.

6.4.4 Decrease in tiering: liquidity costs

The close relationship between values settled and liquidity needs found across all scenarios allows us to attempt a forecasting exercise. Our interest is to try and gauge how much liquidity would CHAPS need if some large (in terms of values of payments processed) customer banks became settlement banks. This boils down to fitting a line to the points shown in 3 and projecting it to the positive quadrant of the horizontal axis. This will give the increase in liquidity needs expected when the degree of tiering in the system decreases and the values settled in CHAPS increase.

Since no data points from our simulations lie in the positive quadrant of the horizontal axis we need to make strong assumptions about the functional form of the relationship between positive increases in values settled and positive increases in liquidity needs. The only information we have at our disposal is the relationship found in the negative quadrant. We try three regression specifications: linear, quadratic and cubic. The criteria for choosing a particular forecasting specification is twofold: goodness of fit to the observed data points and plausibility of the predicted values in the positive quadrant.

The three regressions fitted the data well, with adjusted R^2 of 0.95 for the linear equation, 0.99 for the quadratic and 0.99 for the cubic. But a quadratic form extended into the positive quadrant resulted in implausibly high values of liquidity needs: 100% extra for an increase in values settled of 7%. Our chosen functional forms for forecasting purposes, therefore, are a linear equation and a cubic equation. The cubic equation is used to create a concave mirror image of the convex line on the negative quadrant on the positive quadrant. We believe this is a plausible functional specification: extra liquidity needs become smaller as new banks join CHAPS, just as extra savings become smaller as more banks leave the system.

The points chosen for the forecast are not arbitrary. They correspond to the values that would settle in CHAPS if the largest, second largest and up to five largest customer banks joined CHAPS. We know the values from the 2003 CHAPS Traffic Survey, a dataset that includes the values of transactions settled by current settlement banks on behalf of their (anonymised) customer banks.[17] Transactions originated by the largest customer bank makes up 4.2% of current CHAPS values, and those originated by the five largest customer banks make up 15.7%.

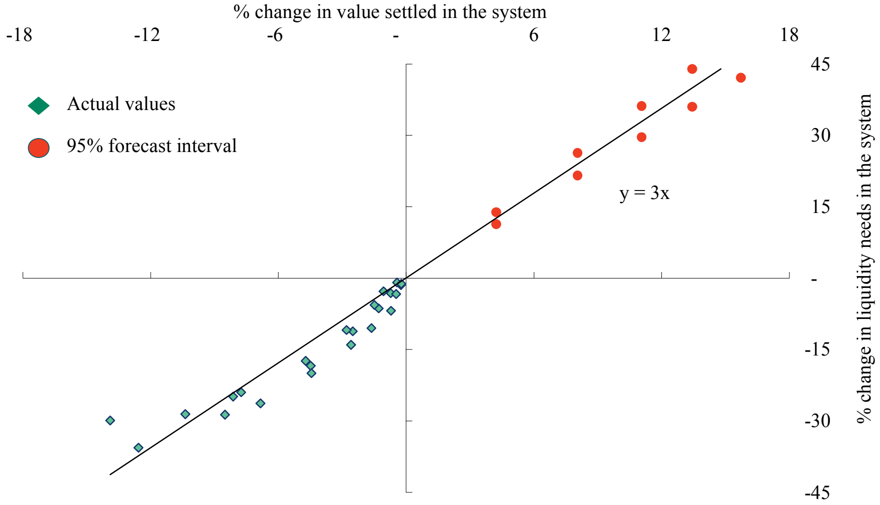
Figures 6.4 and 6.5 depict the linear and cubic equations respectively. In the linear specification, the 95% confidence interval

---

[17] The values used are likely to be downward bias. It is plausible to assume that customer banks have to make payments to their settlement banks and vice versa – these payments will, by definition, be internalised. This internalisation would disappear once the customer banks became settlement banks, so the overall values settled through CHAPS are likely to be higher.

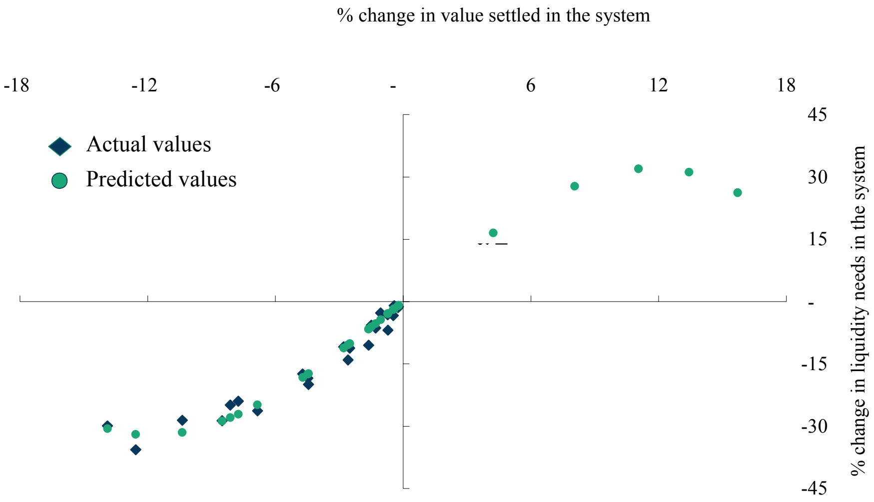
for the coefficient is between 2.7 and 3.3 (with no constant). That implies that for any given percentage increase in values settled, the percentage of extra liquidity needed is three times as high. This assumption becomes less plausible as we predict larger increases – one would expect to find some deceleration in the rate of liquidity needed as tiering tends towards zero. The cubic specification accounts for this deceleration. The linear coefficient lies in a 95% confidence interval between 3.74 and 4.43, and the cubic one between -0.012 and -0.007. In fact, the deceleration captured by the latter negative coefficient is so strong that there is s a slight percentage decrease in liquidity needs for the last two points in our forecast. Figure 6.6 depicts the 95% confidence intervals around the forecast, while Table 6.4 gives the values of extra liquidity needs that an increase in the number of settlement banks would bring about.

Figure 6.4 **Predicted changes in liquidity needs based on changes in value settled – linear prediction**



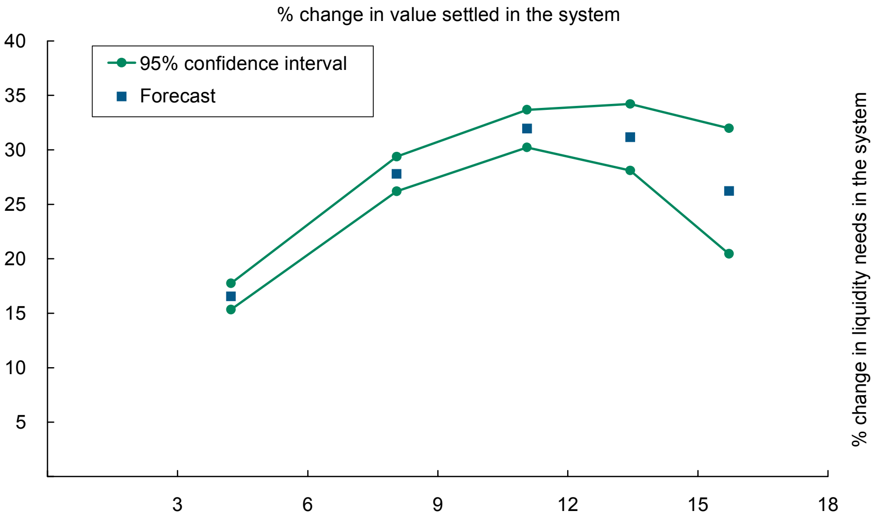
Source: Payments database and Bank calculations.

Figure 6.5 **Predicted changes in liquidity needs based on changes in value settled – cubic prediction**



Source: Payments database and Bank calculations.

Figure 6.6 **Forecast changes in liquidity needs and confidence intervals - cubic prediction**



Source: Payments database and Bank calculations.

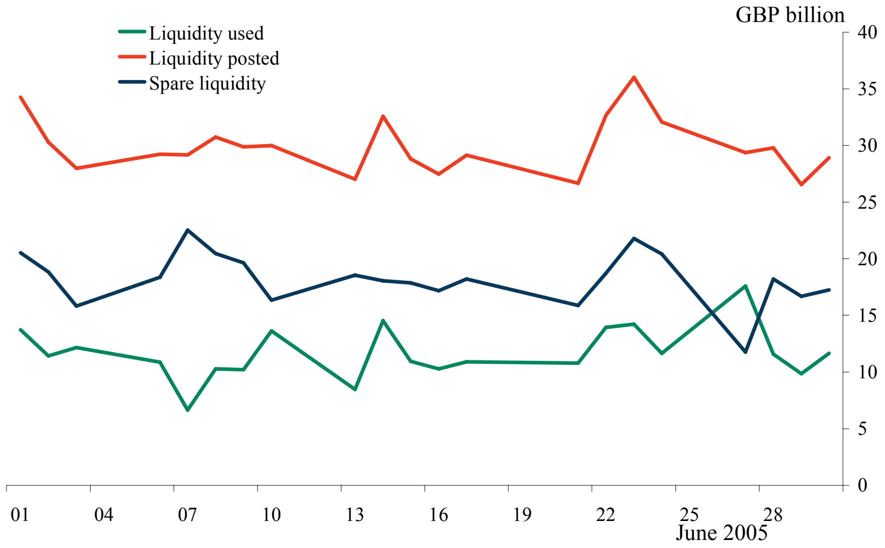
Table 6.4 **Changes in value settled and changes in liquidity needs**

	Change in values settled, per cent (in GBP billion)	Change in liquidity needs, per cent (in GBP billion)	
No. of customer banks joining CHAPS		Linear	Cubic
1	4.2 (GBP 8.2 bn)	12.6 (GBP 2.1 bn)	16.5 (GBP 2.7 bn)
2	8.0 (GBP 5.6 bn)	23.9 (GBP 4.0 bn)	27.8 (GBP 4.6 bn)
3	11.1 (GBP 21.4 bn)	32.9 (GBP 5.4 bn)	32.0 (GBP 5.3 bn)
4	13.4 (GBP 26.1 bn)	40.0 (GBP 6.6 bn)	31.2 (GBP 5.2 bn)
5	15.7 (GBP 30.5 bn)	46.7 (GBP 7.7 bn)	26.2 (GBP 4.3 bn)

Source: Payments database and Bank calculations.

How valuable these liquidity savings are depends on how scarce liquidity is. Data show that there is a lot of spare liquidity (defined as the difference between the maximum liquidity posted intraday and the maximum liquidity used intraday) in the system (see Figure 6.7). Banks posting more liquidity than needed is an indication that liquidity in the system is cheap. This may well be the case, especially for UK-owned banks subject to the Sterling Stock Liquidity Regime (SLR) by the regulator. Under the SLR, banks must hold a stock of eligible liquid assets overnight. The list of eligible assets broadly coincides with assets that can be used as collateral to obtain intraday credit with the Bank of England. If there is practically no opportunity cost in using the eligible assets intraday, then banks may decide to post more liquidity than needed. Foreign banks operating in the United Kingdom, on the other hand, are subject to a maturity mismatch approach, which simply requires that they have incoming liquidity to fund known outflows. Intraday liquidity may be more expensive for them.

Figure 6.7 **CHAPS spare liquidity**



Source: Payments database and Bank calculations.

In June 2005, the average spare liquidity was GBP 17 billion (GBP 18 billion median). The standard deviation over the month was GBP 5 billion. The differences across banks are marked. The settlement bank with most spare liquidity had an average of GBP 5 billion, that with the least just GBP 10 million. Given these figures, the increase in liquidity needs of up to GBP 8 billion suggested by our forecasts do not appear, at first glance, a disproportionate price to pay for the potential reductions in node and credit risk that a decrease in tiering in CHAPS would bring at the system level. On an individual bank basis, and given the different price liquidity has for different banks, this might not be true. Especially if the joining banks have higher liquidity costs that the current members, this would certainly be the case for non-UK banks, not subject to SLR requirements. Those banks would not be ready to trade off credit risk for liquidity costs and some policy action (such as extensions of eligible collateral, use of cross-border collateral, liquidity saving functionalities) might be necessary.

6.4.5 Liquidity pooling vs internalisation

The liquidity gains observed in a more tiered system can stem from either an increase in the pooling of liquidity of from internalisation of payments. To our knowledge, no paper in the literature has estimated which of these two effects drives the savings.

We cannot disentangle the proportion of liquidity gains due to pooling and to internalisation with absolute precision. The reason for this ambiguity is that the amount saved at each transaction will depend crucially on the liquidity position of the settlement bank at the precise point in time when the payment takes place. But we can calculate the upper and lower bound for the proportion of savings due to each effect. This information will give a clear indication of each factor's relative size. The appendix gives a detailed explanation of how the intervals are calculated and why they are the closest we can get to quantifying the relative savings due to pooling and to internalisation.

Our results show that the vast majority of liquidity savings are caused by pooling. Table 6.5 displays the estimated intervals for the proportion of liquidity savings in CHAPS that are due to each factor. The average is calculated as the daily mean over our sample period and across all scenarios. Pooling is, on average, eight times bigger than internalisation. Table 6.5 also presents the upper and lower bound for the minimum (maximum) daily savings for the scenario with the minimum (maximum) savings. The lower bound of the minimum savings due to pooling (63%) is clearly higher than the upper bound of the maximum savings due to internalisation (37%). We can therefore be certain that liquidity pooling accounts for most of the savings observed.

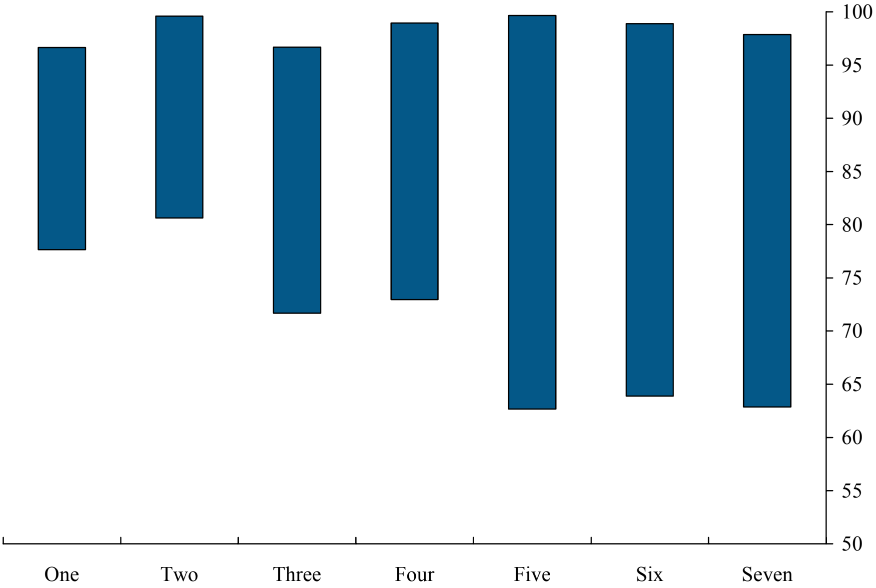
Table 6.5 **Estimated range of savings due to internalisation and to liquidity pooling**

	Internalisation	Liquidity pooling
Average	1%–22%	78%–99%
Minimum	0%–2%	63%–96%
Maximum	4%–37%	98%–100%

Source: Payments database and Bank calculations.

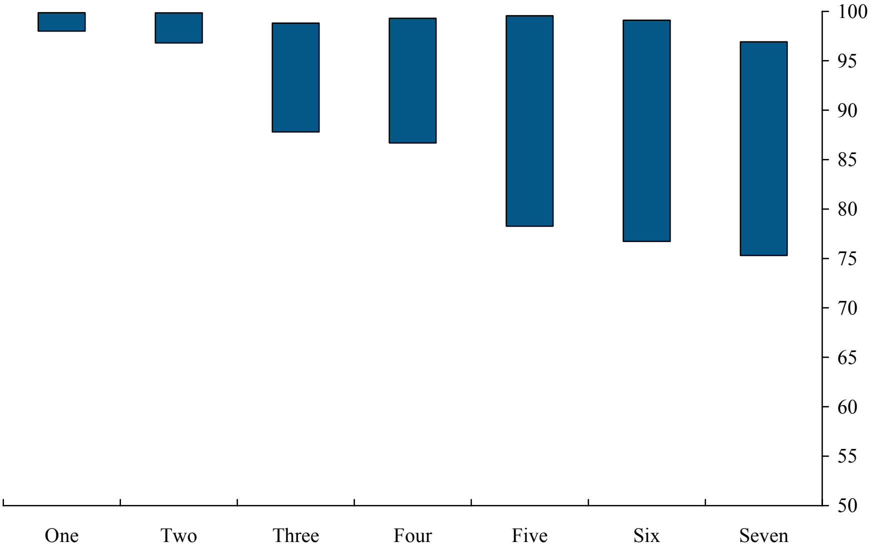
Finally, Figures 6.8 and 6.9 show that the proportion of savings due to internalisation increases with the number of new customer banks assigned to each major settlement bank.

Figure 6.8 **Liquidity savings due to pooling (%) when banks are assigned to major bank 1**



Source: Payments database and Bank calculations.

Figure 6.9 **Liquidity savings due to pooling (%) when banks are assigned to major bank 2**



Source: Payments database and Bank calculations.

6.5 Conclusions

This paper uses a simulation approach to quantify the impact of a change in the degree of tiering in the structure of the United Kingdom large-value payment system on node risk, credit risk and liquidity efficiency. It does so by creating artificial scenarios where the number of direct participants in CHAPS Sterling is reduced one by one, thus increasing the degree of tiering. The new customer banks are assigned to either one or a combination of the large (by value of payments processed) settlement banks.

The results show that node risk would rise substantially in what is already a highly concentrated system. The increase in risk is slightly smaller than the share of payments accounted for by the settlement banks becoming customer banks because of internalisation effects. As for credit risk, our figures have confirmed previous analysis at the Bank that, under normal circumstances, the size of intraday exposures compared with the settlement bank's capital is very small and therefore the likelihood of contagion remote. But, more importantly, our analysis has shown that the increase in credit risk brought to the system by settlement banks leaving CHAPS bears little relationship to the values settled by each individual bank. The key determining factor of the size of intraday credit exposures is the pattern of intraday flows of second-tier banks – a variable that central banks do not observe.

Increasing the degree of tiering in CHAPS leads to substantial liquidity savings. Our analysis has shown that the vast majority of the savings are due to liquidity pooling rather than internalisation. Moreover, the clear relationship between changes in values settled and liquidity needs shown by our simulations make it possible to project what would happen if current customer banks joined CHAPS as settlement banks. We estimate that liquidity needs could increase by GBP 8 billion in aggregate if as many as five large banks (in terms of values of payments processed) joined CHAPS. While this figure is significant, it is only a fraction of the GBP 17 billion spare liquidity posted on average in the system as a whole during the same time period. On an individual bank basis however, given the different cost liquidity has for different banks, this might not be true. Some banks would not be ready to trade-off credit risk for liquidity costs. Liquidity cost/risk and efficiency benefits of not in and of themselves make a tiered system optional from a welfare perspective.

Two possible extensions of this paper stand out as promising. One involves simulating operational outages of individual banks in the more concentrated scenarios in order to analyse how robust CHAPS would be in terms of liquidity compared with the current situation. This exercise will complement the direct effects of an individual bank's operational outage (share of payments usually settled by that bank that are disrupted) by adding second-order disruptions through potential liquidity shortages in the system. A second possible avenue for further work is to model the relationship between settlement banks and the new customer banks as an ancillary system. Such an exercise would provide us with a more detailed picture of the sources of liquidity needs in corresponding banking.

References

Bedford, P – Millard, S – Yang, J (2004) **Assessing operational risk in CHAPS Sterling: a simulation approach.** Bank of England Financial Stability Review, December, 135–143.

Flannery, M J. (1996) **Financial Crises, Payment System Problems, and Discount Window Lending.** In Controlling Risks on Large-Value Transfer Systems. Journal of Money, Credit and Banking, Vol. 28, No. 4, Part 2: Payment Systems Research and Public Policy Risk, Efficiency, and Innovation, November, 804–824.

Harrison, S – Lasaosa, A – Tudela, M (2005) **Tiering in UK payment systems.** Bank of England Financial Stability Review, June, 63–72.

Jackson, J – Manning, M (2006) **Central bank intra-day collateral policy and implications for tiering in RTGS payment systems.** mimeo.

James, K (2003) **A statistical overview of CHAPS Sterling.** Bank of England Financial Stability Review, June, 115–121.

Kahn, C M – Roberds, M (2005) **Payments settlement: tiering in private and public systems.** University of Illinois and Federal Reserve Bank of Atlanta, mimeo.

Leinonen, H (2005) **Liquidity, risks and speed in payment and settlement systems – a simulation approach.** Bank of Finland Studies E:31, 49–57.

Selgin, G (2004) **Wholesale Payments: Questioning the market-failure hypothesis.** International Review of Law and Economics, 24, 333–350.

Appendix 1

Liquidity pooling vs internalisation

As stated in the main body of the paper, concentration in payment systems leads to liquidity savings because of two effects: internalisation of payments and liquidity pooling. When more payments are internalised, the settlement bank does not need to find the liquidity to fund them and the liquidity needs of the system decrease. And a settlement bank with a larger number of transactions is likelier to have incoming payments funding its outgoing transactions. As a result, the maximum liquidity need intraday is likely to be lower.

In practice, though, it is very difficult to disentangle both effects since the origin of the precise liquidity savings at any point in time during a day depends crucially on the liquidity position of the settlement bank at that point in time. A payment sent, say, from settlement bank A to settlement bank B will be internalised if settlement bank A becomes a customer of settlement bank B. This, in theory, could reduce the liquidity needs of settlement bank B. But it will not do so if bank B was already in liquidity surplus when the saving took place. We need to bear in mind that liquidity needs are defined as the intraday *maximum*, not average. Conversely, a potential saving caused by liquidity pooling may not affect the maximum intraday liquidity needs if the settlement bank was already in surplus. In what follows we run through a stylised example that attempts to disentangle these two effects.

Table A6.1 lists the payment structure in a very simple system with four settlement banks and eight transactions at different times within the day. We call this 'the benchmark model'. We are interested in comparing the liquidity needs in the benchmark model with the liquidity needs in a situation when bank AA becomes a customer bank of bank MM and all payments between them are internalised. We have set up the transactions in such a way that we only need to compare the liquidity needs of bank MM before and after AA becomes its customer bank to see how the liquidity needs of the system changes. AA starts receiving money before having to send any payments AA always has surplus liquidity, thus not affecting the maximum intraday liquidity needs of the system. The comparison is not affected by the liquidity needs of BB and CC, either – they are the same before and after AA becomes MM's customer bank.

Table A6.1 **Payments structure – benchmark model**

Time	Payer	Payee	Amount
1	MM	AA	90
2	MM	CC	10
3	BB	AA	50
4	MM	AA	25
5	CC	AA	100
6	AA	BB	10
7	AA	MM	10
8	MM	BB	100

Table A6.2 calculates the liquidity needs under different calculations. As explained we can just focus of MM‘s liquidity needs for comparison purposes.

Table A6.2 **Liquidity needs**

Benchmark [A]		AA customer of MM, payments between them internalised [B]		As benchmark but excluding payments between AA and MM [C]	
Time	Net debit	Time	Net debit	Time	Net debit
1	90	1	0	1	0
2	100	2	10	2	10
3	100	3	-40	3	10
4	125	4	-40	4	10
5	125	5	-140	5	10
6	125	6	-130	6	10
7	115	7	-130	7	10
8	215	8	-30	8	110
Maximum liquidity needs to be able to settle all transactions in RTGS					
215		10		110	

The first two columns in Table A6.2 refer to the benchmark situation (when AA is a direct CHAPS member). Each row gives, after each transaction, the net debit position of bank MM. For example, after transaction 1, MM bank is in a net debit position of 90 (since it has to pay to AA 90 without having received any payment yet). At time 2, MM has to make a further payment of 10, increasing its net debit position to 100. No modifications occur after transaction 3 since it does not involve bank MM. After transaction 4, MM has to make another payment of 25 further increasing its net debit position to 125. Transactions 5 and 6 do not affect MM's net debit position, whereas transaction 7 decreases it to 115 due to the payment of 10 received from AA. There is a further increase in MM's net debit position of 100 to 215 after transaction 8. The maximum liquidity need is therefore 215.

It is plausible to think that the difference between calculations [A] and [C] would give us the liquidity savings due to internalised payments. The only difference between those two calculations is the payments between the settlement bank and the bank to become customer of the settlement bank; this gives a saving in liquidity of 105 (215–110). Likewise, the difference in liquidity needs between calculations [B] and [C] would show the savings in liquidity due to liquidity pooling. The difference between that pair of calculations corresponds to the transactions between any other settlement bank and the new customer bank. This gives a savings in liquidity of 100 (110–10). The total savings in liquidity are thus 205 (215–10), that is, the sum of liquidity savings due to internalisation and the liquidity savings due to liquidity pooling.

Let us consider now one more calculation [D], as in Table A6.3 as shown below: the liquidity needs of the system when AA becomes a customer bank of MM but the payments between them continue to be sent to RTGS. We could assume now that the difference between calculations [D] and [A] would give us the liquidity savings due to liquidity pooling, since in calculations [D] we do not allow for internalisation. Liquidity savings are now 125 (215–90).

Table A6.3 **– extra column: Liquidity needs**

AA customer of MM, payments between them sent to RTGS [D]	
Time	Net debit
1a	90
1b	0
2	10
3	-40
4a	-15
4b	-40
5	-140
6	-130
7a	-120
7b	-130
8	-30
Maximum liquidity needs to be able to settle all transactions in RTGS	
	90

The difference in liquidity needs between calculations [D] and [B] could be interpreted as savings in liquidity due to internalisation only. This yields a value of 80 (90–10). Adding the savings due to liquidity pooling and internalisation savings gives again total liquidity savings.

As we can see, these two methods of disentangling the liquidity savings yield different results. The reason for the difference is that when using columns [A], [B] and [C], we are assigning all savings whose cause we cannot identify to internalisation. When we use columns [A], [B] and [D], by contrast, we assigned them to liquidity pooling. We therefore conclude then that we cannot disentangle the liquidity savings due to internalisation and to liquidity pooling with absolute precision. We are able, though, to calculate upper and lower bounds for the estimates of the savings due to either internalisation or pooling in each of our scenarios. In the example presented above, the lower bound for internalisation savings is given by D–B (80), and upper bound by A–C (105). The lower bound for savings due to liquidity pooling is given by C–B (100) and the maximum by A–D (125).

Chapter 7

Risk concentration, network structure and contagion in the Austrian Real Time Interbank Settlement System

Stefan W Schmitz – Claus Puhr

7 Risk concentration, network structure and contagion in the Austrian Real Time Interbank Settlement System

Abstract

The objectives of this paper are twofold: First, we present a statistical analysis of liquidity, risk concentration and network structure in the Austrian Real Time Interbank Settlement system (ARTIS). Second, we quantify the contagion effect of an operational incident at one of the participants' sites on the other participants of ARTIS by use of simulations. The main result of the statistical analysis is that in general the value and number of payments received and submitted were quite concentrated among the top three banks and the top transfer account in the ARTIS system during the sample period. With respect to the second objective of the paper, the simulation analysis focuses on operational problems occurring in the realm of one of the participants. The main finding of the simulations was that the contagion effect on the smooth functioning of the payment system was substantial in all three scenarios. In contrast to the literature, we operate with actual rather than simulated liquidity data and study the contagion effect based on the individual bank level in addition to the aggregate level of unsettled payments. Moreover, the paper provides results on features of large value payment systems that have hitherto gone unstudied in the literature: the stop sending rule and debit authorisation.

7.1 Motivation and objectives

The first objective of this paper is to present descriptive statistics on ARTIS and a statistical analysis of liquidity and risk concentration as well as of the network structure in ARTIS. Despite the growing research literature on payment systems, very little statistical analysis of large value payment systems is available internationally.[1] This

---

[1] Notable exceptions are James (2003) and the papers published in Leinonen (2005).

paper helps to fill this gap. Furthermore, data on the concentration of risk is an important input in payment system oversight. As such it forms the basis for the assessment of operational risk in ARTIS – the second objective of this paper. Aggregate liquidity in the system is comparatively high, so operational incidents might not have a severe contagious impact. The purpose of this paper is to take this conjecture to the data and quantify contagion in ARTIS. The simulations employed have to focus on operational risk at institutions with most payment activity and liquidity. The analysis of operational risk is a basic task of the European System of Central Banks (ESCB) as well as of the Oesterreichische Nationalbank (OeNB), namely the promotion of the smooth operation of the payment system (Maastricht Treaty Article 105 (2) and ESCB Statute Article 3 (1)). The OeNB is in charge of payment system oversight in Austria. The mandate includes oversight over ARTIS, which is also the Austrian component of the Trans-European Automated Real-time Gross Settlement Express Transfer System (TARGET). The analysis is based on transactions and collateral data for November 2004.[2]

The second objective of this paper is to quantify the impact of an operational incident at one of the participants in ARTIS on the ability of other participants to settle transactions by use of simulations. In this project we do not simulate an operational failure of the ARTIS platform itself. Operational risk is defined as 'the risk that operational factors such as technical malfunctions or operational mistakes will cause or exacerbate credit or liquidity risk'; CPSS (2001). The simulations utilise real data for the sample period November 2004 and map the basic functionalities of ARTIS onto the simulation tool as closely as possible. The scenarios are designed according to the estimation of potential risk concentrations based on actual data for the sample period. Nevertheless, one must bear in mind that the results reported are the output of simulation experiments based on stylised operational failures rather than historical events. Operational incidents where disruptions exceeded a few hours among ARTIS participants have occurred too rarely and with too little impact on payment activity to provide a reliable data basis for an empirical assessment of operational risk based on the system's and the participants' data history.

The paper is structured along the following lines. In Section 7.2 we present the data on participation, transactions and liquidity in ARTIS;

---

[2] November 2004 was chosen as typical month of ARTIS activity; the results reported are not time sensitive.

in Section 7.3 we analyse concentration risk and the network structure in ARTIS; Section 7.4 describes the scenarios, presents the respective results and compares these across scenarios; Section 7.5 discusses their implications and Section 7.6 summarises the results and concludes the paper.

7.2 Participation, transactions and liquidity in ARTIS

This section provides basic statistical data for ARTIS for the sample period to lay the foundations for the simulation analysis in Section 7.3. More detailed statistical data for ARTIS can be found in Schmitz and Puhr (2006).

7.2.1 Participants

In November 2004 the system had a total of 575 accounts, which were held by banks, the federal government, non-financial companies and the OeNB itself. A large number of accounts were offset accounts (eg accounts of Geld Service Austria (GSA), a partial OeNB subsidiary in charge of cash distribution in Austria) and transfer accounts (eg transfer accounts that link ARTIS to the other national components of TARGET).[3] Austrian and international banks held 234 transaction accounts (excluding international institutions like the Bank for International Settlement, the International Monetary Fund and non-EU central banks as well as banks' offset accounts).

7.2.2 Transactions

Throughout November 2004 the average daily value of payments submitted in ARTIS amounted to EUR 32.6 billion.[4] The value was quite volatile, with a standard deviation of EUR 7.7 billion, which can be partly explained by three days significantly below the mean: 1

---

[3] Transfer accounts are ARTIS accounts held by other ESCB central banks at OeNB. All national TARGET components are directly linked by transfer accounts. All transactions to and from the respective country and Austria are routed via these transfer accounts.

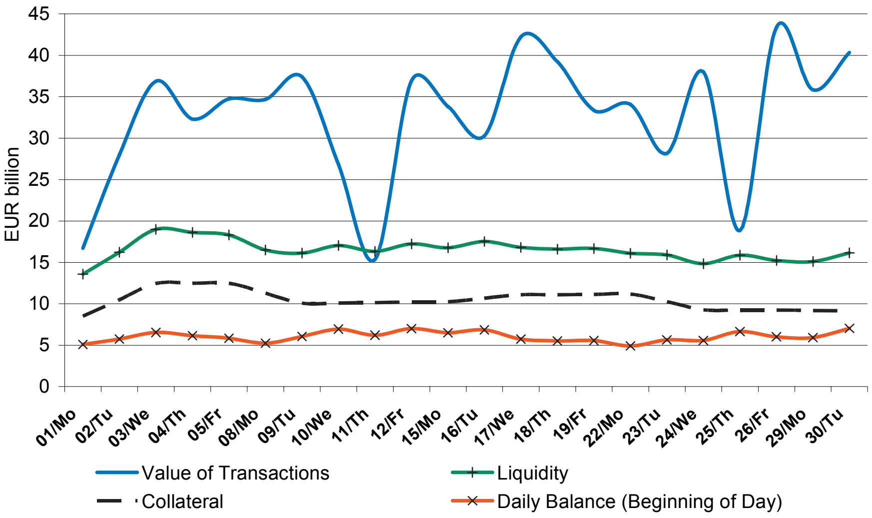
[4] Throughout the paper all values are rounded to one decimal place, which may lead to rounding errors at some points.

November (All Saints' Day, a public holiday in Austria), 11 November (US Bank Holiday, Veterans' Day) and 25 November (US Bank Holiday, Thanksgiving). The total amount of transactions submitted in the period was EUR 717.4 billion, which amounted to roughly three times nominal GDP in 2004.

7.2.3 Aggregate liquidity

The aggregate liquidity in the system exceeded actual use of liquidity, which could lead to the conclusion that it was sufficient, especially as all transactions submitted were settled and no accounts experienced liquidity shortages that would have led to unsettled transactions at closing time (06:00:00 pm). The average daily aggregate liquidity in the system – defined as beginning of day balances plus collateral available – equalled EUR 16.8 billion (Figure 7.1).

Figure 7.1 **Daily values for aggregate liquidity and its components – collateral, daily balance at the beginning of the day and value of transactions – in November 2004 (in EUR billion)**



Total liquidity was the sum of two components: daily beginning-of-day balances with a mean of EUR 6 billion and available collateral with a mean of EUR 10.4 billion. We interpret available collateral as a component of total liquidity in the system, although system participants must apply for daylight overdraft limits to liquidise it. However, pecuniary transaction costs of this procedure are zero and in terms of non-pecuniary transaction costs (ie time delay) they are effectively zero. In order to assess the aggregate liquidity stance of the system, we calculated indicators of liquidity usage. Firstly, aggregate collateral usage is defined as the share of available collateral that was liquidised by applying for daylight overdraft limits. With an average daylight overdraft limit of EUR 3.5 billion actually liquidised, average aggregate collateral usage amounted to 33.7%.

Secondly, the liquidity usage indicator measures the share of submitted transactions, which were settled by running down available liquidity rather than by received payments.[5] This ranges from 0 to 1, as its numerator is the difference between the beginning-of-day balance and the minimum balance during the day, and its denominator is the sum of all payments settled (see appendix). In our sample the indicator had a mean of 30% and a standard deviation of 3%. On average (across participants and across days), about one third of all settled transactions were covered by available liquidity and about two thirds by liquidity from received payments.

7.2.4 Disaggregate liquidity

Despite sufficient aggregate liquidity, individual accounts were occasionally illiquid and payment delays occurred frequently.[6] Throughout an average day payments with a total value of EUR 1.4 billion were queued (standard deviation EUR 0.6 billion or about 40% of the mean). They could not be settled immediately due to liquidity shortages of the submitting accounts. The settlement delay indicator averaged 0.16 across days, so that on average a submitted payment was queued for 16% of its potential queuing time.

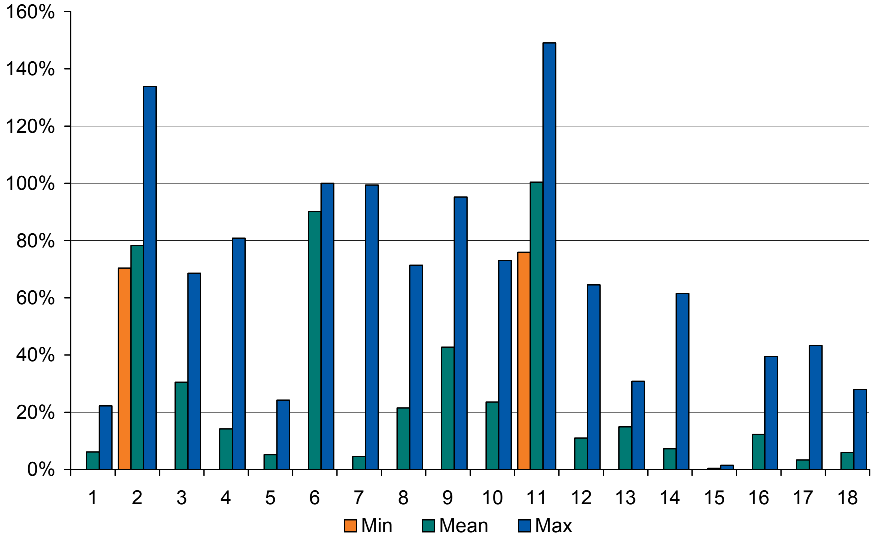
The disaggregated analysis of liquidity usage revealed that liquidity usage was highly heterogeneous across participants. Only 18 banks (out of 234) made active use of daylight overdraft limits on at least one day in November 2004 (Figure 7.2).

---

[5] The calculations of the liquidity usage indicator were conducted by using the Bank of Finland's Payment System Simulator BoF-PSS2.

[6] These results are based on the simulation of normal payment activity with real transactions and liquidity data for the sample period as reported queuing times were not available.

Figure 7.2 **Disaggregated analysis of daily usage of collateral by individual banks in ARTIS in November 2004 (in percentage of collateral available)**



NB Values in excess of 100% are due to additional short-term collateral supply by individual banks, which are not included in the daily averages of collateral posted on which the analysis is based. All daylight overdrafts must be fully collateralised at all times. The bank codes have been changed for this figure to make inference from the data on extent of collateral usage by individual banks impossible.

On average, 14.4 banks applied for daylight overdrafts per day. While the average overdraft limit across days and across the 18 banks corresponded to 26.3% of available collateral, the standard deviation was 31.2%. Eleven banks applied for daylight overdraft limits of up to 20% of their available liquidity on average, four banks of 20% to 50% and only three for more than 50% of their available collateral.

As liquidity is usually held as a buffer in case of unexpected large outflows of payments, the maximum daylight overdraft limits by individual banks throughout the month provide a better indication of individual collateral usage. Indeed, the average maximum daylight overdraft limit across banks amounted to 65.9% of individual available collateral (standard deviation 35.6%). Only one participant had a maximum value of below 20%, six had values between 20% and 50% and the remaining eleven values above 50%. Two participants even had maximum values of above 100%.[7] To sum up, the disaggregated analysis of collateral usage suggests that some banks actually used their individual liquidity reserves intensively. These results demonstrate that conclusions drawn from aggregate liquidity data do not necessarily apply to the individual participant level.

7.3 Risk concentration and network structure in ARTIS

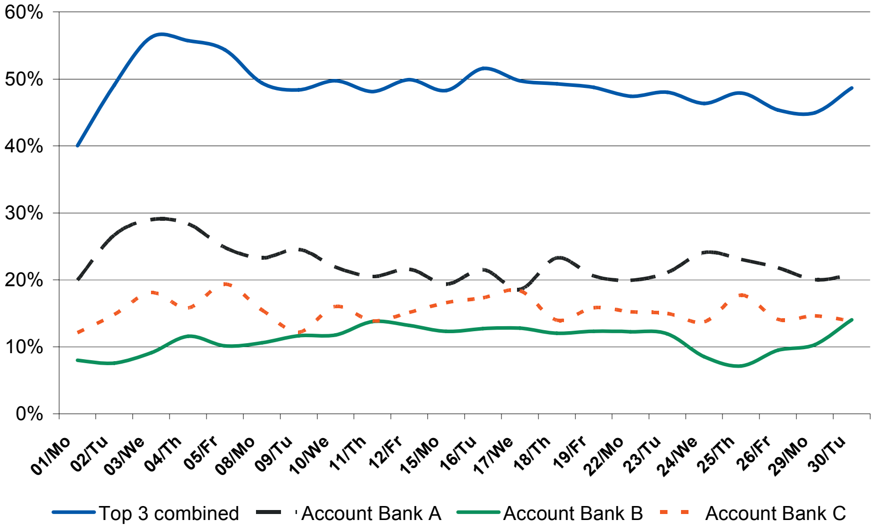
An analysis of risk concentration and network structure lays the foundation for the scenario design in Section 7.4. To identify the accounts that carried most risk, we first looked at measures of risk concentration: the value of liquidity concentrated at these nodes (liquidity concentration channel), the number and value of payments submitted and received (payment concentration channel) and the Herfindahl index of concentration of payment flows (both based on the number of payments and the value of payments received and submitted). Second, we analysed the network topology of the system. We did so for the monthly network as well as the daily networks.

Aggregate liquidity was highly concentrated in ARTIS. The liquidity concentration risk focuses on the share of liquidity (beginning-of-day balances plus collateral) a participant holds at the beginning of the day. The top three banks held almost half of the liquidity in the system, with individual values ranging from 11.1 to 22.5% of total liquidity (Figure 7.3).

---

[7] Despite the strict principle of full collateralisation of daylight overdraft limits, this is possible for the following reason: while we do have data on maximum daylight overdrafts, we do not have data on maximum daily available collateral. Data on the latter is available for the stock at 00:00:00 am for each day. Changes during the day in collateral that are reversed on the same day are therefore not reflected in the data; changes that carry over to the next day are measured ex-post. In order to capture these at least partly on the day they occur, we use daily averages for available collateral over two consecutive days. In the very few cases of large relative changes in available collateral, this leads to daylight overdrafts that seemingly exceed available collateral, although this in fact cannot take place.

Figure 7.3 **Liquidity concentration risk for the top three banks in ARTIS in November 2004**



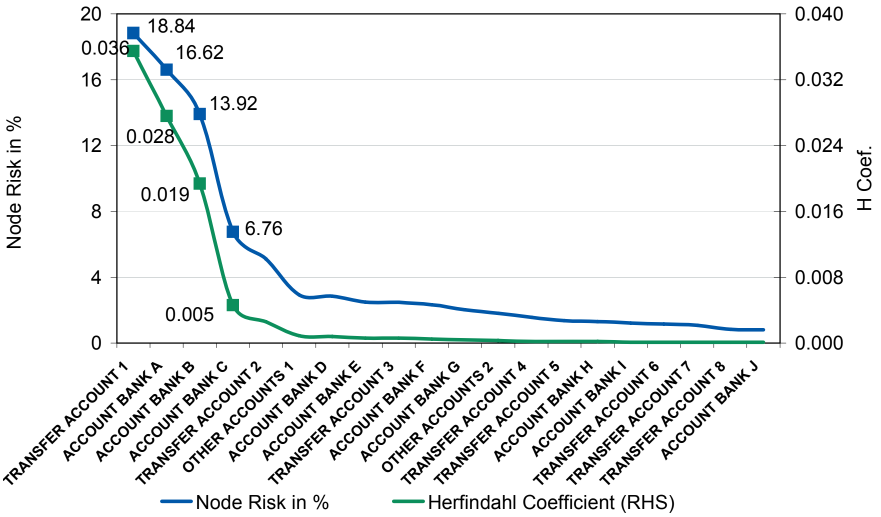
In general, payments were quite concentrated in the ARTIS system in the sample period. Payment concentration risk ('node risk') can be measured by reference to the value or to the number of transactions.[8] The share of the top three accounts of the total value of payments amounts to 49.4% and the share of the top five to 61.3%. In terms of the number of payments, concentration was much lower. The top three submitted and received 31.9% of the number of all payments and the top five 45.1%. While the values for the share of an individual bank in the total value of payments for the top three banks ranged from 13.9% to 18.8%, the corresponding individual node risks based on the total number of payments were only 8.4% to 13.8%. This indicates that the payments submitted and received by the most active accounts were also larger than those submitted and received by the less active accounts.

The Herfindahl Index (see appendix) for the value of payments for all 575 accounts was 0.0955. If the values of payments had been distributed uniformly, the index value would have been 0.0017 (or 1/575). The index was 56 times as large, whereby the conclusion of a non-uniform distribution and a concentration of payments was supported (Figure 7.4).

---

[8] James (2003).

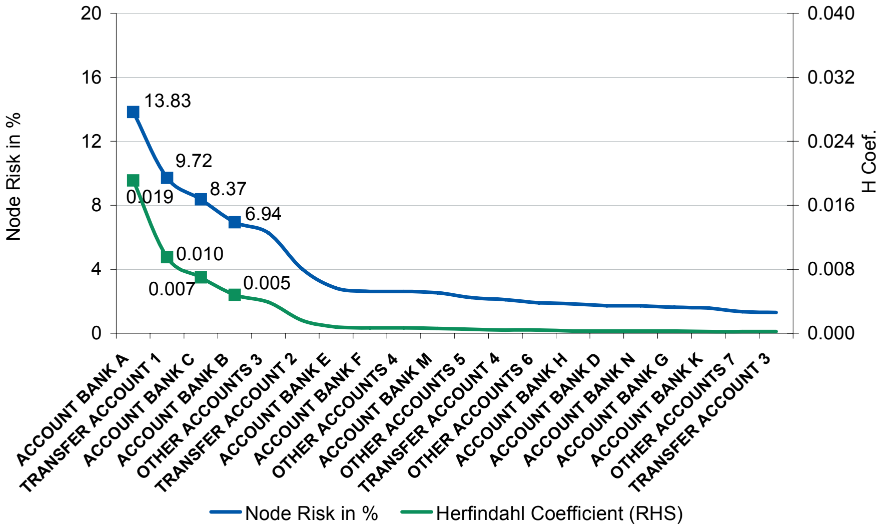
Figure 7.4 **Node risk and Herfindahl coefficient based on the total value of payments**



NB Only the top 20 accounts are included in the graphs.

The Herfindahl Index for the number of payments was 0.0530 and 31 times as large as the value compatible with a uniform distribution. The number of payments was therefore not uniformly distributed among accounts either (Figure 7.5). Additionally, the lower value of the Herfindahl Index for the number of payments is another indication for the conjecture that the more active nodes also processed higher-value payments.

Figure 7.5 **Node risk and Herfindahl coefficient based on the total number of payments**



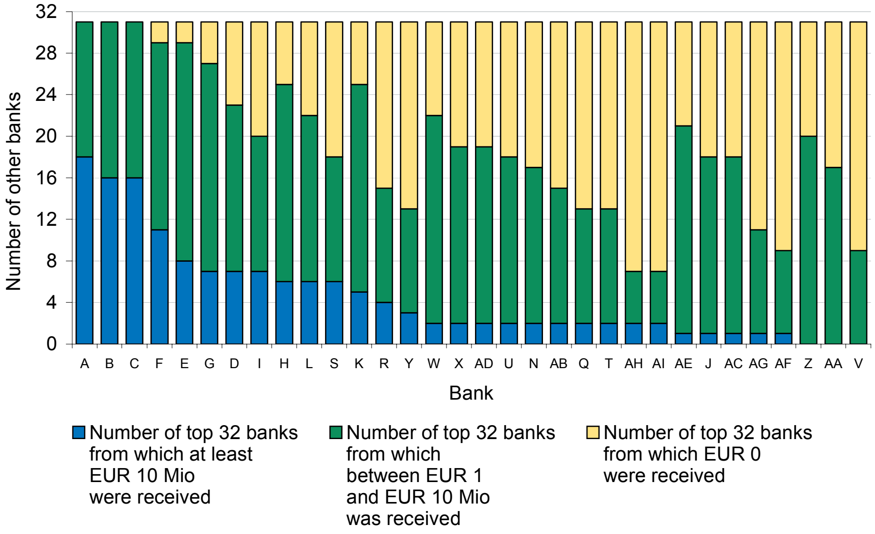
NB: Only the top 20 accounts are included in the graphs.

The finding of a high concentration of payment activity is supported by the analysis of the network structure among the top 32 participating banks.[9] Only the three most active accounts (Banks A, B and C) received payments from any of the other 31 banks among the top 32 on at least one day throughout the month, while the other top 32 banks received payments from an average of 17.9 other banks (Figure 7.6). The top three also received payments in excess of EUR 10 million from 16 to 18 other banks in the subsample on average (across days), while the average for the other banks was roughly 3.3.

---

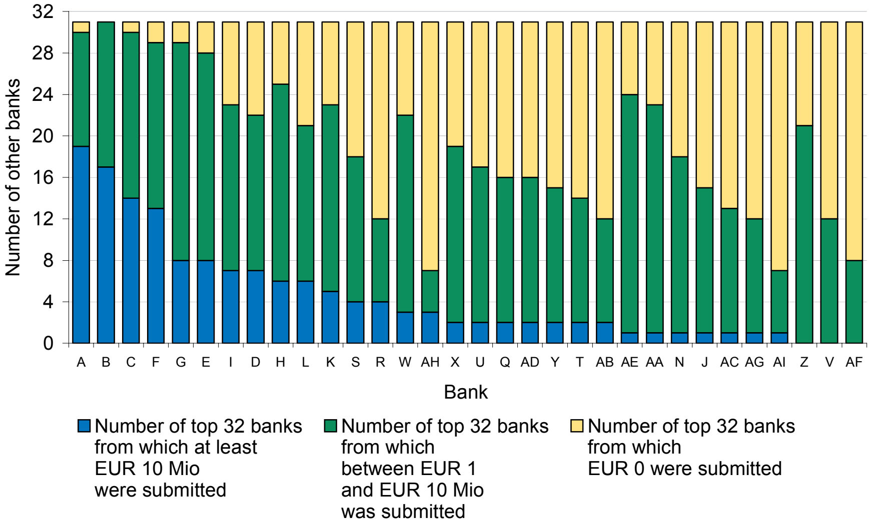
[9] Only the most active 32 banks (with a Herfindahl index exceeding 0.000049) were included in this analysis.

Figure 7.6 **Network structure of payments received within the network of the 32 most active banks (monthly network)**



A similar picture was presented by the network analysis of the payments submitted. Bank B submitted payments to any of the other 31 banks among the top 32 on at least one day throughout the month and Banks A and C to 30. The average of the remaining banks in the subsample was 18 (Figure 7.7). The top three also submitted payments in excess of EUR 10 million to 14 to 19 other banks in the subsample on average (across days), the other banks only to 3.3 on average.

Figure 7.7 **Network structure of payments submitted within the network of the 32 most active banks (monthly network)**



The analysis revealed that the most active banks, A, B and C, formed the core of the network structure among banks in the ARTIS system in terms of the value and number of payments received and submitted as well as connections to other banks throughout the month. A similar analysis of a subsample of the 51 most active accounts (including offset/transfer accounts) revealed that the most active transfer account (Transfer Account 1) also held a special position within the network structure.[10] In terms of other accounts it transacted with as well as large value payments (in excess of EUR 10 million) it was far less central to the network than the top three banks but still a cut above the rest.

The network structure throughout the month is a good first indication for scenario design. However, as the scenarios apply to individual days, we take a closer look at daily network topology.[11] Table 7.1 summarises the results for the entire network at the network level.

---

[10] Only the most active 51 accounts (with a Herfindahl index exceeding 0.000049) were included in this analysis.

[11] In the following analyses, we use the set of indicators of network topology and the notation of Soramäki et al (2006) in order to facilitate the comparison of the results.

Table 7.1 **Network Topology in ARTIS (across days for the entire network)**

	Mean	Median	Minimum	Maximum	Standard Deviation
Payments					
Volume of transactions	15,246.1	15,654	5,532	19,800	2,646.6
Value of transactions (EUR million)	32,608.6	34,376.5	15,354	43,403	8,701.4
Average value of transactions (EUR million)	2.2	2.1	1.2	3.0	0.4
Size					
Nodes	575	–	575	575	–
Active nodes n	193.5	192	172	218	9.3
Links m	1,421.6	1,447	700	1,556	167
Connectivity p (in %)	3.8%	3.9%	2.4%	4.2%	0.4%
Reciprocity r (in %)	21.2%	21.4%	18%	22.6%	1.1%
Distance measures					
Average path length l	2.2	2.2	1.8	2.3	0.1
Average eccentricity ε	3.0	3.0	2.3	3.8	0.3
Diameter D	5.7	6.0	5.0	7.0	0.6
Degrees					
Average degree k	7.3	7.5	4.1	7.8	0.8
Max out-degree	78.1	78.5	63	87	5.5
Max in-degree	77.8	74.5	71	130	12.1

NB Mathematical definitions of the indicators are provided in the appendix.

The average volume of transactions per day was 15,246 (standard deviation 2,647). The average value of transactions per day came to EUR 32,609 million (standard deviation of EUR 7,701 million). The average transaction size was EUR 2.2 million (standard deviation EUR 0.4 million). The size of the network is defined by the number of nodes n. There were 575 accounts in ARTIS in November 2004, but the average number of active nodes was only 193.5 (standard deviation 9.3). The active nodes were linked by an average of 1,422 links (m). The connectivity p of the network is captured by the number of actual links relative to the number of possible links. Connectivity p averaged 3.8% (of all possible links). It ranged from 2.4% to 4.2%, and its standard deviation was 0.4. Reciprocity r captures the extent to which existing links are bidirectional rather than unidirectional. In ARTIS about one-fifth of the existing links were bidirectional, the rest were active in one direction only.

An indicator of the distance between nodes is the lowest possible number of links that connect each (active) node with each other (active) node in the network. It is referred to as shortest path length. We calculated the average path length for each (active) originating node by averaging across terminating nodes and then averaged across originating nodes to derive the average path length l of the entire network. Across days this value equals 2.2. This means it only takes slightly more than two links to reach any terminating node in the network from any originating node in the network. The network is compact, mostly because almost all active nodes are linked to the largest banks. This network structure is quite stable across days as the standard deviation is low at 0.1. This conjecture is supported by the average eccentricity ε which is defined as the maximum path length averaged across nodes. In our system it is 3, hereby – it takes – at most – three links from the average originating node to reach any terminating node in the network (standard deviation 0.3). As to maximum eccentricity across nodes, defined as diameter D, picking an originating node at the very fringe of the network and counting the lowest possible number of links to the terminating node that is furthest away from it yields 5.7 links (standard deviation 0.6).

How well are the nodes connected to each other in the network? This is captured by the average degree k of the network which is calculated by summing across all (active) links originating from each node and then averaging across nodes.[12] This results in the measure m/n. Averaged also across days, it amounts to 7.3 in the ARTIS system (standard deviation 0.8). Pick an active node on a random day in November 2004 and it can be expected to have 7.3 links originating (or terminating) at it. However, the most active nodes have a much larger number of links originating and terminating at them. The maximum out-degree averages 78.1 across days, so that the most active node on each day has about 11 times as many links originating from it than the average node. The maximum in-degree (77.8) is correspondingly much higher than the average degree.[13]

In Figures 7.6 and 7.7 we presented the number of active links between banks within the network of the most active 32 banks across the whole month, ie the link was active on at least one day during the month. How does the picture change when we consider the average number of active links across days? For the subsample, the average degree k increases to 9.8. The maximum out-degree is 25.3 and the maximum in-degree 26.1. Thus on an average day the most active

---

[12] The out-degree refers to the number of links originating at the node while the in-degree is based on the number of links terminating at the node. Across the network the average out- and in-degree are equal to m/n.

[13] The large standard deviation of the in-degree is due to All Saints Day which is a public holiday in Austria but not all federal states of Germany where some markets and banks remain open.

bank submits payments to 81% of the other banks in the subsample and receives payments from 84% of them.

Which are the most active accounts in the network based on daily indicators of network topology for individual nodes? In order to design the scenarios for the simulations in section four, we calculated a number of indicators of network topology for individual accounts. The three most active banks in Figures 7.6 and 7.7 also had the highest daily average of links originating (out-degree k_i^{out}) and terminating at them (in-degree k_i^{in}). They submitted payments to between 64 and 76 other banks in the entire network on an average day, while the next largest submitted to only 41. Furthermore, they received payments from between 63 and 72 other banks, while the fourth largest bank only linked to 50. Even more significant the out-/in-degrees of the most active bank were about 10 times the average for all banks (7.3). The most active transfer account showed lower values than the top three banks but was still a cut above the other accounts (out-degree 69, in-degree 45). The four most active accounts featured average path lengths in the entire system across days of below 1.7, while the next lowest value was 1.9. The average path length for the whole network was 2.2.

To sum up, on the basis of the various measures of activity – the value of liquidity concentrated at these nodes (liquidity concentration channel), the number and value of payments submitted and received (payment concentration channel), and the Herfindahl index of concentration of payment flows (both based on the number of payments and the value of payments received and submitted) as well as network topology – the most active accounts were the same three banks and the most active transfer account.[14] This assessment holds true for the entire month as well as for each individual day.

7.4 The simulations

Potential contagion within payment systems is one of the reasons for the liquidity regulation of banks and for the lender-of-last-resort role of central banks. It is argued that the illiquidity of an individual bank can lead to liquidity problems at other banks. Illiquidity causes a non-individually attributable negative externality, a kind of market failure. Section 7.2 provided evidence that aggregate liquidity is quite high in

---

[14] See Schmitz and Ittner (2007).

ARTIS, which might lead to the hypothesis that the contagious impact of operational problems would be quite low. The purpose of the simulations is to investigate the empirical relevance of contagion in ARTIS and to quantify it in terms of value and volume of payments unsettled and in terms of the number of banks affected.

What does the data reveal about the contagion risk within the system with respect to the operational failure of one of the participants? Two channels via which operational incidents at one of the participants can have contagious effects on other participants can be distinguished: the payment concentration channel and the liquidity concentration channel.[15] The former focuses on the number of payments a participant is involved in as either submitter or receiver, the latter on the share of liquidity (beginning-of-day balances plus collateral) a participant holds at the beginning of the day.

To quantify either risk and its adverse effects, we conducted a large number of simulations based on three different scenarios for all transaction days in November 2004 with the Bank of Finland Payment System Simulator (BoF-PSS2). The simulator thereby recalculates each day's transactions by adding incoming payments to and subtracting outgoing payments from the participants' respective accounts. As transactions in the input data set provide time stamps, the simulator recalculates the balances of all participants to the system throughout the day, depending on the institutional features of the system (eg settlement algorithm, queue release mechanism). We could implement many of these directly by the parameterisation of the BoF-PSS2. However, some of the institutional features of the system could not be accounted for in the simulator and had to be mapped into the input data set. In addition, the simulator cannot take into account the behavioural reactions of system participants. Consequently, all relevant behavioural reactions of system participants must be determined exogenously by mapping them into the input data set. Nevertheless, this tool is widely used in studies of operational risk. For example, Bedford, Millard and Yang (2004) present findings in which the contagion effect of operational shocks that hit participants in CHAPS-Sterling is quite low. Most of the comparable studies are based on simulated aggregate liquidity levels, while ours utilises

---

[15] An operational incident at a participant who transacts with many other participants is expected to have a larger contagious impact as it is likely to lead to a larger withdrawal of liquidity from the system. For the same reason, an operational incident at a participant who holds a large share of aggregate liquidity is likely to have a larger contagious impact. Bedford, Millard and Yang (2004).

actual liquidity data and analyses the impact of operational risk on the system as well as individual banks.

7.4.1 The scenarios

The scenarios were designed on the basis of analysis of actual payment flows in ARTIS in the previous sections, with the objective of estimating the contagion effect within the system. That is, the extent to which those accounts not experiencing operational problems become subject to liquidity shortages and the extent to which risk could be systemic. We designed the scenarios along the following four dimensions.

First, we determined the nature and impact of operational failure. The impact of an operational incident was assumed to be the incapacitation of the participant from processing outgoing payments, ie the inability to submit transactions.[16]

Second, we selected the node(s) of the network of payment flows affected by the operational failure, ie the account(s) experiencing operational problems. We selected the most active nodes in the network based on the value of liquidity concentrated at these nodes (liquidity concentration channel), the number and value of payments submitted and received (payment concentration channel), the Herfindahl index of concentration of payment flows (both based on the number of payments and the value of payments received and submitted) as well as network topology. Here we built on the results of the previous section.

Third, we determined the duration of the operational failure of a participant, that is, for how many hours the participant is incapacitated by the incident. We conducted the simulations on the assumption of a one-day failure[17] to submit payments. Scenario design was guided by the principle that the shocks to the system should be exceptional but plausible. Anecdotal evidence in ARTIS suggests the shocks simulated are indeed exceptional but nevertheless plausible. Additionally, ARTIS provides business continuity arrangements for participants: in the case of operational failure they can submit

---

[16] It is assumed that the resulting illiquidity of the participant is not interpreted as potential insolvency by other participants in the payment system and the financial system at large.

[17] See also Bedford, Millard and Yang (2004).

payments by phone, fax, courier service or eKonto[18] if their internal systems remain fully functional.[19] As these methods are more costly, they are only employed for critical and/or large-value payments. In order to assess the impact of such back-up options, we re-ran simulations under the assumption that back-up options would be employed before the close of the system, ie after 10 hours of operational failure. The re-runs are based on the (rather restrictive) assumption that even very large numbers of payments can be processed via back-up options in a timely manner, ie before closing and on the assumption that the stricken bank's internal systems are fully functional.[20]

Fourth, since the simulator does not simulate the behavioural reaction of other system participants or the system operator to an operational incident, behavioural reactions must be assumed to be exogenous in two areas. First, other participants might want to stop submitting payments to the participant experiencing operational problems. In TARGET a stop sending rule applies if a transfer account of a central bank in the system experiences an operational problem. In such a case no further payments are transferred to the stricken transfer account. ARTIS operators, however, provided evidence that in all other cases participants continue to submit payments to participants experiencing operational problems – even if the latter cannot submit payments themselves for many hours. Although this is a restrictive assumption, it is well-supported by anecdotal evidence. According to ARTIS operators, banks explicitly state that they prefer to submit payments to stricken banks because they want to fulfil their (and their customers') obligations with respect to the stricken bank (or its customers) in a timely manner, irrespective of operational problems at the stricken bank. We are not aware of any evidence suggesting that banks impose bilateral sending limits. Consequently, we adopt the corresponding assumption for the scenarios concerning banks, even though this limits the plausible duration of an operational incident in the simulations to one day. For any longer operational failure, participants are more likely to react by discontinuing payment submittance to the participants with operational problems. Second,

---

[18] The eKonto is an alternative access mode to the ARTIS operating desk available to some but not all participants via an online account. Payments are submitted manually via the eKonto by the participant and must be further processed manually by the ARTIS operating desk.

[19] Otherwise the stricken participants would lack information on their respective payment obligations.

[20] A delayed closing is, in principle, possible on ECB approval.

participants could react to operational incidents by increasing available collateral. Anecdotal evidence suggests that participants already hold large shares of their assets that qualify as collateral at the OeNB. Depositing eligible assets with the OeNB is no more costly for system participants than depositing them with the Austrian central securities depository and can be even cheaper than depositing international assets with the respective foreign central securities depository. Provision of additional eligible collateral is likely to involve portfolio readjustment by participants and is therefore likely to be costly. Consequently, we assume that system participants do not increase collateral for durations of operational incidents of up to one day. This also limits the plausible duration of operational failure.

Finally, we defined three scenarios with the highest expected impact and the highest expected contagion effects according to step 2: (1) the first scenario assumes that the most active transfer account cannot submit payments to the system, (2) the second scenario assumes that the most active bank cannot submit payments to the system and (3) the third scenario assumes that the three most active banks experience operational failure simultaneously (eg due to a communication infrastructure breakdown) and cannot submit payments to the system. In all three scenarios we assumed that the operational incident would last from 1 day or 10 hours, respectively, according to step 3. We assume that the participants who are still fully operative continue to submit payments to the participants hit by operational problems, with the exception of the first scenario where payments continue to be submitted but the stop sending rule would be applied in accordance with the basic functionalities of ARTIS/TARGET. The simulations are based on actual liquidity data for November 2004, interpreted as binding liquidity constraints for banks.

7.4.2 Scenario 1 – top transfer account failure

The national TARGET operator in charge of the most active transfer account experiences an operational incident at 07:15:00 am. It cannot submit or settle payments for the rest of the day until the system closes at 06:00:00 pm. In response, ARTIS imposes a stop sending at

08:00:00 am in line with ARTIS/TARGET business continuity arrangements.[21]

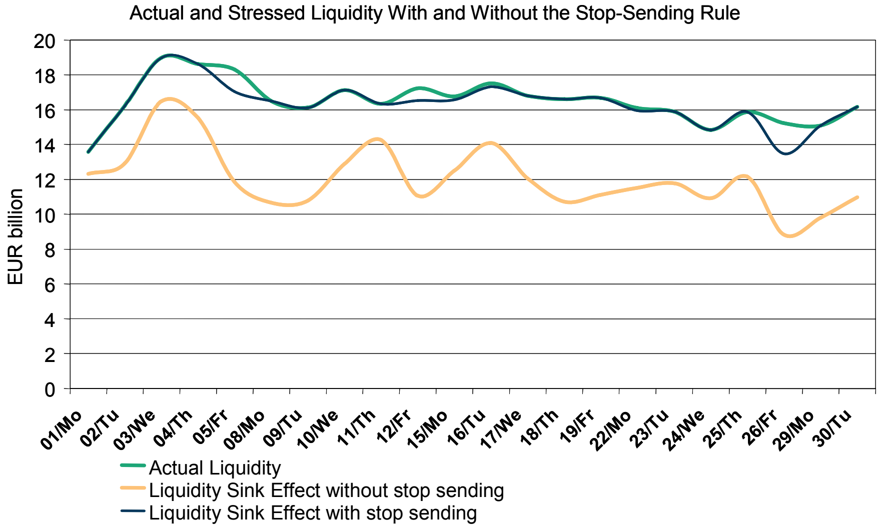
The impact of the scenario on aggregate liquidity and on the smooth functioning of the payment system

Aggregate liquidity in the first scenario was equal to actual aggregate liquidity at the beginning of the day as the transfer account did not hold a beginning-of-day balance or collateral. Consequently, the operational problems at this account did not lead to a liquidity drain effect, defined as a reduction of aggregate liquidity due to the liquidity reserves of the failed participant not being available for circulation in the system. But, the account's central position in the network structure of payments in ARTIS could lead to a liquidity sink effect (also referred to as a liquidity trap). This is defined as a reduction of aggregate liquidity available for circulation in the system that results from ongoing transfers of liquidity to the stricken account, where it discontinues circulating due to the account's operational problems. The more value is submitted to the stricken account, the higher the liquidity sink effect. The stop sending rule can therefore reduce the liquidity sink effect in this scenario. Ongoing transactions before the imposition of the stop sending rule (but after operational problems occurred) reduced aggregate liquidity available by an average of 1.2% of aggregate daily liquidity (Figure 7.8). The impact of the scenario on aggregate liquidity was quite volatile as the standard deviation of the impact was about 240% of the mean. The source of volatility is the liquidity sink effect, which differed substantially from day to day. As the analysis suggested that the stop sending rule would substantially reduce the contagion effect within the system to operational risk outside the system, we re-ran the simulations for all 22 days in the sample period without the stop sending rule. The results are presented after the results for the scenario with the stop sending rule.

---

[21] After about 30 minutes, national TARGET operators exchange information on the operational incident at the stricken central bank in a conference call and decide whether to impose the stop sending rule. So, it is a sensible estimate that it takes about 45 minutes in total to actually apply the stop sending rule.

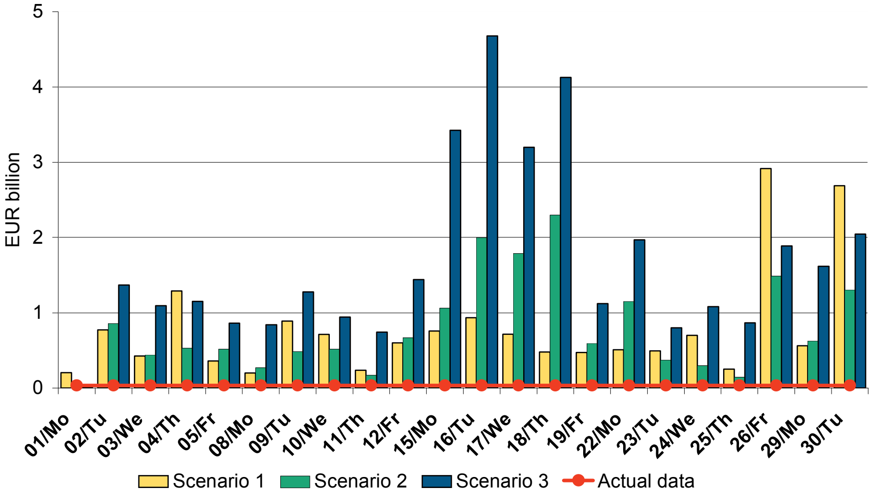
Figure 7.8 **Actual and stressed liquidity with and without the stop sending rule (based on Scenario 1)**



The value of payments submitted to the system amounted to an average of EUR 22.4 billion – with a standard deviation of EUR 5.8 billion. This corresponded to a decrease of 31.5% relative to the unstressed system. This value can be attributed to two sources: (1) the node risk (defined as the share of an individual bank of the total value of transactions) of the stricken account (18.8% of total value of payments submitted or received) as it cannot submit payments and (2) the stop sending rule (12.7%). The average daily value settled was EUR 21.6 billion. Relative to the unstressed scenario this corresponded to a reduction of 33.8%. The number of payments submitted averaged 12,832 per day during the sample period, corresponding to a reduction of 16.3%. This reduction is substantially larger than the node risk of the transfer account in terms of the number of payments (9.7% of total number of payments submitted or received), which can be attributed to the impact of the stop sending rule that kept banks from transferring money to a stricken account.

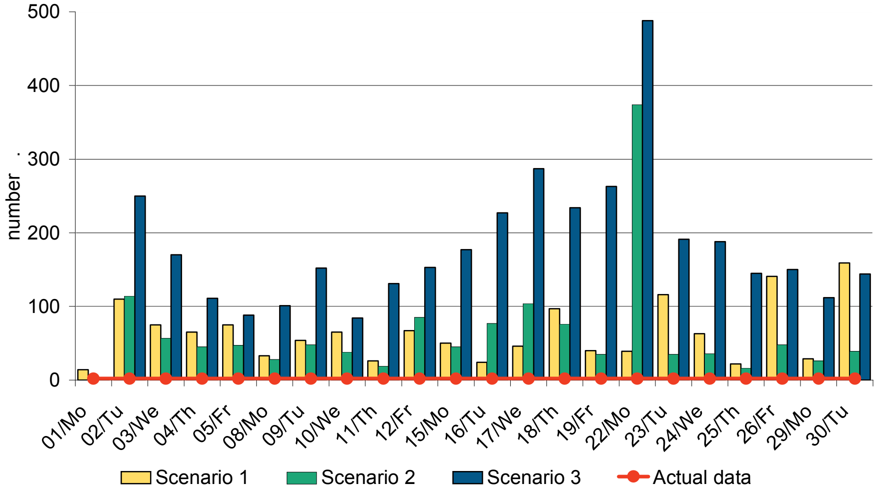
The contagion effect of the scenario on the other participants of the payment system was substantial in terms of aggregate value unsettled. The value unsettled was EUR 0.8 billion on average, or 3.5% of the average value submitted in the stressed system (Figure 7.9). In the unstressed system all payments submitted were settled. The value unsettled only refers to the payments submitted by the other participants (including those to the stricken transfer account). Thus, it does not include the payments from the stricken transfer account which could not be submitted. With a standard deviation of EUR 0.7 billion, it was rather volatile. The minimum value unsettled was EUR 0.2 billion and the maximum EUR 2.9 billion. The number of payments submitted but unsettled amounted to 64 per day on average (within a range of 14 to 159) (Figure 7.10). The large variations in value unsettled demonstrate that the same operational incident can impact the system in different ways on various days.

Figure 7.9 **Value of unsettled payments**



NB Actual refers to real data in November 2004.

Figure 7.10 **Number of unsettled payments**



NB Actual refers to real data in November 2004.

How much additional liquidity would be required to settle all transactions on each day? The value of transactions unsettled provides a first indication. However, it overstates the need for liquidity assistance as it fails to take into account the circulation of liquidity once injected into the system. The continuous liquidity usage indicator provides another estimate for the ratio of payments submitted that was covered by reserves. In Scenario 1 the indicator had an average value of 0.37 (compared to a value of 0.30 in the unstressed system), so that on average across days and across participants 37% of the total value submitted was paid for out of individual participants' liquidity reserves and 63% from payments received. Multiplying daily continuous liquidity usage with daily value unsettled provides an estimate of liquidity assistance required throughout the sample period that takes into account the circulation of liquidity. On average an injection of EUR 0.3 billion would indicate a lower bound of additional aggregate liquidity that enabled all accounts to settle.[22] This corresponded to 1.8% of liquidity available during the sample period. Across the sample period the necessary minimum liquidity assistance ranged from 0.1 (0.4% of actual aggregate liquidity available on that

---

[22] For the lower bound to suffice for settlement of all transactions, additional liquidity must be provided to those participants in the system that experience settlement failure, ie that actually need additional liquidity. Furthermore, the circulation of additional liquidity must equal the circulation of aggregate liquidity.

day) to EUR 1.1 billion (7.5% of actual aggregate liquidity available on that day). The average value of unsettled payments (EUR 0.8 billion or 4.7% of average aggregate liquidity in the unstressed system) provides an indication of an upper bound – the maximum amount – of additional liquidity required to prevent a contagion effect.

The impact of the scenario on individual banks

The contagion effect was substantial in terms of the number of individual banks with values unsettled. The number of banks subject to contagion averaged 12.1 per day and ranged from 8 to 18 per day of a total of 234 banks among the 575 accounts (Table 7.2). The total number of banks that failed to settle payments submitted on at least one day totalled 36, of which 2 could not settle on all 22 days and 10 on 50% of all days or more. Seven accounts failed on one day only. The impact of the scenario on individual banks therefore varies widely between days and among banks.

Table 7.2 **Number of banks with value unsettled**

Number of banks with unsettled payments	Actual	Scenario 1	Scenario 2	Scenario 3
Daily average	0	12.1	8.7	22.8
Minimum	0	8.0	0.0	1.0
Maximum	0	18.0	12.0	30.0
Standard deviation	0	2.4	2.8	5.9
Total	**0**	**36.0**	**38.0**	**56.0**

The impact of business continuity arrangements

In order to assess the impact of the employment of back-up options we re-ran the simulations assuming the duration of the operational failure to last until 04:00:00 pm rather than the rest of the day. This implies that the available back-up options are employed in a timely manner such that all payments can be processed before the closing of the system (06:00:00 pm). Furthermore, it must be assumed that the participant's internal systems are fully operational, so that it knows which payments to process. Under these assumptions all payments submitted were actually settled and no adverse effects in terms of payments unsettled were recorded for the stricken account or any other participant. However, the payments from the stricken account were delayed by up to 10 hours, which could increase the queuing times of other participants' payments. We did not further investigate this potential effect at this stage.

The impact of the stop sending rule

The stop sending rule substantially reduced the adverse impact of the operational shock and increased the resilience of the system. In order to assess the relative impact and thereby the efficacy of the stop sending rule, we replicated Scenario 1 without the stop sending rule while keeping all other features identical. Without the stop sending rule, the liquidity sink effect would increase from 1.2% to 26.9% of unstressed aggregate liquidity and the mean of the value submitted by EUR 4.2 billion (or 19.3%) (Table 7.3). This implies that the value of payments to the transfer account after 08:00:00 am (when the stop sending rule was imposed) must have equalled EUR 4.2 billion on average. Without the stop sending rule, the average value unsettled would increase from EUR 0.8 billion to EUR 1.3 billion. The average number of payments unsettled would almost double from 64.1 to 120.8.

Table 7.3 **Selected indicators in Scenario 1 with and without stop sending rule (daily values/ averages across November 2004)**

Indicator	Scenario 1 with stop sending rule (1)	Scenario 1 without stop sending rule (2)	Difference (1)–(2)
Aggregate liquidity (in EUR billion)	16.3	12.1	4.2 (26%)*
Liquidity reduction (in % of aggregate liquidity)	1.2	26.9	-25.7*
Value submitted (in EUR billion)**	22.4	26.7	-4.3 (-18.9%)
Value unsettled (in EUR billion)***	0.8	1.3	-0.5 (-71.8%)

* Differences in percentage of value with stop sending rule. ** Value submitted refers to the value of the payments submitted by participants not affected by operational problems. It excludes payments not submitted by the stricken bank due to operational problems. If the stop rule applies (column 1), the payments redirected in the queue due to the stop sending rule are not included in the value of payments submitted: the respective liquidity is still available to the banks, who can cancel submissions as long as they are queued. *** Value unsettled refers to the payments submitted by participants not affected by operational problems.

7.4.3 Scenario 2 – top bank account failure

The second scenario assumes that the most active bank cannot submit or settle payments from 06:00:00 am until 06:00:00 pm due to an operational incident. The mapping of the scenario included debit authorisation by the bank for a number of other participants in ARTIS.[23] Consequently, many payments from the stricken bank could be submitted (from the stricken bank's account via the participants to whom debit authorisation was granted) and settled despite operational problems at the stricken bank. Thus, debit authorisation can reduce the liquidity drain effect. To assess the impact of debit authorisation on

---

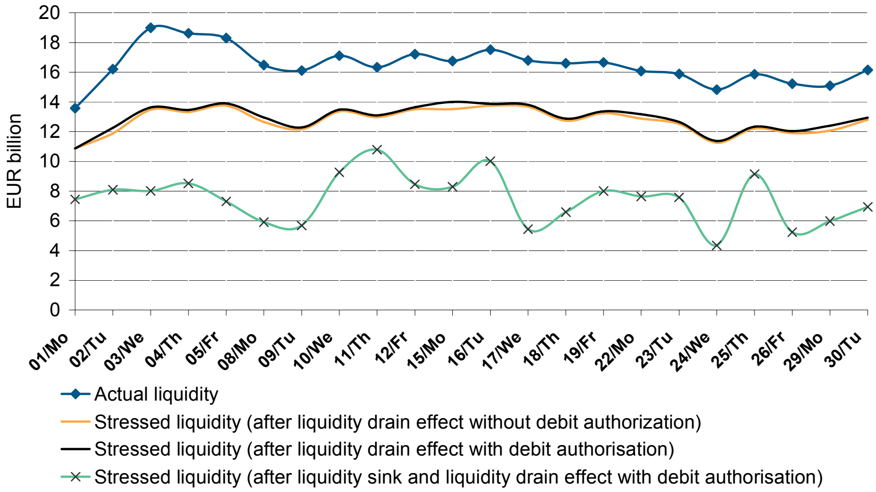
[23] Participant A can grant participant B a debit authorisation according to the Terms and Conditions Governing the OeNB's ARTIS system (Section 9). Debit authorisation is defined as the right of participant B to initiate (certain pre-agreed) payments from the account of participant A. Debit authorisations are granted to a small number of participants for prearranged purposes (very frequent recurring standard operations) and cannot be interpreted as a crisis mitigation instrument available at short notice in the case of an operational incident.

the contagion effect within the system, we re-ran the simulations based on a replicated scenario without debit authorisation.

The impact of the scenario on aggregate liquidity and on the smooth functioning of the payment system

The operational incident at the most active bank decreased aggregate liquidity available for circulation in the system (ie excluding the liquidity accumulating at the stricken bank) by an average of 54.6% to a daily average of EUR 7.5 billion – of which 21.6 percentage points were due to the liquidity drain effect and 33.16 percentage points due to the liquidity sink effect (Figure 7.11). The average daily value of payments submitted amounted to EUR 27.4 billion (standard deviation EUR 6.4 billion), a reduction of EUR 5.2 billion of the actual value in November 2004. The decrease of 16% corresponds to the stricken bank's usual share of payments submitted (as these cannot be submitted due to the operational incident) minus the value of payments submitted by debit authorisation – which can still be submitted as long as the accounts hold positive balances. The average value settled was EUR 26.6 billion, with a standard deviation of EUR 6.1 billion.

Figure 7.11 **Actual and stressed liquidity with and without debit authorisation (based on Scenario 2)**



Source: OeNB and own calculations.

The operational incident had a substantial negative contagion effect on aggregate payment activity as the value unsettled amounted to EUR 0.8 billion or 2.9% of the value submitted (Figure 7.9).[24] However, the impact of operational risk varied markedly from day to day as the value unsettled ranged from EUR 0.0 billion to EUR 2.3 billion. The number of payments unsettled rose to 63.3 on average and accounted for an average of 0.4% of payments submitted (Figure 7.10). A substantial value of payments could not be settled by participants that were not subject to operational problems, ie the contagion effect was substantial. Taking into account the circulation of liquidity, we estimate the lower bound of average liquidity injection required to settle all payments submitted to be in the order of EUR 0.3 billion or 1.9% of average aggregate liquidity in the unstressed system. The upper bound would correspond to EUR 0.8 billion or 4.9% of average aggregate liquidity in the sample period.

The mean of the continuous liquidity usage corresponded to 40%, so that under stress 40% of the payments submitted were settled from liquidity reserves. Compared to the unstressed scenario this implied an increase of about 10 percentage points. Despite a substantial contagion effect, the circulation of liquidity did not come to a complete halt.

The impact of the scenario on individual banks

The scenario had a substantial impact – that varied substantially from day to day – on other banks' ability to settle payments submitted. A total of 38 banks were affected by contagion throughout the month in this scenario (Table 7.2). This corresponded to 16.2% of the total number of banks in the system. On average 8.7 banks were unable to settle payments submitted each day, with a minimum of 0 and a maximum of 12. Four banks could not settle on 21 days, and seven on 50% of all days or more. Fourteen banks were affected on one day only. The scenario affected different banks in a different manner.

The impact of business continuity arrangements

We re-ran the simulations under the assumption that business continuity arrangements were invoked by 04:00:00 pm and all

---

[24] The value unsettled refers only to the payments submitted by the participants not hit by the operational incident and does not include the payments of the stricken bank (as these are not submitted). But it can include payments by other participants to the stricken bank.

payments of the stricken bank were settled before the close of the system. Under these assumptions all payments were settled and no contagion effect materialised in terms of unsettled payments at the end of the day. However, the resilience of the system rests on the assumption that the participant's internal systems are fully operational, so that the bank knows which payments to process, and that between 534 and 1,655 payments submitted via phone, fax, courier service or eKonto can be processed manually before 06:00:00 pm.[25]

The impact of debit authorisation

In addition to Scenario 2 with debit authorisation as reported above, we re-ran the scenario without debit authorisation. Relative to the latter, debit authorisation slightly attenuated the contagion effects within the system to operational shocks. Debit authorisation enabled payments to be settled that would otherwise not have been because of the operational problem at the stricken bank (as long as its account is sufficiently liquid, which is usually the case as the bank is unable to submit payments). Consequently, the average liquidity drain was lower than in a system without this feature. Without debit authorisation, the liquidity drain in Scenario 2 would have corresponded to the liquidity concentration at the stricken bank (22.5% of aggregate liquidity). Debit authorisation reduced the liquidity drain to 21.4%. As a result the value unsettled decreased from an average of EUR 1 billion to EUR 0.8 billion, which corresponds to a reduction of 15.6% of value unsettled (without debit authorisation). The number of unsettled payments was reduced from 137.3 to 63.3 on average per day. The average number of banks with value unsettled decreased from 10.3 to 8.7. The total number of banks affected by contagion was reduced from 42 to 38. Compared to its impact on the system, debit authorisation had a stronger impact on individual participants who had the right to access the account of the stricken bank. They were effectively shielded from direct effects of the operational incident.

---

[25] As noted above, a delayed closing is possible with ECB approval.

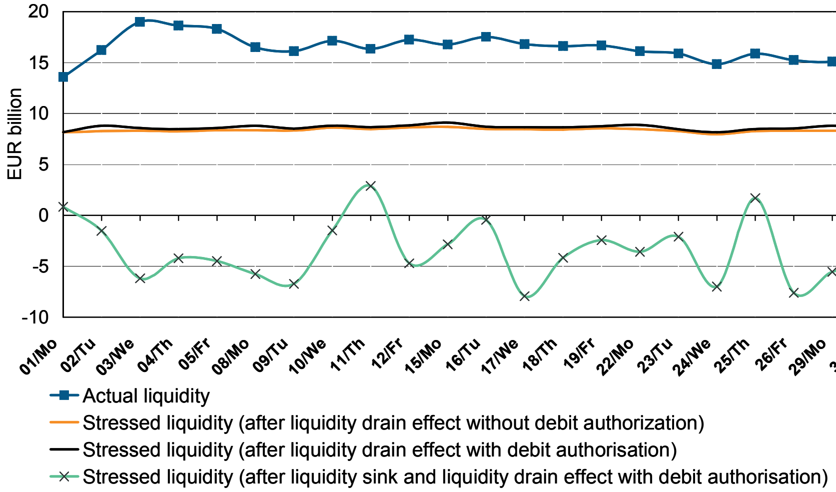
7.4.4 Scenario 3 – simultaneous failure of the three most active bank accounts

This scenario assumes that the three most active banks cannot submit payments from 06:00:00 am until 06:00:00 pm due to an operational incident. All three stricken banks granted debit authorisation to a number of other participants in ARTIS. To gauge its impact on the smooth functioning of the system, we re-ran the simulations based on a replicated scenario without debit authorisation.

The impact of the scenario on aggregate liquidity and on the smooth functioning of the payment system

Aggregate liquidity available for circulation in the system (ie excluding the liquidity accumulating at the stricken banks) decreased by (a theoretical) 121.5% of the unstressed level. Liquidity drain accounted for 47.4 and liquidity sink for 74.1 percentage points (Figure 7.12). In the case of all payments to the three stricken banks being settled, liquidity would have turned negative. In reality, aggregate liquidity available for circulation in the system (ie excluding the liquidity accumulating at the stricken banks) is bound below by zero. In Scenario 3 the liquidity sink basically withdrew all remaining liquidity from circulation and the adverse impact on the smooth functioning of the payment system due to contagion was very strong.

Figure 7.12 **Actual and stressed liquidity with and without debit authorisation (based on Scenario 3)**



Source: OeNB and own calculations.

The average value submitted was EUR 20.7 billion, which corresponded to a reduction of 36.43% relative to the unstressed system. This reduction equalled the share of the three stricken banks in the value submitted in the unstressed system (as their payments were not submitted due to operational incidents) minus the share of the value of payments submitted under debit authorisation. On average, value settled amounted to EUR 19.1 billion, which constituted a reduction of 41.6% from the unstressed value. Daily value unsettled was EUR 1.7 billion on average and ranged from EUR 0.2 billion to EUR 4.7 billion (Figure 7.9). On average, 175 payments could not be settled (with a range from 3 to 488) (Figure 7.10). Value and payments unsettled refer only to payments submitted by other participants (ie excluding the stricken banks' payments as these were not even submitted but including payments of the other participants to them). Taking into account the circulation of liquidity, we estimated the lower bound of additional liquidity necessary to settle all payments submitted to be EUR 1.1 billion on an average day (ranging from EUR 0.1 billion to EUR 3.2 billion). That amounted to 6.8% of aggregate liquidity in the unstressed system. These results also indicated that the impact of the scenario varied substantially across days. The upper bound of additional liquidity would correspond to EUR 1.7 billion (10% of aggregate liquidity in the unstressed system).

The participants in the system had to rely much more on their liquidity reserves rather than on incoming payments to settle outgoing payments. The continuous liquidity usage indicator increased from 29.9% in the unstressed scenario to 67.8%. Participants covered two-thirds of the value of payments submitted and settled by their liquidity reserves and only one-third by liquidity from incoming payments.

The impact of the scenario on individual banks

On average 22.8 banks failed to settle all payments submitted per day (Table 7.2). The minimum number of banks that experienced serious liquidity shortages was 1 and the maximum 30. The total number of banks that were unable to settle all payments on at least one day was 56. In the sample period, 1 bank could not settle on all 22 days and 24 banks on 50% or more. Ten failed to settle on a single day only. The effect of the scenario on individual banks varied markedly between days and banks.

The impact of business continuity arrangements

In order to assess the impact of alternative submission channels we re-ran the scenario under the assumptions introduced in Scenario 2. Here all payments were settled and the scenario had no negative effects in terms of unsettled payments at the end of the day. Under these assumptions the system proved to be resilient even to a very strong negative shock. For the business continuity arrangements in place this implied that between 1,440 and 4,022 payments would have to be processed manually before closing at 06:00:00 pm.[26]

The impact of debit authorisation

In addition to the scenario with debit authorisation reported above, we re-ran Scenario 3 without debit authorisation. Compared to the latter, debit authorisation reduced the liquidity drain effect on aggregate liquidity by an average of EUR 0.3 billion or 1.5% of aggregate liquidity per day. Value unsettled decreased from EUR 1.9 billion to EUR 1.7 billion. The number of unsettled payments on average went

---

[26] As noted above, a delayed closing is possible on ECB approval.

down from 267 to 175. The average number of banks affected by contagion was reduced to 22.8 (down from 24.6). The number of banks with value unsettled on at least one day in the sample period decreased from 60 (without debit authorisation) to 56 (with debit authorisation). Debit authorisation slightly decreased the impact of the scenario on the system under stress. The impact on the liquidity position of the participants with the right to access the accounts of the stricken banks was more substantial. Debit authorisation insulated them effectively from any direct impact of the operational incident (as long as the stricken bank's account was sufficiently liquid).

7.4.5 Comparison across scenarios

Taking into account the business continuity arrangements, no scenario had an adverse impact on the smooth functioning of the payment system (all payments were settled at the end of the day, even if queuing times might have increased). Given the very restrictive assumptions underlying the efficacy of the business continuity arrangements, we compared the impact of the scenarios without business continuity arrangements. Among the three scenarios, Scenario 3 had the strongest impact on aggregate liquidity, on value unsettled and on the number of banks with unsettled payments as well as on the frequency of settlement failure (Table 7.4). However, one must bear in mind that Scenario 3 was designed as a worst-case scenario. Scenarios 1 and 2 featured very similar values unsettled, numbers of unsettled payments and total numbers of banks with unsettled payments. This similarity of impacts is quite surprising, taking into account the large differences in liquidity reduction (Scenario 1 with 1.2% and Scenario 2 with 54.8% of aggregate liquidity). In addition, the stop sending rule applied only to Scenario 1.

Table 7.4 **Selected indicators in all three scenarios and in the actual data (daily values/ averages across November 2004)**

Indicator		Actual	Scenario 1	Scenario 2	Scenario 3
Aggregate liquidity (in EUR billiom)		16.5	16.3	7.3	-3.8
Liquidity reduction (in % of aggregate liquidity)		0.0	1.2	54.8	121.5
of which	Liquidity drain (in % points)	0.0	0.0	21.6	47.4
	Liquidity sink (in % points)	0.0	1.2*	33.2	74.1
Value submitted (in EUR billion)		32.6	22.4	27.4	20.7
Without business continuity arrangements					
Value unsettled (in EUR billion)		0.0	0.8	0.80	1.66
Value unsettled (in % of value submitted)		0.0	3.3	2.7	7.7
Number of payments unsettled		0.0	64.1	63.3	175
*With business continuity arrangements***					
Value unsettled (in EUR billion)		0.0	0.0	0.0	0.0
Value unsettled (in % of value submitted)		0.0	0.0	0.0	0.0
Number of payments unsettled		0.0	0.0	0.0	0.0

* With stop sending rule – without stop sending rule the respective value would be 26.91%. ** One has to bear in mind that the assumption that all payments can be submitted by the stricken bank via back-up options and can further be processed by ARTIS operators manually in time is rather restrictive.

7.5 Implications

The implications of the results for payment system design and payment system oversight need to take into account the issue of practicability and efficiency, as stipulated in Core Principle VIII.[27] The marginal costs of implementing additional security features and business continuity arrangements must not outweigh the marginal (pecuniary and non-pecuniary) return from increased reliability.

The simulations take into account the available business continuity arrangements by reopening the submission channel for the stricken

---

[27] 'Core Principle VIII – The system should provide a means of making payments which is practical for its users and efficient for the economy.' (CPSS 2001).

bank(s) at 04:00:00 pm. In the simulations many transactions were queued until that point of time and the system settled all these transactions between 04:00:00 pm and 06:00:00 pm. However, this implied that for business continuity measures to be effective and for service levels to be met even under stress, some 1,500 to 3,400 payments (depending on the scenario) would have to be processed manually. On peak days in the worst case scenario this rises to about 4,000 payments. This assumption is very restrictive. The time available to complete the task crucially depends on the point in time when the stricken bank switches to alternative submission procedures and on the processing capacities available at the central platform. Assuming that about 30 payments per hour can be processed manually by one staff member, business continuity would require substantial additional human resources and equipment to reach the required payment throughput before closing (06:00:00 pm) while maintaining a high level of processing quality.[28]

In order to reduce contagion within the system even under stress, existing contingency procedures could be complemented by a stop sending function, comparable to the one employed in Scenario 1. This rule would consist of informing other participants that a particular account cannot submit payments and the option for them to redirect their payments to the stricken bank to a queue. In principle, the queued payments remain available to the sending bank in ARTIS. Once the stricken bank has resolved its operational problems, all payments in the queue are released and settled. A stop sending function would substantially reduce the liquidity sink effect. Additionally, it would be simple and practical, as suggested by the interpretations of Core Principle VII; CPSS (2001). Nevertheless, in order to assess the exact impact of a stop sending function, further simulations based on Scenarios 2 and 3 would have to be conducted.

7.6 Summary

The first objective of this paper was to analyse the liquidity stance, the risk concentration and the network structure in ARTIS. The aggregate liquidity in the system exceeded actual use of liquidity, whereby it seemed sufficient. All transactions submitted were settled and no accounts experienced liquidity shortages that would have led to

---

[28] In Basel Committee of Banking Supervision (2005) Principle 6 calls for tests of business continuity plans and for the evaluation of their effectiveness.

unsettled transactions at closing time (06:00:00 pm). Despite sufficient aggregate liquidity, individual accounts were occasionally illiquid and payment delays occurred frequently. The disaggregated analysis of liquidity usage revealed that liquidity usage was highly heterogeneous across participants. Furthermore, it suggested that some banks actually used their individual liquidity reserves extensively. These results demonstrate that conclusions drawn from aggregate liquidity data do not necessarily apply to the individual participant level. In general, the value and the number of payments received and submitted were quite concentrated among the top three banks and the top transfer account in the ARTIS system during the sample period. In addition, the most active accounts also received and submitted larger payments than the less active accounts. This conclusion is supported by the analysis of the network structure of the entire network, the subsample of the top 32 participating banks and that of the most active 51 accounts: the four most active accounts (Banks A, B and C as well as Transfer Account 1) formed the core of the network structure. In addition to payments, liquidity was rather concentrated among the top three banks in the system too, as these held about 50% of available liquidity. An analysis of operational risk in the system should focus on operational problems at the institutions with high payment concentration risk and high liquidity concentration risk to test for high-impact scenarios. It would, therefore, have to focus on the top three banks and the top transfer account.

The second objective of this study was to quantify the contagion effect of an operational incident outside the system on the ability of other participants not hit by an operational problem to settle payments. The methods applied were model simulations of operational shocks for the sample period November 2004.

In the unstressed scenario sufficient aggregate liquidity guaranteed the smooth functioning of the system. All transactions submitted were settled and no account experienced liquidity shortages that would have led to unsettled transactions at closing time (06:00:00 pm) on any day in the sample period.

We conducted simulations based on three different scenarios which we derived from the risk concentration and network analysis in this paper. These scenarios were designed to take into account the two main sources of contagion risk in payment systems: the payment concentration channel and the liquidity concentration channel. The shocks were exceptional but plausible operational incidents. The main findings of the simulations were as follows:

1. ARTIS is highly reliable, as long as the existing business continuity arrangements prove effective. Under the restrictive assumptions that (1) the stricken bank(s) had information concerning their payment obligations and (2) all payments submitted by the stricken bank(s) via phone, fax, courier service or eKonto could be settled in time, no adverse effects on its own payments or any other participant's payments were recorded. The system functioned smoothly even under severe stress.

2. The simulations provide a quantification of the high demands on the business continuity arrangements in times of distress. The numbers suggest it would be unlikely they would be fully effective. So we demonstrated how the results of the simulations could be used also to test the efficacy of the business continuity arrangements.

3. Without the use of business continuity arrangements, or in a case where they do not prove fully effective, the contagion effect on the smooth functioning of the payment system was substantial in all three scenarios. Contagion can occur despite the high level of aggregate liquidity in the system. A non-negligible number of banks failed to settle payments. We also provided estimates of the amount of Emergency Liquidity Assistance (ELA) needed to support the smooth functioning of the payment system. In these estimates we took account of the estimates of the flow of liquidity even under stress, which substantially lower the amount of ELA required.

4. The simulations revealed that the same operational incident had very different impacts on the system on various days. They also affected individual banks to a very different extent. In addition, each scenario's impact on individual banks differed from day to day. More research is called for to better understand the determinants of the impact of shocks on the system, on its participants and across days.

5. Furthermore, we investigated the impact of two noteworthy features of ARTIS on the contagion effect – the stop sending rule and debit authorisation: The stop sending rule substantially reduced the contagion effect of the operational shock and increased the resilience of the system. Currently, the stop sending rule applies only to operational problems at another TARGET central bank. Our research indicates that a similar rule for

operational incidents at commercial banks would strongly increase the resilience of the system. Further research is needed to put this hypothesis to test. Debit authorisation is not a risk mitigation tool per se as it is granted for special purposes and not available at short notice. Nonetheless, it has an impact on contagion. Debit authorisation attenuated the reaction of the system to operational shocks much less than the stop sending rule but still to a non-negligible extent. More importantly, it proved effective in insulating the participants who had access to the stricken bank's account via debit authorisation from the operational incident.

References

Bank of Finland (2005) **Descriptions of BoF-PSS2 Databases and Files.** Version 1.2.0. February.

Basel Committee on Banking Supervision (2005) **High-level principles for business continuity.** Consultative document by the Bank for International Settlements. Basel.

Bech, M L – Garatt, R (2003) **The Intraday Liquidity Management Game.** Journal of Economic Theory 109, 198–219.

Bedford, P – Millard, S – Yang, J (2004) **Assessing operational risk in CHAPS Sterling: a simulation approach.** Bank of England Financial Stability Review, June, 135–143.

CPSS – Committee on Payment and Settlement Systems (2001) **Core Principles for Systemically Important Payment Systems.** Bank for International Settlements. Basel.

Galos, P – Soramäki, K (2005) **Systemic risk in alternative payment system designs.** ECB Working Paper No. 508.

James, K R (2003) **A statistical overview of CHAPS Sterling.** Bank of England Financial Stability Review, June, 115–121.

Koponen, R – Soramäki, K (2005) **Intraday liquidity needs in a modern interbank payment system – a simulation approach.** In Leinonen, H (ed.) Liquidity, risk and speed in payment and settlement systems – a simulation approach, Bank of Finland Studies E:31, 73–116.

Leinonen, H (ed.) (2005) **Liquidity, risk and speed in payment and settlement systems – a simulation approach.** Bank of Finland Studies E:31.

Leinonen, H – Soramäki, K (2005) **Optimising liquidity usage and settlement speed in payment systems.** In Leinonen H (ed.) Liquidity, risk and speed in payment and settlement systems – a simulation approach. Bank of Finland Studies E:31, 117–150.

Mazars, E – Woelfel, G (2005) **Analysis, by simulation, of the impact of a technical default of a payment system participant.** An illustration with the PNS system. Banque de France Financial Stability Review No. 6, June, 113–124.

Schmitz, S W – Ittner, A (2007) **Why central banks should look at liquidity risk.** Central Banking, Vol. XVII, No. 4, 32–40.

Schmitz, S W – Puhr, C (2006) **Liquidity, Risk Concentration and Network Structure in the Austrian Large Value Payment System.** http://ssrn.com/abstract=954117

Schmitz, S W – Wood, G E (eds.) (2006) **Institutional Change in the Payments System and Monetary Policy.** Routledge. London.

Soramäki, K – Bech, M L – Arnold, J – Glass, R J – Beyeler, W E (2006) **The Topology of Interbank Payment Flows.** Federal Reserve Bank of New York Staff Report No. 243.

McVanel, D (2005) **The impact of unanticipated defaults in Canada's Large Value Transfer System.** Mimeo, Department of Monetary and Financial Analysis Bank of Canada. Ottawa, Ontario.

Appendix 1

The *lower bound of liquidity* is defined as the theoretical minimum of the aggregate amount of liquidity in the system that enables all participants to settle all payments submitted in end-of-day partial or multilateral netting. It is below the corresponding value for pure RTGS systems without queuing as the latter are less liquidity efficient. However, it is identical to the corresponding minimum liquidity in a RTGS with queuing. The theoretical minimum assumes that all liquidity in the system is allocated in perfect accordance with individual liquidity demands. It is calculated in the following manner

$$\max\left(0, \sum_{k=1}^{d} a_{i,k} - \sum_{j=1}^{n}\sum_{k=1}^{d} a_{j,k} \middle| \left(r_{j,k} = i\right)\right)$$

where $a_{i,k}$ is the value of the payment. The first summation is the value of all payments submitted and the second one that of all payments received. If the participant experiences a net inflow of liquidity during the day, its lower bound is zero. If he experiences a net outflow of liquidity, this amount defines the lower bound. The corresponding lower bound at the system level is calculated as the sum of individual values.

The *turn over ratio* indicates how often each euro of the aggregate stock of liquidity is spent during one day. It is calculated according to the following formula

$$\frac{\sum_{i=1}^{n} S_i}{\sum_{i=1}^{n} L_i}$$

where S_i and L_i denote the sum of settled payments submitted by participants i during the day and the available liquidity of participant i, respectively.

The *liquidity usage indicator* measures the share of submitted transactions, which were settled by running down available liquidity rather than by received payments. It is calculated according to the following formula

$$\frac{\sum_{i=1}^{n}\left(b_i^0 - b_i^{min}\right)}{\sum_{i=1}^{n} S_i}$$

where b_i^0 denotes the beginning of day balance of participant i b_i^{min} denotes the minimum balance of participant i during the day and S_i the sum of settled payments submitted by participants i during the day. Its range is from 0 to 1. Unsettled transactions are not included in the calculation.[29]

The individual *node risk* is defined as the share of an individual bank in the total value of transactions (or in the total number of transactions) according to the formula for each participant i

$$\frac{\text{Payments}_{i,\text{Submitted}} + \text{Payments}_{i,\text{Received}}}{\sum_{i=1}^{n} \text{Payments}_{i,\text{Submitted}} + \text{Payments}_{i,\text{Received}}}$$ [30]

The *Herfindahl Index* measures the concentration of the number of payments (or similarly of their value or of the liquidity of participants) among all n participants based on the following formula

$$\sum_{i=1}^{n}\left(\frac{\text{Payments}_{i,\text{Submitted}} + \text{Payments}_{i,\text{Received}}}{\sum_{i=1}^{n} \text{Payments}_{i,\text{Submitted}} + \text{Payments}_{i,\text{Received}}}\right)^2$$

If payments are uniformly distributed across all participants, the index value is 1/n, which is also its minimum value. Its maximum value is 0.5, which implies that all transactions take place between two participants only.[31]

---

[29] Koponen and Soramäki (2005).

[30] Bank of Finland (2005).

[31] James (2003) and Bedford, Millard and Yang (2004).

The network topology indicators[32]

An indicator of the *distance* d_{ij} between nodes is the lowest possible number of links that connects each (active) node i with each other (active) node j in the network. It is referred to as shortest path length.

We calculated the *average path length* for each (active) originating node l_i by averaging across terminating nodes j and than averaged across originating nodes i to derive the average path length l of the entire network.

$$l_i = \frac{1}{n-1}\sum_{j\neq i} d_{ij}$$

$$l = \frac{1}{n}\sum_i l_i$$

Average eccentricity ε is defined as the maximum path length (eccentricity ε_i which in turn equals the maximum distance between the originating node i and the terminating node j) averaged across nodes

$$\varepsilon_i = \max_j d_{ij}$$

$$\varepsilon = \frac{1}{n}\varepsilon_i$$

Considering the maximum eccentricity across nodes defines the *diameter D*

$$D = \max_i \varepsilon_i$$

The *average degree* k of the network is calculated by summing across all (active) links originating from each node (out-degree k_i^{out}) or terminating at each node (in-degree k_i^{in}) and than averaging across nodes

$$k = \frac{1}{n}\sum_i k_i^{out} = \frac{1}{n}\sum_i k_i^{in} = \frac{m}{n}$$

---

[32] We follow the notation of Soramäki et al (2006).

Chapter 8

How can banks control their exposure to a failing participant?

Elisabeth Ledrut

8 How can banks control their exposure to a failing participant?

Abstract

This paper assesses the effect of counterparties' reaction to an operational failure at one of the biggest participants in the Dutch interbank payment system. Counterparties react according to two basic rules: they stop sending payments to the stricken bank either after some pre-determined time or after their exposure to the stricken bank reaches a certain level. The simulations are based on historical liquidity levels. The impact of the disruption is quantified in terms of the additional liquidity needed to settle all payments that can settle given the banks' intraday reserves and collateral facilities. Assuming the disruption lasts for the remainder of the day, banks are faced with costs as they need to borrow this additional liquidity overnight from the market or from the central bank. From a cost perspective, response seems to be more effective when determined by the individual exposure of the stricken banks' counterparties than when triggered by the elapsed time after the disruption. However, even an immediate reaction does not prevent banks from running losses following the failure of a major participant. How much each payment system participant can control its exposure to the stricken bank partly depends on the degree of reciprocity in the value of bilateral payments.

8.1 Introduction

The Payment and Settlement Simulator ('the simulator') developed by the Bank of Finland has proven to be a useful tool to gain insight in risks and efficiency in payment systems. It has been used to estimate the trade-off between the liquidity needs of different types of systems and their level of risk (Koponen and Soramäki, 1998, or Leinonen and Soramäki, 2005, among others) and to assess the systemic impact of an operational disruption, either at a major market participant or at the system itself (McVanel, 2005, among others). These latter studies rely on historical data, adapted in order to create various scenarios for the disruption, implicitly assuming that, were a disruption to occur, (other) payment system participants would not adapt their behaviour to the new environment – an exercise that could be subject to a form of Lucas critique.

Behavioural aspects have received less attention. This can be explained by a lack of knowledge about participants' reactions in the face of a disruption or about the factors affecting participants' reaction patterns. There is some historical evidence with regard to wide-scale disasters like September 11, but one should be cautious when generalising findings from such exceptional events. There are also some observations by payment system operators, mainly about participants' reactions to failures to pay by other participants. For example, work by Bedford et al (2005) is based on the assumption that participants in the United Kingdom's payment system would react to such a failure within ten minutes. Empirically, one study in particular – Conover (2005) – researched occurrences of disruptions and estimated the strength of other participants' reactions. Theoretically, the reaction of payment system participants to (the absence) of incoming payments was expressed in the form of a reaction function (McAndrews, 2002), which assumes that part of participants' outgoing payments is triggered by payment receipts and that payment orders exhibit a certain degree of synchronicity.[1]

If one of the major participants in the Dutch interbank payment system failed to send payments, how would other participants' reaction influence the impact of that disruption? How can payment system participants mitigate their vulnerability to such an event? Insight in the influence of payment system participants' behaviour on the system would allow central banks to identify and encourage stabilising behaviours, with a view to lowering the potential systemic impact of operational disruptions.

This paper assesses the impact of an operational failure[2] and the effect of different possible counterparty reactions. Counterparties react according to two basic rules: they stop sending payments to the stricken bank either after some pre-determined time or after their

---

[1] McAndrews empirically estimated the slope of that function, both under 'normal circumstances' and for the days following September 11, and found it to be generally quite steep (indicating a high level of coordination) and significantly less so in the days following the attacks (coordination broke down).

[2] The terms 'operational failure', 'failure', 'disruption' etc refer to the bank's inability to send payment messages. They do not in any case refer to the institution's possible insolvency. An insolvency of a major bank would indeed not come as a surprise, and its counterparties would have taken measures in advance of the default. The aim of this exercise it to assess the impact of an unexpected failure to send payments, and to see whether counterparties can control their intraday exposures to a major participant. This bank is called 'failing' or 'stricken' (as in Bedford et al, 2005).

exposure to that bank reaches a certain level. On the contrary to other work, the liquidity available to banks is considered as fixed. The impact of the disruption is explicitly quantified in terms of the additional liquidity needs faced by payment system participants and in terms of the costs of additional overnight lending. Indeed, when a participant fails to pay, other participants bear the burden of that failure by having to rely on intraday credit from the central bank. If banks have not received the expected payments at the end of the day, they will need to borrow overnight from the central bank or from the market.

8.2 Payment system participants' reaction to a disruption

8.2.1 How do counterparties react?

To my knowledge there are no thorough empirical analyses of payment system participants' reactions following an operational disruption at one of their counterparties. Part of this stems from the fact that banks' reaction is related to the intraday counterparty credit limits that they have built into their internal systems. These limits are not communicated to central banks. Payment system operators' interactions with market participants and their observations of day-to-day payment activity yield some valuable information on participants' behaviour. Some central banks even keep an 'operational events database' (eg the Bank of Canada), including information on participants' reactions. But it is difficult to generalise banks' behaviour, because disruptions have different causes and effects and behaviour can change over time.

One of the most interesting empirical studies of payment system participants' behaviour following a disruption, by Conover and Amanuel (2005), was presented at the Bank of Finland's simulator seminar in 2005. In that work, the authors analysed real failures in the US interbank payment system Fedwire, defined as abnormally long periods of time between payments from individual participants. It seems that, during such an event, [3] the value of payments received by the outage bank is down by 15.3%, while the number of payments is

---

[3] This excludes the period between 11 to 14 September 2001, ie an exogenous event clearly not related to operational disruptions at specific participants.

not significantly affected. Payments also seem to stay longer in the queue as the average delay of payments to the outage bank increases by 3 minutes.[4] These findings would indicate that the counterparties of the failing institution tend to retain their payments to that bank, while payments to other banks are not voluntarily affected. Such conclusions are in line with the reaction function proposed by McAndrews (2002), which implies that banks send part of their payments in reaction to payment receipts.[5] In case of a disruption, counterparties would behave strategically and retain their payments to the failing bank. Reaction functions will be further discussed in 8.2.3.

8.2.2 Determinants of counterparties' reaction

When considering whether to send or retain payments, banks can be expected to take into account the trade-off between settlement delay and intraday liquidity costs. This trade-off has been used to explain the choice participants face when deciding to send their payments relatively early or relatively lately during the day (Bech and Garratt, 2003). When choosing to pay early during the day (ie before having received any payments themselves), system participants incur liquidity costs related to the funding of these payments. When choosing to delay their outgoing payments, they incur reputation risk. However, when determining the course of action vis-à-vis a failing participant, liquidity costs can be expected to carry more weight than reputational considerations. Therefore, in the simulations and in the subsequent calculations of the costs related to an operational failure, I have assumed the costs of delaying payments to the stricken bank to be negligible.

How rapidly banks will react and stop sending payments to the stricken bank is likely to be determined by the rapidity with which they can identify a disruption at one of their counterparties. This will be influenced by the structure of the domestic banking market (eg the overall number of banks and the relationships between banks) and by factors related to payment system activity (eg the structure and regularity of payment flows and the modes of participation in the system). Some factors might have conflicting effects. For example,

---

[4] In addition to that, it would be very interesting to know how long participants tend to wait before delaying their payments to the outage bank and how this reaction changes for longer disruptions.

[5] Received payments make up for an important part of banks' funding liquidity in payment systems, in addition to reserves and intraday credit.

with direct participation, a more complete diffusion of information on a disruption could be achieved more rapidly as more banks would be directly exposed to the failing participant. However, an increase in the number of direct participants could lead to increased informational ’noise’. In a tiered payment system, all direct participants can be expected to be informed more rapidly of a disruption at one of their peers, while in a direct participation model the most active counterparties (possibly, but not necessarily, the same group of banks) would be up-to-date first.

8.2.3 Direct and indirect consequences

The analysis of reaction curves can be used to illustrate the effect of counterparties’ responses to a single bank failure on that bank and on other institutions.

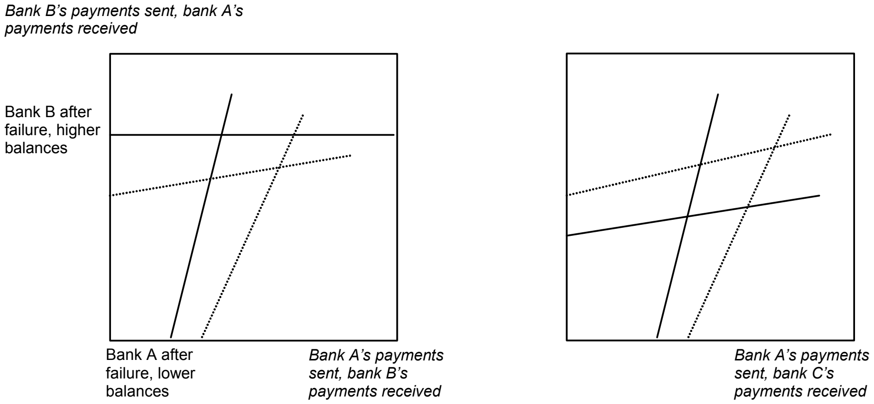
Figure 8.1 shows how reaction functions can move in response to a disruption at a payment system participant. The left-hand side of the figure is borrowed from McAndrews (2002).[6] It contains two sets of reaction functions. The dashed set shows the reaction functions of banks A and B under normal conditions. For bank A, this function can be expressed as $P_t^A = a + bR_t^a + \varepsilon_t$, where bank A’s payments in time t equal bank A’s autonomous willingness to send payments (a), its aggregate receipts and an error term. b is the slope of the reaction function. The autonomous willingness to pay is related to each bank’s reserves at the central bank. Under normal conditions, the slopes are positive, and payments are coordinated. When bank B proves unable to send payments, the slope of its reaction function takes value zero, receipts are not followed by payments. Bank A’s reserves decline as a consequence, as does its autonomous willingness to pay and the slope of its reaction function. This is shown by the solid lines. A similar fate would await other counterparties of bank B.

The right-hand side of Figure 8.1 introduces another participant, bank C, which has been affected by bank B’s failure in a similar way to bank A. These two healthy banks now have lower available reserves at the central bank, leading to an overall decline in their autonomous willingness to pay and in the slope of their reaction curves.

---

[6] See McAndrews (2002), p. 71. Here the figure is inverted as this paper refers to the disrupted bank as ‘bank B’.

Figure 8.1



In practice, if we assume that payment system participants react by delaying or cancelling all payments to bank B, the autonomous willingness to pay needs not be affected. Banks that are net debtors to bank B could control their exposures and prevent their balances at the central bank from declining by cancelling payments to B. Nonetheless, banks that would be net creditors would miss out on expected funds; their expected reserves would always be affected by B's failure. Such banks would therefore likely retain non-urgent payments to other participants in order to prevent important payments from remaining unsettled. Some important payments might indeed involve not only reputational costs but also costs caused by the de facto roll-over of the underlying deals.

Such third-round effects have not been included in the simulations. This paper only presents the results of simulations of first-order reactions to a counterparty's failure to pay. As it also works with historical data – be it adapted to fit various response scenarios – it can also be subject to a form of Lucas critique as not all behaviours in the reaction chain are modelled. This paper did not assess the implications of certain behavioural rules (mainly the use of exposure limits) under normal conditions either. Internal limits applied by banks would indeed have the potential to prevent some payments from being sent to the system at all, although we might expect banks to adjust their limits or change their payment flows in this situation.[7] As to centralised limits in the payment system, provided they are indeed used by banks, they might delay settlements. If such delays happen regularly, they might be harmful for banks' reputation and, by increasing the overall settlement delay of the payment system, they might negatively affect its resilience to operational events. Centralised limits will not be discussed further in this paper.

8.3 Methodology for the simulations

8.3.1 Replicating the Dutch large-value payment system TOP

Analysing the effects of a disruption on a payment system with the help of the simulator necessitates, first, to replicate the functioning of the given payment system under normal conditions. This is done by calibrating the algorithms and parameters defining payment system design in the simulator, with a view to obtaining the best possible fit between the pattern of payment flows in the resulting benchmark simulation and in the historical settlement data.

TOP, the Dutch payment system, is relatively easy to approximate with the simulator: it is a simple RTGS system, with queues and priorities, but no liquidity-saving mechanisms. TOP is part of the pan-European payment system TARGET and settled over EUR 30 trillion in 2005. The system has 100 participants, with most banks located in the Netherlands participating directly in the system. The banking sector in the Netherlands is quite concentrated, with four major banks accounting for most of the traffic in TOP. The large bank failing to send payments in the simulations accounts for about one fifth of TOP's volume and value. These large banks tend to have branches in most if not all other EU countries and therefore participate in TARGET via numerous channels. These bank branches also make liquidity transfers between each other.

---

[7] A quick glance at TOP participants' historical net debit positions in June 2004 vis-à-vis the stricken bank shows that, on average, 1.7 participants saw their net debit position at the end of the day exceed 25% of their capital; nearly 3 saw it exceed 10%; and 3.5 saw it exceed 5%. As explained in chapter 2.3, these are the levels of limits used for the simulation exercises. This means that, under normal conditions, such limits would have led these banks to retain some of their payments to that bank.

Simulations with historical payments data serve as a benchmark against which to assess the impact of disruptions and the effectiveness of counterparties' reaction.[8] However, this could only be done for banks located in the Netherlands as the level of reserves and collateral facilities available to banks located in other countries and participating in TOP via TARGET were not known. Furthermore, at the time at which data was gathered and simulations carried out, the only information available concerned the country from which or to which payments were made. The European counterparties that were branches of the failing bank could not be identified. This could be of importance for simulating a reaction from the stricken bank's counterparties, as counterparties belonging to the same group are likely to have more rapid and more thorough information on the event and therefore exhibit different reaction patterns than other banks. They are also less likely to retain payments from a perspective of exposure control. The failing institution in our simulations does have many branches in other EU countries.

8.3.2 Data used for the simulations

The data used for the simulations is from the month of June 2004 (22 days). That month displayed a volume and value of transactions that were close to the daily average for that year, with a slightly higher number of transactions (24.4 thousand a day, compared to an average of 23.5 thousand) and a slightly lower value (EUR 160 billion a day compared to 164 billion). In that year June offered the additional advantage of not having any special days, holidays or American bank holidays. Many smaller European banks are closed on holidays, which leads to a reduced activity in TOP. US bank holidays also tend to have a significant impact on payments in the Netherlands as most Dutch banks manage their liquidity by euro-dollar swaps.[9]

---

[8] In other types of work, benchmark simulations have also been used to compare different system designs and to assess the efficiency of a system's design compared to other possible designs.

[9] This impact is mainly felt on the day of the US bank holiday itself, with a decrease of 35% and 23% in the value and number of transactions respectively. On the day after the US bank holidays both the value and number of transactions increase by 23% and 17% (van Oord and Lin, 2005).

8.3.3 Benchmark simulations and three scenarios

A total of 242 simulations were carried out: 66 for the benchmark, the higher and lower bounds of liquidity and the worst-case failure scenario, with a disruption occurring at start-of-day;[10] 110 for Scenario 2, with the non-failing participants reacting within 10, 30 minutes, 1, 2 or 4 hours; and finally 66 for Scenario 3, with participants reacting when their individual exposure reach three different thresholds.

In the worst-case failure scenario (Scenario 1), payments from a major TOP participant to other banks are removed from the data as of start-of-day. The stricken bank does not recover within the day, which means that the disruption lasts for the remainder of the day. Although this would be an exceptional event, it is not totally impossible and the impact of such a disruption would certainly be significant. In 2003 one of the major European banks, accounting for slightly less than 10% of the overall TARGET volume, failed to send payments between 11:30 and 19:30 on a single day. Its counterparties were expecting to receive EUR 40 billion on that day.[11] In comparison, the stricken bank in the simulations accounted for slightly more than 20% of all TOP transactions in volume and slightly less than 20% in value.

Scenarios 2 and 3 introduce a reaction from other payment system participants. In Scenario 2, banks cease their payments to the stricken bank in order to prevent its accounts from acting as a liquidity sink. That way they are able to reduce the impact of the operational disruption on the system. A reaction time of 10, 30 minutes, 1, 2 or 4 hours is simulated. The 10-minute interval is based on work done by Bedford et al (2005) for the UK payment system CHAPS. Moving away from this reaction time allows us to assess the influence of a reaction delay on the impact of the disruption. In Scenario 3, counterparties' reaction is individually-tailored as they stop sending payments once their exposure to the failing bank (the net amounts paid) reach a certain percentage of their regulatory capital, namely 25%, 10% or 5%. 25% of regulatory capital corresponds to the maximum allowed exposure toward any individual counterparty. Under what is called the 'large positions rule' (*grote postenregeling*),

---

[10] An additional 242 simulations were carried out for the failure scenario, varying the time at which the disruption occurs. But as this is of little influence on the analysis of payment system participants' reaction to a counterparty failure, these will not be discussed in this paper. For detailed information on the time-dependency of the impact of an operational disruption, see Ledrut (2007).

[11] As reported by a former TOP operations manager.

banks are not allowed to exceed this threshold. Although this rule applies to overnight and longer exposures, I have used it as a maximum benchmark for intraday exposures in this exercise.[12] Only 8 banks send payments to the stricken bank (sometimes significantly) in excess of 25% of their regulatory capital. Banks are furthermore obliged to report any (overnight) exposure exceeding 3% of their regulatory capital (see Lelyveld and Liedorp, 2004). For intraday liquidity, I have set the lowest limit at 5%; when carrying out this exercise, lowering that threshold to 3% did not seem to add meaningful information as most payments remaining under the 5% threshold are already sent by a few banks with such high levels of regulatory capital that lowering the threshold to 3% would not generate different results. See Table 8.1 for a summary of these different scenarios.

Table 8.1 **Summary of the different scenarios**

	Failure major participant	Reaction counterparties	Variations on reaction	Number of simulations
Benchmark	No			22 days
Scenario 1	Yes	No	–	22 days
Scenario 2	Yes, at 12:00	Yes, timed stop sending	10 minutes 30 minutes 1 hour 2 hours 4 hours	22 days * 5 variations = 110
Scenario 3	Yes, at 12:00	Yes, stop sending after exposure reaches % of regulatory capital	25% 10% 5%	22 days * 3 variations = 66

The benchmark simulation, carried out with historical data, was also run with a view to assess the upper and lower bounds of liquidity that banks will need in order to settle all their payments (see Leinonen and

---

[12] When setting their intraday limits, banks do not seem to take into account any additional overnight lending to the institutions concerned. Such overnight lending does not need to be subtracted from the maximum allowed exposure in order to calculate a threshold for intraday lending.

Soramäki, 2005, for a more thorough explanation of the concept of liquidity bounds).[13]

In June 2004 in TOP, the upper bound of liquidity, which would have been needed by all payment system participants to allow all payments to settle immediately, amounted to EUR 50.18 billion on average and to nearly EUR 74 billion at a maximum. The upper bound of liquidity needed by banks only amounted to slightly more than half these figures. The lower bound of liquidity, which would have been enough for all payments to settle at end-of-day, was EUR 10.78 billion.[14] The liquidity available de facto in the system to banks[15] for the same month was EUR 6.7 billion in start-of-day balances and EUR 53.85 billion in deposited collateral (minus haircuts), ie a total of more than EUR 60 billion. This is roughly twice the upper bound of liquidity for banks only, which suggests that banks' endowment in liquid funds and collateral is very comfortable for the payments that are scheduled to be settled on the days of that month. However, the distribution of liquidity among banks is also important for the system to function well. In theory, as reserves are costly and collateral has an opportunity cost, banks would be tempted to reduce their liquid holdings at the central bank down to the level at which they would be able to settle all their payments within an acceptable timeframe. As we will see later, however, other factors play a role in the distribution of payment system participants' liquidity. In the simulations, this distribution is important as it influences the queuing and possible rejection of payments. It is implicitly assumed that there is no interbank market where banks with excess reserves could trade with banks lacking some necessary funds.

---

[13] These bounds of liquidity can be computed by allowing banks, starting the day without any balances on their accounts, to draw on unlimited amounts of intraday credit. All payments will thus settle immediately and queues will be empty. The sum of all the *lowest negative balances* on banks' accounts during the day is the upper bound of liquidity at system level. The sum of all *negative* net balances at the end of the day is the lower bound of liquidity at system level; it would be sufficient to allow participants to settle all their obligations in a DNS system.

[14] These are theoretical lower and higher bounds as the system is a 'pure' RTGS system. No consideration is taken of time-critical payments which absolutely need to be settled at certain times during the day. This includes payments not originated or received by banks, which may be sent from accounts that are not bound by a maximum overdraft facility, eg the accounts of the central bank itself.

[15] By banks is meant banks located in the Netherlands, which need to rely on either balances on their accounts at De Nederlandsche Bank or on collateral deposited at the central bank. This does not include banks located in other European countries which send payments to TOP through TARGET.

On the contrary to a number of other papers which estimate the relationship between the liquidity level and the impact of a disruption (Bedford et al, 2005, and others), all simulations in this paper are based on the historical amounts of reserves and collateral at the central bank, so as to assess the potential effect of a disruption under the conditions that held in June 2004. These conditions were relatively comfortable, which suggests that a disruption occurring under more stressed liquidity conditions would undoubtedly have a stronger impact on the system, assuming the distribution of reserves and collateral is kept constant.

8.3.4 Quantifying the impact of a disruption

The impact of a disruption can be measured by the number of unsettled payments sent by healthy system participants (secondary-round effect), by the settlement delays incurred by payments submitted to the system and by the cost of additional credit that healthy participants need to borrow overnight. These measures allow a comparison of the different scenarios.

Unsettled payments / secondary-round effects

The first-round effects of an operational disruption stem directly from the payments not sent by the failing institution while the secondary-round effects can be measured as the number and value of payments that other banks will not be able to make due to a lack of funds caused by the fact that none of the payments from the stricken bank were received. They indicate the potential for the disruption to spread out and lead to additional unsettled payments. Such unsettled payments will be used in this paper as the prime measure of system-wide risk.

Settlement delays

When payments cannot be settled due to a lack of funds, they remain queued. The longer they remain in the queue, the longer settlement is delayed. From a system perspective, the higher the number (and value) of payments that cannot be settled immediately, and the longer their stay in the queue, the higher the settlement delay. RTGS systems with ample available liquidity and no queuing will show no settlement delay at all. The introduction of the possibility of queuing and reduction of the funds available increases settlement delay, with DNS systems that settle all payments at the end of the day exhibiting the highest settlement delay. Such an aggregate settlement delay can be expressed by way of a settlement delay indicator (Koponen and Soramäki, 1998)

$$\rho = \frac{\sum_{i=1}^{T} Q_i}{\sum_{t=1}^{T} \sum_{i=1}^{t} V_i}$$

where, in the numerator, the value of payments in the queue is summed over each minute of the day and, in the denominator, the *cumulative* value of all sent payments is summed over each minute of the day. Settlement delay can be calculated for a portion of time (up to a whole day) for some or all payment system participants. It is between 0 and 1, with 0 indicating that all payments are settled immediately (as in a pure RTGS system with no queues and ample liquidity) and 1 that all remain delayed until the end of the period under consideration (this will typically be exhibited by a DNS system). Settlement delay means settlement risk. As long as queued payments have not been settled with finality, the recipient remains exposed to credit risk (which would materialise if the sender proves unable to fund its account), to liquidity risk (if the payment arrives later than the recipient would have expected given its own funding needs) and to operational risk (a disruption at either the sending institution or the system operator can prevent the payment from being settled in due time).

The cost of an operational failure

Given the high cost of delaying payments to the next day (eg due to the roll-over of some deals), I assumed for this exercise that payment system participants would fund their outgoing payments up to the maximum during the day with central bank intraday credit. At the end of the day, as the disruption would not be resolved, remaining negative positions would need to be funded overnight, either on the overnight interbank market or at the central bank, making use of what is called the marginal lending facility. The additional liquidity that banks incorporated in the Netherlands would need at the end of the day, compared to the benchmark simulation, is equal to the difference between the sum of all negative values on central bank accounts at the end of the day (resulting from the simulations) and the use of the marginal lending facility in the month of June 2004, which averaged EUR 159.76 thousand. For the same month, the interest rate of the marginal lending facility was 3% pa, while the EONIA rate, the market rate, amounted to 2.03% pa. Although I have calculated the costs of additional liquidity in accordance with these rates, a major disruption could influence the market rate and push it upward as demand for overnight liquidity would significantly rise compared to the available liquidity in the market. However, such an effect would be dampened by the width of the euro market, as the liquidity needs of Dutch banks might be fulfilled by other euro area institutions. This effect can be expected to be stronger in a small closed market where all participants need to weather the same disruption in their main payment system.

8.4 Simulating the influence of banks' reaction on the impact of a disruption

8.4.1 Counterparties do not react

Scenario 1 assumes that one of the major payment system participants in the Dutch interbank payment system TOP is unable to send payment orders as from 07:00, the opening time of the TARGET system. The default lasts for the rest of the day. If the operational inability to send payments were to be solved within a short period of time, the payment system would only be affected during the disruption; queues would arise for the duration of the disruption and disappear rapidly after the failing participant has resumed its operations. At the end of the day no unsettled payments would remain. In this exercise, unsettled payments will be used in order to assess the impact of the disruption and participants' ability to weather a counterparty's failure to pay.

Results

The simulations show that, on average over the month of June 2004, were a major Dutch bank to retain all its payments as from the opening time of the system, about 16 payments from other participants would remain unsettled as a consequence of liquidity shortages. This would represent more than EUR 1 billion. Additional simulations have shown that, as expected, the secondary-round effects decrease when the disruption occurs later during the day (see Ledrut, 2007, for a discussion of the time-dependency of the impact of a disruption). The strength of the secondary-round effects can substantially differ from one day to another, varying between 3 and 45 unsettled payments, and a value between about EUR 40 million and slightly less than EUR 2.2 billion (see Table 8.2). An operational disruption at a major payment system participant also clearly increases settlement delay from 0.08 in the benchmark simulation to 0.13 in the worst-case failure scenario.

Table 8.2 **Impact of an operational disruption (excluding failing participant's payments) on unsettled payments and settlement delay**

Number unsettled				Value unsettled (EUR million)				Delay ind.
Av	Min	Max	St. dev	Av	Min	Max	St. dev	Av
16.62	3.00	45.00	11.38	1,075.18	38.95	2,196.00	520.36	0.13

In June 2004, counterparties to the failing bank were expected to receive, on average, about 3.4 thousand payments and EUR 30 billion on each day. Following an operational failure, the value trapped on the stricken bank's accounts (due to the liquidity sink effect) would reach about 5.5 times its normal end-of-day value. On average, the maximum level of intraday borrowing that would remain outstanding at the end of the day would be more than EUR 1.5 billion. The maximum cost of additional overnight funding would be about EUR 200 thousand for the marginal lending facility and nearly EUR 140 thousand for interbank lending.

8.4.2 Timed stop-sending

There is some observed evidence that payment system participants would react within a certain timeframe following an operation disruption at one of their counterparties. A paper by Bedford et al (2005) is based on such evidence for CHAPS, the UK payment system.[16] Because their work incorporates the effect of counterparties' timed reaction, the worst-case moment for a disruption to occur is not

---

[16] 'Anecdotal evidence from CHAPS Sterling suggests that the time-lag between an individual bank experiencing an operational failure and the flow of payments to that bank slowing significantly is typically of the order ot ten minutes'.

at the beginning of the day but at a moment at when the ‘virtual credit balance’ (for each minute, the sum of the stricken bank’s reserves and its expected payments in the next ten minutes) is at its highest and a lot of payments still need to be made. Such a virtual credit balance gives an indication of the potential for the stricken bank to develop into a liquidity sink. At the moment when the operational disruption is meant to occur, the stricken bank is due to send or receive 46.000 payments with a total value of GBP 45.7 billion. As a share of the total value processed by CHAPS Sterling, this is in the same order of magnitude as the bank considered in this paper, although the UK bank accounts for a much higher relative volume of payments. Ten minutes after the disruption, all payments to the failing participant cease. Varying the quantity of liquidity available, Bedford et al find that the secondary-round effects of a disruption are only significant at levels of liquidity which are much below the liquidity available in the system.

With regard to TOP, observations by payment system operators suggest that Dutch banks would be less prompt to react than UK banks, and that, for a disruption occuring early in the morning, they would probably not delay payment orders before noon. Foreign banks are expected to be the first to stop sending payments. In our example, given the reputation of the failing bank and its habit of sending payments early in the day, and therefore of supplying other banks with liquidity, domestic banks would probably not rapidly cancel transactions. We will see later in this paper that differentiating between domestic and foreign banks can be justified by the fact that all banks wish to limit their relative exposure to the failing institution.

Scenario 2 is based partly on the work by Bedford et al as I simulate that participants stop sending payments to the stricken bank some time after the start of the disruption. However, it focuses on two new questions: By how much can participants reduce the impact of a disruption? And is this a function of the reaction time? The stricken banks’ counterparties stop to send payments 10, 30 minutes, 1, 2 or 4 hours after the disruption. Note that such a reaction pattern assumes perfect information in the market about the occurrence and timing of a disruption. Even if, de facto, not all banks react simultaneously, as some did not schedule payments to the stricken bank within the reaction period, all banks retain their payments as from the same point in time.

Results

When participants do not react following a disruption and continue to send payments to the stricken bank, about 16 additional payments get stalled due to liquidity shortages. When participants do react, the secondary-round effect decreases to an average of about 7 rejected payments at the end of the day for a reaction time of 4 hours (a reaction time in concordance with payment system operators' observations for TOP). It thus seems that even a relatively long reaction time would be effective in limiting the secondary-round effects of a failure. Counter-intuitively, reducing the reaction time from 4 to 2 hours or less has a much more limited effect. Indeed, for reaction times ranging between 10 minutes and 2 hours, the secondary-round effects are quasi-similar, with about 5 stalled payments.

Reacting also appears to be quite efficient in reducing the costs of additional liquidity. Where such costs reach about EUR 202 thousand and slightly less than EUR 140 thousand for central bank and money market lending respectively (for an additional EUR 1.7 billion) in the worst-case scenario, they can be reduced to less than half these amounts if banks react within ten minutes (see Table 8.3). But no matter how fast they react, banks appear unable to fully eliminate the need to rely on additional liquidity following a disruption; even when they stop sending payments to the failing institution within 10 minutes after a disruption, they still incur costs of about EUR 100 thousand or EUR 67 thousand for central bank and money market lending respectively. Between 10 and 30 minutes, the costs of additional liquidity increase, as expected, while the unsettled payments in the secondary-round effects remain constant: for these levels, banks can apparently continue to incur higher overdrafts; their payments are only rejected for a longer reaction time. But between 30 minutes and 1 hour the additional liquidity needed is reduced as more payments remain unsettled. The difference between days is relatively important (see Table 8.4). On the day with the lowest second-round effects, all payments settle as from a reaction time of 2 hours while on the day with the highest effects even a reaction time of ten minutes does not reduce the number of unsettled payments below 12.[17] Changes in settlement delays will be discussed separately at the end of 8.4.3.

Table 8.3 **The impact of timed stop-sending on secondary round effects and liquidity costs**

Time before participants react	Unsettled payments		Additional liquidity		
	Number	Value (EUR million)	Value (EUR million)	Costs of overnight overdraft (EUR thousands)	
				EONIA	marginal lending
10 min	4.86	303.37	824.49	66.95	98.94
30 min	4.86	303.37	824.71	66.97	98.96
1 hour	4.90	319.23	823.83	66.89	98.86
2 hours	5.00	329.65	901.77	73.22	108.21
4 hours	7.43	510.46	1400.30	113.70	168.04
Unlimited (> 11 hours)	16.62	1075.18	1687.96	137.06	202.56

Table 8.4 **Daily variation of results for timed stop-sending**

Reaction time	Number unsettled				Value unsettled (EUR millions)			
	Av	Min	Max	St. dev	Av	Min	Max	St. dev
10 mn	4.86	0.00	12.00	3.55	303.37	0.00	1,264.98	322.64
30 mn	4.86	0.00	12.00	3.55	303.37	0.00	1,264.98	322.64
1 h	4.90	0.00	13.00	3.66	319.23	0.00	1,264.98	352.62
2h	5.00	0.00	13.00	3.63	329.65	0.00	1,264.98	345.75
4h	7.43	2.00	32.00	6.69	510.46	20.28	1,353.02	347.24

In view of the described observations with regard to the behaviour of TOP participants, it would be interesting to differentiate between participants (for example, letting only a fraction of participants react or different types of participants react at different time intervals, with smaller or foreign banks reacting immediately and bigger or domestic banks being less rapid in their reaction). It would also be possible to simulate a timed stop-sending where the time lag is calculated not as from the occurrence of the operational disruption – which assumes perfect information in the market about the operational health of any given bank – but as from the moment at which each bank expects to

---

[17] On a certain day, an interesting situation can be witnessed as participants' reaction leads to an increase in the number of unsettled payments compared to no reaction. On that day, the payments sent by the defaulting participant before disruption, but waiting in the queue for incoming payments, can remain unsettled as these expected incoming payments do not arrive.

receive payments from the failing institution – which assumes that banks are only informed about an operational failure when they miss payments due to them. However, instead of differentiating in terms of participants' reaction time, I have worked on another scenario, where differentiation was a natural consequence of the assumption: participants react when their exposure reaches a certain threshold.

8.4.3 Exposure control

The hypothesis underpinning Scenario 3 is that once the stricken bank's counterparties' liquidity is getting scarcer, they block outgoing payments to that bank. Doing that, they rely on their internal limits. Their real limits are not known. For this exercise, they have been approximated by setting bilateral credit limits vis-à-vis the failing bank at 25%, 10% and 5% of counterparties' regulatory capital.

A practical difficulty related to this exercise concerned bank branches operating under a European passport; these branches do not need to maintain separate regulatory capital for their Dutch activities. There were 7 such branches participating in TOP in June 2004. They accounted for about 2.6% of all payments sent by banks in volume and 3.8% in value. I have assumed that their payment activity in the Netherlands was indicative of their strength in terms of capital in that country and have created a peer group for each bank, made up of 4 to 6 Dutch banks with comparable payment patterns in TOP. I have then taken the average regulatory capital of the Dutch banks in the peer group as a proxy for the regulatory capital of the foreign banks. Although this may not be totally correct in theory, as the solvability of these banks is related to the strength of their mother institution; from a payment system perspective it seems reasonable to assume that a bank's ability to manage its payment activities in a foreign country is related to its payment flows. Discussions with banking supervisors and payment system operators confirm that view.

This exercise bears resemblance to the work of Mazars and Woelfel (2005), who simulate the technical failure of a major participant in the French PNS system at start-of-day. Up to 10% of payments get rejected in the second round. The authors try out different bilateral limits – these limits are managed centrally in the system – and conclude that 'the consequences of a technical default could be greatly reduced if the participants set their bilateral sender limits at a lower level than that currently observed and if they reacted rapidly to information indicating a technical default by reducing their bilateral limits with the defaulting participant'. By reducing the level of exposure from 25% to 10% and 5%, I have tested if this conclusion also holds for TOP.

Results

In TOP, compared to no threshold at all, the existence of a threshold makes a difference as even a threshold of 25% reduces the secondary-round effects nearly by half to slightly more than 9 unsettled payments, with a value of EUR 850 million (see table 8.5). But there too reducing the level of that threshold to 10% and even 5% does not bring about the reduction of rejected payments that would be expected. Even when payment system participants do not allow for an exposure higher than 5% of their regulatory capital, the secondary-round effects appear to be stronger than when they only stop sending payments four hours after the start of the disruption. But for a similar level of secondary-round failures, the additional liquidity needed is lower, with EUR 1.1 billion for the 5% threshold and about 1.5 billion for the 4-hour stop-sending scenario. However, the large variations between days (see Table 8.6,), with a minimum of 2 and a maximum of 20 unsettled payments and a high standard deviation at a 5% threshold, as well as the relatively small number of simulations carried out limit the statistical significance of the results.

This exercise shows that foreign banks (mainly subsidiaries of foreign banks that do need to hold capital in the Netherlands) and smaller banks will tend to delay payments earlier as their regulatory capital is relatively low. With a threshold of 5%, most of the payments of such foreign participants to the stricken bank remain unsent as the value of these payments would amount to more than 5% of their regulatory capital.

Table 8.5 **The impact of exposure control on secondary round effects and liquidity costs**

Exposure limited to (in % of regulatory capital)	Unsettled payments		Additional liquidity		
	Number	Value (EUR million)	Value (EUR million)	Costs of overnight overdraft (EUR thousands)	
				EONIA	marginal lending
5%	8.71	756.96	1137.35	92.35	136.48
10%	9.00	783.71	1286.13	104.43	154.34
25%	9.38	850.42	1522.94	123.66	182.75
Unlimited	16.62	1075.18	1687.96	137.06	202.56

Table 8.6 **Daily variation of results for exposure control**

Reaction threshold	Number unsettled				Value unsettled (EUR millions)			
	Av	Min	Max	St. dev	Av	Min	Max	St. dev
5%	8.71	2.00	20.00	4.83	756.96	7.62	1,560.84	457.45
10%	9.00	2.00	20.00	4.65	783.71	7.62	1,560.84	435.19
25%	9.38	3.00	20.00	4.77	850.42	41.25	1,886.70	467.64

Settlement delays

Compared to the simulations carried out under Scenario 1, settlement delays increase when the failing bank's counterparties stop sending payments to that bank under Scenarios 2 and 3. The indicator has a value of 0.14 for all reaction times under Scenario 2, except for the longest (4 hours), for which it takes the value 0.12. Counter-intuitively, a longer reaction time *reduces* settlement delay. The delay indicators for the third scenario confirm these findings, with longer delays (a value of 0.16) under all sub-scenarios. As counterparties' reactions are expected to limit the liquidity strains in the system, these results are somewhat surprising. But they may be explained by the way the indicator is calculated. Indeed, the number (and value) of payments decreases very steeply when participants react while the values in the queue are much less affected. For example, the value of queued payments declines by 21% in the 5% exposure control scenario, while the overall value of all sent payments is reduced by 32% as a consequence of participants' reaction.

8.4.4 Limitations of these behavioural simulations

The simulation of payment systems participants' behaviour makes it necessary to rely on a number of debatable assumptions. Assuming all participants react in an identical matter is one of them. For example, the nationality of the counterparty would almost certainly play a role, as Dutch banks are more likely to continue sending payments to each other for a longer time after a disruption than would foreign banks. This is mainly due to foreign banks' lower regulatory capital, but it could also find its roots in a different relationship with the failing institution. Such differences can also be found among Dutch banks, with close cooperative relationships leading to less retained payments and vice versa. The size of the bank could also matter.

More specifically for the exercise with exposure reduction, the relation between regulatory capital and payments is not straightforward. Some banks have a high capital ratio and make few payments and vice versa. There are also some special banks which are close to single purpose institutions and make many payments of a very high value in spite of a relatively low capital ratio. The construction of the proxy for foreign banks could also be discussed as it would not reflect their solvability, which would depend on their mother institution.

Not only have I assumed identical behaviour among payment systems participants, but no differentiation has been made among payments sent to the stricken bank. However, the type of payment, its value and its originator could all influence the sending bank's likeliness to retain it in the face of a disruption. Some payments need to be made on a given day in order to discharge an obligation incurred for a certain amount of time, such as the repayment of money market loans. These payments are not likely to be retained longer than the end of the day in order to avoid overnight penalties. Other payments may be 'coupled' to payments that the failing bank is about to send; these payments would only be released after receipt of the incoming payment. The value of the payment also matters as small value payments, which account for an important number of transactions settled by large-value payment systems,[18] would be most likely to pass in spite of limits. Very large payments would be the first to be delayed. Note that this does happen under Scenario 3 as limits happen to be often reached once a very large payment needs to be sent. With regard to the originator, payments for important customers are more likely to be sent independently of the operational health of the receiving bank.

Finally, the order of payments is only of marginal importance in Scenario 2 but can play a significant role in Scenario 3. Due to the construction of the reaction function, all payments above the threshold remain unsent. This means that very large value payments that exceed the threshold automatically get blocked, regardless of their timing. In practice, some banks send such payments very early in the day to the failing institution, before receiving any payments from that bank. These payments would unlikely be stopped at start-of-day. However, in the simulations, when very large payments precede smaller ones, both are blocked. This is not true when the smaller payments come first, so the rule can have different effects on different days

---

[18] As much as 25% of all TARGET transactions have a value of less than EUR 1,250.

(depending on the order of, among others, recurring payments). The influence of these variations on the results is however likely to be minor as the large payments are the ones that matter.

8.5 Conclusion

This paper presents simulations carried our with the payment and settlement system simulator developed by the Bank of Finland. It shows how participants' behaviour in the face of a disruption can mitigate the impact of such a disruption, both for themselves – in terms of the costs of fulfilling their own payment obligations – and for the system as a whole – in terms of the payments rejected following a liquidity shortage.

It seems that the secondary effects of a disruption at a major TOP participant are relatively limited at current liquidity levels. Liquidity levels that are higher than the upper bounds of liquidity, which correspond to the liquidity needed to settle all payments immediately, are helpful to reduce the impact of a disruption – but do not guarantee that such a disruption will be weathered without further payment rejections as the distribution of this liquidity in the system makes a difference (abstracting from an interbank money market). The cost of such a disruption could be large for some healthy banks as these banks normally rely on incoming payments to fund part of their outgoing transactions. These would then need to be funded with intraday credit and with overnight credit if the stricken bank were to prove unable to resume operations at the end of the day.

Payment system participants can control their exposure to their counterparties by way of internal bilateral limits or by way of monitoring the activity of other participants and reacting to outages. Simulations seem to show that even a slow reaction (eg 4 hours) might be effective in reducing the secondary effects of a disruption. However, in some exceptional situations, participants' reactions could slightly increase the impact of the disruption as payments sent by the failing participant in advance of the disruption, and waiting in the queue, could remain unsettled due to lack of funding. Although participants' actions seem to be effective, they do not completely eliminate the secondary effects of a disruption or eliminate the costs related to additional overnight funding. Furthermore, simulation results seem to indicate that stringent bilateral limits effectively mitigate the individual costs of operational outages. A timed stop sending rule seems less able to limit additional liquidity costs but more effective in mitigating systemic effects. Even a reaction delay of 1 or 2 hours would substantially reduce the second-round impact of disruptions. In order to achieve this, central banks could share information about participants' outages so as to allow banks to respond timely – although such transparency might be incompatible with confidentiality obligations. Banks might also be encouraged to communicate in a timely manner about their operational difficulties.

These results are based on a relatively small sample, and the sometimes large variations between days further limit their statistical significance. In order to confirm these findings, additional simulations would need to be carried out over longer periods of time. It would also be interesting to see whether these conclusions hold for other large players in TOP and for other payment systems. The combination of limits and timed stop sending rules could also be tried out as some banks are likely to have relatively high counterparty limits intraday while still being capable of reacting to a detected disruption within a certain timeframe. Furthermore, the cost of intraday credit, not taken into account in this exercise (collateral being deposited ex ante at the central bank), could be used to assess the impact of a disruption in a system which offers such credit against a direct fee (eg Fedwire). Finally, the implications of the use of (centralised) limits by payment system participants under normal conditions could be assessed. Limits would indeed have the potential to delay settlements and might even prevent some payments from being settled at all, although we might expect banks to adjust their limits or change their payment flows if this risk were to materialise. Delays might lead to various costs for payment system participants (reputation costs but also an increased risk of payments not settling), which were assumed negligible for the purpose of this paper. Before encouraging central banks to introduce centralised limits and banks to actively use or reduce their limits, it might be useful to compare these costs to the costs imposed on participants by the failure of one of their largest counterparties, as calculated in this paper.

References

Amanuel, D – Conover, D (2005) **Operational Disruptions in Fedwire: Simulating Liquidity Needs and Understanding Counterparty response.** Presentation at the BoF-PSS Seminar, August 25.

Bedford, P – Millard, S – Yang, J (2005) **Analysis the impact of operational incidents in large-value payment systems: a simulation approach.** In Leinonen, H (ed.)

Koponen, R – Soramäki, K (1998) **Intraday Liquidity Needs in a Modern Interbank Payment System.** Bank of Finland Studies E:14.

Ledrut, E (2007) **Simulating retaliation in payment systems: can banks control their exposure to a failing participant?** DNB Working Paper.

Leinonen, H (ed.) (2005) **Liquidity, risks and speed in payment and settlement systems – a simulation approach.** Bank of Finland Studies.

van Lelyveld, I – Liedorp, F (2004) **Interbank contagion in the Dutch banking sector.** DNB Working Paper, July.

Lin, H – van Oord, A (2005) **Modelling inter- and intraday payment flows.** DNB Working Paper, December.

Mazars, E – Woelfel, G (2005) **Analysis, by simulation, of the impact of a technical default of a payment system participant.** In Leinonen, H (ed.)

McAndrews, J – Potter, S (2002) **Liquidity effects of the events of September 11, 2001.** FRBNY Economic Policy Review, November.

McVanel, D (2005) **The Impact of Unanticipated Defaults in Canada's Large-Value Transfer System.**

Chapter 9

The impact of unanticipated defaults in Canada's Large Value Transfer System

Darcey McVanel

9 The impact of unanticipated defaults in Canada’s Large Value Transfer System

Abstract

Canada’s Large Value Transfer System (LVTS) is designed to meet international risk-proofing standards at a minimum cost to participants in terms of collateral requirements. It does so, in part, through collateralised risk-sharing arrangements whereby participants may incur losses if another participant defaults. The LVTS is designed to be robust to defaults. Its rules, however, do not ensure that individual participants are robust to defaults. The author studies participants' robustness to default empirically by creating unanticipated defaults in LVTS and finds that all participants are able to withstand their loss allocations that result from the largest defaults she can create using actual LVTS data.

9.1 Motivation

Canada’s Large Value Transfer System (LVTS) is designed to meet international risk-proofing standards at a minimum cost to participants in terms of collateral requirements. It does so, in part, through collateralised risk-sharing arrangements whereby participants may incur losses if another participant defaults.[1,2]The system is designed so that there is sufficient collateral prepledged by participants to cover at least the largest possible payment obligation to the system. Therefore, the rules ensure that the system is robust to defaults. The system’s rules, however, do not ensure that individual participants are robust to defaults; it is up to participants to manage their own risks to make sure they can withstand potential losses that result from a default. In this

---

[1] A participant is deemed to be in default if it cannot meet its end-of-day net debit position.

[2] The Bank of Canada guarantees settlement in the extremely unlikely event that more than one participant defaults on a single day and the sum of the exposures exceeds participants’ prepledged collateral, so the system is robust to even multiple defaults in a single day. This provides for intraday finality of payments.

paper, I study participants' robustness empirically by creating unanticipated defaults in LVTS.

The system's rules create the incentive for participants to prudently manage their risks vis-à-vis other participants. The majority of LVTS transactions go through the survivors-pay component of LVTS, where participants' losses in the event of a default are governed by the bilateral credit limit (BCL) that they grant to the defaulter. The granting of BCLs, as well as their size, is completely voluntary. Given the possibility that another participant could default, participants have an incentive to set BCLs at a size that would create manageable losses for themselves if a default were to occur. Furthermore, in a situation where participants believe that another participant may be in danger of defaulting, it could be in their interest to reduce the BCLs they grant to this participant, to minimise their loss exposures.

In this paper, I generate a series of unanticipated defaults by individual participants to estimate whether surviving participants would be able to withstand their allocated losses.[3] I first estimate the frequency with which survivors must contribute to cover a defaulter's shortfall and the relative size of these loss allocations, and then assess the ability of participants to withstand these losses by calculating individual survivors' capital positions following a default.

Results are based on an eight-month (from March to October 2004) sample of LVTS transactions, collateral holdings and data on bilateral and multilateral credit limits, provided by the Canadian Payments Association. Each participant's maximum net debit position and the time at which it was incurred are found using the Bank of Finland Payment and Settlement Simulator ('the simulator'). Survivors' additional settlement obligations (ASOs) are calculated according to LVTS rules.

I find that the shortfalls resulting from the theoretical defaults are generally small, but there exists substantial heterogeneity in how often individual participants incur shortfalls, each participant's average shortfall size, and the size of shortfalls over different days. Large participants generally incur shortfalls that are much larger than those incurred by small participants. These factors create a large degree of variability in participants' loss allocations. In both absolute and relative terms, participants' loss allocations are generally small, but when losses are compared with assets and capital, small participants

---

[3] I believe that loss allocations would likely be larger when a default is unanticipated than anticipated since, as already described, participants that anticipate a default may have incentives to reduce the BCLs they grant to that participant.

take on relatively much more risk than large participants. Nevertheless, I find that all participants are robust to the defaults generated.

This paper contributes to the literature in two respects. First, most previous work has focused on losses to survivors in uncollateralised netting systems; I consider losses in a risk-proofed and collateralised system. Second, I apply the simulator to a default analysis, whereas most previous studies have focused on questions of liquidity usage or operational risk.[4]

This paper is organised as follows. Section 9.2 provides an overview of LVTS's risk controls and default-resolution procedures. Section 9.3 compares this study with the previous literature. Section 9.4 explains the procedure for generating defaults. Sections 9.5, 9.6, and 9.7 describe the findings, and the impact of key assumptions. Appendix 1 contains proofs of the efficacy of LVTS's risk controls and Appendix 2 shows the times of participants' maximum net debit positions.

9.2 LVTS framework

To understand the default and loss-allocation procedures used in this paper, it is useful to review the main concepts and risk controls within LVTS:

- LVTS is a real-time electronic payments system that provides certainty of settlement on a continuous basis for all payments that have passed the risk controls. It uses caps, collateral, loss-sharing arrangements, and a residual guarantee by the Bank of Canada to provide intraday finality and irrevocability of payments.
- LVTS is a collateralised, deferred net settlement system. Unlike in real-time gross settlement systems, in which settlement defaults cannot occur inside the system (because settlement of payments involves the immediate transfer of funds across the books of the settlement institution), settlement default is possible in LVTS.[5]

---

[4] See the Bank of Finland website (http://www.bof.fi/eng/3_rahoitusmarkkinat/3.4_Maksujarjestelmat/3.4.3_Kehittaminen/3.4.3.3_Bof-pss2/) for links to studies that use the simulator.

[5] However, the collateralised risk controls in the system and the residual Bank of Canada guarantee provide for certainty of settlement even where there are multiple defaults.

- LVTS consists of one fully defaulter-pays payment stream and one partially survivors-pay stream. In the partially survivors-pay stream, a participant is able to incur a larger net debit position than the collateral it pledges to support LVTS activity. Thus, survivors may have to contribute to cover losses in the event of a default. Approximately 88% of LVTS value goes through the partially survivors-pay payment stream.
- The risk controls are designed so that there will be sufficient aggregated collateral in the system to cover at least the largest net debit position possible, or, put differently, at least the default of the largest single net debtor.
- The system will always settle, because all participants contribute to a collateral pool and the Bank of Canada provides a residual guarantee that, in the unlikely event of multiple defaults on a single day, if survivors' prepledged collateral does not cover the defaulters' losses, the Bank will cover the difference.

9.2.1 Collateralisation

Participants first pledge collateral to the Bank of Canada to support their LVTS activity, and they then apportion parts of it to collateralise each of the defaulter-pays Tranche 1 (T1) and the partially survivors-pay Tranche 2 (T2).[6] Collateral pledged to the Bank by a participant but not apportioned to LVTS is referred to as excess collateral.

In T1, each participant, i, apportions collateral to cover its own obligations. Its maximum allowed net debit position, referred to as its T1 net debit cap ($T1NDC_i$), is set equal to the value of the collateral (minus haircuts) that it has pledged to cover these obligations ($C1_i$). Thus, each participant fully collateralises its own T1 obligations

$$C1_i = TINDC_i \tag{9.1}$$

In a default, this collateral would be used to cover the defaulter's position, so this stream is referred to as defaulter-pays.

In T2, participants determine how much exposure they are willing to take on vis-à-vis other participants and extend lines of credit accordingly. Each participant i must then apportion collateral ($C2_i$)

---

[6] Eligible LVTS collateral includes Bank of Canada funds and government and highly rated corporate bonds. The usable value of collateral is the market value of each security less a certain amount (a 'haircut'), to account for market risk.

equal to a percentage (θ) of the largest BCL it has extended to any other participant j ($\max_j(BCL_{ij})$).[7] This value is called the participant's maximum additional settlement obligation ($\max ASO_i$), which is the maximum amount that the participant will have to contribute if one or more participants to which it has granted a BCL defaults

$$C2_i = \max_j(BCL_{ij}) \cdot \theta = \max ASO_i \tag{9.2}$$

In the event of a default, the defaulter's own T1 and T2 collateral will be used first to settle its net debit position. If there is a shortfall, however, then survivors' collateral will be used to cover the defaulter's residual T2 obligations. Thus, although T2 is considered to be a survivors-pay tranche, it has a defaulter-pays element as well.

Each participant can incur a net bilateral debit position equal to the BCL that has been established for it by the grantor. As well as BCLs, each participant has a multilateral net debit cap. Each participant i's maximum permitted multilateral T2 net debit position, its T2 net debit cap ($T2NDC_i$), is set equal to the sum of the credit lines received from all participants, multiplied by the systemwide percentage

$$T2NDC_i = \sum_{j=1}^{N-1} BCL_{ji} \cdot \theta \tag{9.3}$$

where there are N LVTS participants.

9.2.2 Settlement

Throughout the day, individual payments that have passed the risk controls are netted, novated, and replaced by a net obligation to receive or pay funds.[8] At the end of the day, participants' T1 and T2 positions are combined to yield a final multilateral net position that they must settle. The Bank of Canada facilitates settlement by debiting the settlement accounts of the participants that are in a multilateral net debit (short) position and crediting the accounts of participants that are in a multilateral net credit (long) position. Through this settlement

---

[7] The percentage, referred to as the systemwide percentage, takes into account the effect of netting.

[8] The Bank for International Settlements (BIS) (2003) defines novation as the 'satisfaction and discharge of existing contractual obligations by means of their replacement by new obligations'.

process, net debtors discharge their credit obligations and net creditors receive Bank of Canada funds.

Solvent participants that are short funds at the end of the day in LVTS may trade with participants that are long funds to borrow the funds needed for settlement. As well, such participants may obtain the funds necessary to settle by taking a fully collateralised discretionary advance from the Bank of Canada (at the Bank Rate). Under this option, the participant pledges collateral to the Bank of Canada with a value equal to its deficit position at the close of LVTS, and the Bank of Canada credits its settlement account with the funds. The duration of this loan is one day (to be paid back by 6 pm the following day).

Participants are allowed to use all the collateral that they have apportioned in LVTS to cover their discretionary advance: namely, the collateral that they have apportioned to support their own T1 obligations and the collateral that they have apportioned to T2 to cover the lines of credit they have granted to others. As well, they may apportion their excess collateral in support of their discretionary advance.

9.2.3 Default

A participant is deemed to be in default if it cannot meet its end-of-day net debit position. A default can occur under two circumstances:

(i) The participant is in a net debit position at the end of the day and has insufficient collateral to cover this position; ie, it has a collateral shortfall.

(ii) The participant has been suspended from further participation in LVTS during the current LVTS cycle and has a net debit position that must be settled.[9] This will occur if a participant is closed by its regulator.

In the event of the default of any participant i, the Bank of Canada will seize the defaulter's apportioned collateral and grant a non-discretionary advance to participant i (NDA_i) equal to the lesser of (i) the absolute value of the participant's combined Tranche 1 and Tranche 2 multilateral net positions ($T1MNP_i$ and $T2MNP_i$,

---

[9] If a participant is suspended from further participation in LVTS, but is shut down with a positive position, it will not be declared in default, because it does not owe funds to the system.

respectively), less any funds the participant is holding in its settlement account at the Bank of Canada (SF_i), or (ii) the participant's apportioned collateral[10]

$$NDA_i = \min[(|T1MNP_i + T2MNP_i| - SF_i),(C1_i + C2_i)] \qquad (9.4)$$

In other words, the Bank of Canada will lend the lesser of the actual position that the participant must settle or the collateral the participant has apportioned to cover its position. For the latter case, survivors will be required to cover the shortfall.[11]

9.2.4 The ability of participants to generate a shortfall

A participant can incur a larger net debit position than the collateral it pledges for LVTS purposes. Note that participant i's maximum net debit position ($maxNDP_i$) is the sum of its T1 and T2 net debit caps[12]

$$\max NDP_i = T1NDC_i + \left(\sum_{j=1}^{N-1} BCL_{ji} \cdot \theta\right) \qquad (9.5)$$

The minimum value of collateral pledged by participant i to cover its position is

$$C1_i + C2_i = T1NDC_i + (\max_j(BCL_{ij}) \cdot \theta) \qquad (9.6)$$

Thus, a participant could incur a position exceeding the value of its own collateral. The maximum own-collateral shortfall for any participant i ($maxOCS_i$) is equal to equation (9.5) minus equation (9.6), or

$$\max OCS_i = \left(\sum_{j=1}^{N-1} BCL_{ji} - \max_j(BCL_{ij})\right) \cdot \theta \qquad (9.7)$$

---

[10] The balance of a participant's settlement account at the Bank of Canada will normally be zero.

[11] Appendix 1 provides proof that there will be sufficient collateral to cover one but not necessarily multiple defaults.

[12] Recall from equation (9.3) that each participant's T2NDC is equal to the sum of BCLs received multiplied by the systemwide percentage.

The maximum own-collateral shortfall also represents the maximum losses to be divided among survivors. In the case of a default where the defaulter has a collateral shortfall, a non-discretionary advance will be granted with a value equal to the defaulter's apportioned collateral ($C1_i + C2_i$), and the survivors will contribute funds to cover the residual shortfall, where the residual shortfall will have an upper bound of $maxOCS_i$.

9.2.5 Loss allocation to survivors

If any one participant i defaults, each participant that granted a BCL to that participant will have to contribute funds to cover participant i's shortfall. Participant j's additional settlement obligation (ASO_j) is calculated according to the following formula[13]

$$ASO_j = OCS_i \cdot \frac{BCL_{ji}}{\sum_{j=1}^{N-1} BCL_{ji}} \tag{9.8}$$

Therefore, survivors cover the defaulter's shortfall, with each survivor contributing in proportion to the BCL that it has granted to the defaulter.

9.2.6 Feature of LVTS under analysis

I have shown that, in most circumstances, the BCLs granted to a participant can allow collateral shortfalls (equation 9.7) and defaults. The system is robust to defaults. LVTS's rules give participants the ability and incentive, not the requirement, to limit their maximum potential losses to a size that they can manage from a solvency perspective. However, the impact of losses on participants' capital adequacy is not known with confidence. This study estimates that impact.

---

[13] In this formula, BCLji represents the largest BCL participant j has granted to defaulting participant i at any time during the day of default. This is important, because participants can increase or decrease their BCLs granted during the day and contribute based on their maximum BCLs granted to the defaulter during the day.

9.3 Comparison with previous literature

In his 1992 and 1993 papers, Engert considers risk controls in payments systems that provide for a system's robustness to default at a minimum cost in terms of collateral requirements. Through a theoretical model, Engert finds that when a payments system is designed such that survivors share in a defaulter's losses, a system's robustness to default does not necessarily indicate that the same is true for individual participants. Since LVTS is a system that falls into this category, it is important to empirically test whether participants are robust to defaults.

Previous researchers have studied the potential for contagion following an initial default in uncollateralised netting-based payments systems (good examples are Northcott, 2002, Humphrey, 1986, and Angelini, Maresca, and Russo, 1996). They assume that the defaulter cannot pay the funds owed to cover its position, and that the participant does not have any collateral to use to fulfill payment of its obligation. The researchers make assumptions about key factors such as unwind rules, provisional credit granted to customers, and the ability of the remaining participants to withstand the losses resulting from the initial default to determine whether there are any subsequent defaulters. Sensitivity analysis is performed, and the researchers are able to determine frequencies and magnitudes of knock-on defaults.

As with the previous studies, my aim is to study the effects of initial defaults on the payments system. A number of differences exist from the previous studies, however, based mainly on the fact that LVTS is a collateralised netting system:

- Rather than assuming that each participant that ends the day with a net debit position defaults, I find each participant's largest net debit position during the day and assume that it is shut down at that time. If the participant has a net debit position, the participant will be a defaulter. Accordingly, a larger number of defaults, and larger net debit positions, occur using this method than if participants' end-of-day positions had been used.
- Because a defaulters' own collateral is first used to cover its net debit position, losses do not accrue to survivors in all cases of default, as they do in the previous studies. Defaulters' collateral is taken into account when determining the losses to survivors.
- Payments that have cleared in LVTS are not unwound. Therefore, losses are determined based on actual LVTS loss-allocation rules

and are not offset by funds recovered from the accounts of customers.

Based on these differences, I do not expect knock-on defaults to occur in this study.

In a recent paper, Galos and Soramäki (2005) explore what the potential for systemic risk in TARGET2 would be if it were designed as either an uncollateralised deferred net settlement (DNS) system or a collateralised DNS system much like LVTS. They find that, under all scenarios, the potential for systemic risk is low, but that the loss-sharing rule is important. The most effective loss-sharing rule is one in which banks share in losses relative to their size.

9.4 Methodology and choice parameters

If a participant is closed by its regulator during the LVTS day, it will immediately become ineligible for further participation in LVTS. Other participants will continue to clear and settle payments among themselves for the remainder of the day. At the end of the day, the net position of the closed participant as of its time of closure will have to be settled. If this is a net debit position, the participant will be declared in default. The Bank of Canada will grant a non-discretionary advance equal to the lesser of the defaulter's net position or the value of its collateral apportioned to T1 and T2. The Bank of Canada will then acquire the collateral as remuneration for the advance. In the latter case, survivors that granted a BCL to the defaulter will have to contribute to cover the shortfall according to the formula used in equation (9.8).

In this study, I create defaults by assuming that each participant is closed by its regulator at the time it incurs its largest combined T1 and T2 net debit position on each day. I find each participant's largest combined net debit position rather than its largest T2 (survivors-pay) net debit position because, at settlement, each participant must settle its combined T1 and T2 position and can use all its collateral to do so. For settlement purposes, a net credit position in T1 will offset a net debit position in T2 or vice versa. The maximum potential shortfall between a participant's net debit position and its collateral occurs when the participant incurs its largest combined T1 and T2 net debit position.[14]

I run T1 and T2 transactions together through the simulator and obtain each participant's maximum net debit position, and the time at which it occurs, from the simulator's output statistics.[15,16] If this is a net debit position, this is an instance of default. The net debit position is then compared with the participant's collateral and, if the former is greater, the participant has incurred a shortfall. The number and value of shortfalls for each participant are recorded. In each case, survivors' losses (ASOs) are calculated. The average and maximum losses of each surviving participant are compared with their assets and regulatory capital requirements to assess whether the survivor can withstand the loss.

9.5 Findings

9.5.1 Data

The period of study spans the 170 business days from 1 March to 29 October 2004. Over this period, the average daily volume and value of payments were 17,063 and CAD 130.2 billion, respectively.

The names and abbreviations of the fourteen institutions that participated in LVTS during the sample period are listed in Box 1.[17] This group contains eight domestic banks (ATB, BMO, BNS, CIBC, LAR, NAT, RBC, and TD), two foreign bank subsidiaries (HSBC and BNP), one foreign bank branch (BOA), one co-operative financial group (CCD), one central finance facility for Canadian credit unions (CUCC), and Canada's central bank (BOC). Participants are classified into 'large' and 'small' participants, with the threshold being assets of CAD 200 billion. Total assets of each participant are reported in Table 9.1.

---

[14] Each participant's end-of-day collateral holdings are used each day for simplicity and in most cases represent its maximum collateral holdings for that day.

[15] The data contain only transactions that have passed the LVTS risk controls, so I can run simulations without incorporating credit limits. I can combine T1 and T2 in one simulation to find participants' combined maximum net debit position, because credit limits are not applied.

[16] To view patterns in the time at which participants incur their largest shortfalls, see Figure A2.1 in Appendix 2.

[17] State Street Bank and Trust Company is excluded from the analysis because it joined LVTS only in October 2004.

Box 1: LVTS participants

Alberta Treasury Branches Financial (ATB)
Bank of America National Association (BOA)
Bank of Canada (BOC)
Bank of Montreal (BMO)
Bank of Nova Scotia (BNS)
BNP Paribas (Canada) (BNP)
Caisse centrale Desjardins du Québec (CCD)
Canadian Imperial Bank of Commerce (CIBC)
Credit Union Central of Canada (CUCC)
HSBC Bank Canada (HSBC)
Laurentian Bank of Canada (LAR)
National Bank of Canada (NAT)
Royal Bank of Canada (RBC)
Toronto-Dominion Bank (TD)

Transactions, collateral and BCL data are used to determine participants' maximum positions, shortfalls, and loss allocations.

- The transactions data contain the sender, recipient, value, tranche, and submission time for each transaction that was successfully cleared by LVTS during the data sample.
- Collateral data contain each participant's value of collateral apportioned and pledged, and the date and time effective.
- BCL data contain the grantee, grantor, value, date and time effective of each BCL, which are used to calculate participants' ASOs.

Table 9.1 **Assets of LVTS participants**

Rank	Participant	Assets (CAD billion)
1	ROYAL	429.26
2	TD	310.55
3	BNS	284.89
4	CIBC	281.72
5	BMO	267.73
6	CCD	103.57
7	NAT	83.45
8	CUCC	74.77
9	HSBC	40.71
10	LAUR	16.50
11	ATB	14.59
12	BOA	4.90
13	BNP	4.26

Note: All participants with assets exceeding CAD 200 billion are classified as large participants. This methodology results in five 'large' participants and eight 'small' participants.

Source: The Office of the Superintendent of Financial Institutions and participants' websites.

As previously described, I also benchmark shortfalls against total assets and capital:

- Information on federally regulated deposit-taking institutions and foreign bank subsidiaries is obtained from the website of the Office of the Superintendent of Financial Institutions (OSFI (2005a) and (2005b)). Information on monthly assets is obtained from participants' consolidated balance sheets, and information on quarterly Tier 1 and total capital is obtained from participants' capital adequacy reports.
- For ATB, information on annual total assets, Tier 1, and total capital is obtained from its 2004/2005 Annual Report (Alberta Treasury Branches Financial (2005)).[18]
- For BOA, information on annual total assets, Tier 1, and total capital is obtained from its 2004 Annual Report (Bank of America National Association, 2005).[19]

---

[18] The time period for these data is from 31 March 2004 to 31 March 2005.

[19] It is relevant to use figures for the Bank of America National Association rather than the Canadian bank branch.

- For CCD, information on total assets and equity (the best estimate of regulatory capital for this institution) is obtained from its 2004 Annual Report (Desjardins Group, 2005).
- For CUCC, information on total assets and members' equity (an estimate of regulatory capital) is obtained from its 2004 Annual Report (Credit Union Central of Canada, 2005). These figures represent aggregates of the credit unions and caisses populaires affiliated with CUCC.

9.5.2 Results

The sample contains 170 days and 13 potential defaulters.[20] Recall that, since participants are assumed to be closed at the time of their largest net debit position, the methodology is expected to yield defaults in almost all cases; that is, for most participants on most days. Indeed, defaults occur in 2,167 of the 2,210 potential cases. These defaults result in 1,026 shortfalls.

Result 1: Shortfalls are relatively frequent and, on average, small. However, there is considerable variability across participants and days.

Recall that a participant is considered to have incurred a shortfall in each instance that its position at the time of closure exceeds its apportioned collateral. I find that shortfalls occur relatively frequently – in 46 per cent of cases. Individual participants' instances of being in a shortfall position range from 0% to 95% of days. Large participants incur shortfalls 15% more frequently than small participants.

Figure 9.1 illustrates the size distribution of shortfalls for all participants in the 46% of cases where shortfalls are incurred. As shown, most shortfalls are relatively small. Considering the size of shortfalls more closely provides the following conclusions:

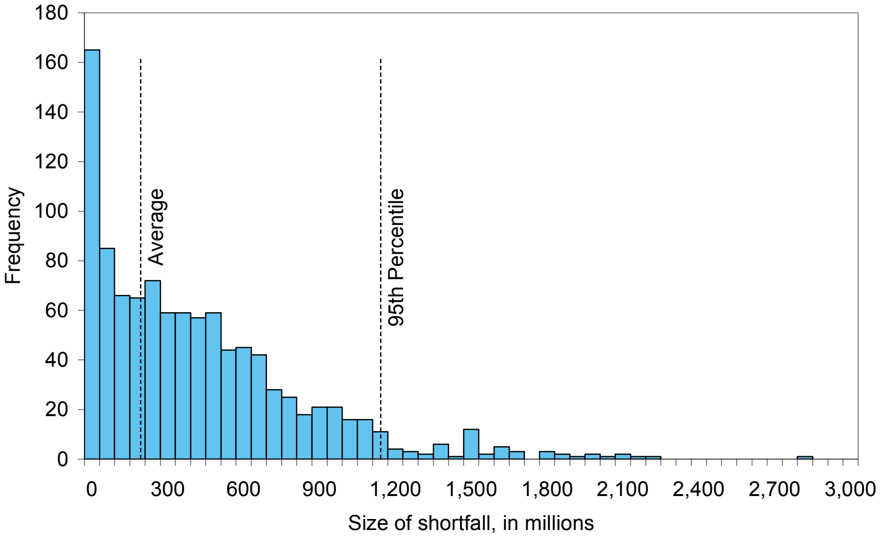
- The average shortfall size for all participants is CAD 210.4 million, with a standard deviation of CAD 181.7 million.
- Shortfalls are, on average, four times larger on participants' worst days than on average days.

---

[20] The Bank of Canada is not a potential defaulter.

- Large participants incur shortfalls that are, on average, nearly three-and-a-half times larger in absolute terms than those of small participants.
- The average shortfall size for the participant that incurs the largest shortfalls is approximately four times that of all participants.
- The largest single shortfall in the sample is nearly CAD 2.9 billion. However, 95% of all shortfalls are under CAD 1.2 billion.

Figure 9.1 **Size distribution of all participants' shortfalls**



Result 2: The shortfalls incurred are much smaller than the maximum shortfalls possible.

Recall from equation (9.7) that participants can incur a maximum shortfall equal to a fixed percentage of the amount by which their T2 net debit cap exceeds their T2 collateral. On average, participants incur actual shortfalls that are very small – just 18.1% of the maximum possible. On each participant's worst day, shortfalls are, on average, 81.3% of the maximum possible. Accordingly, average stresses on the system are small. However, at times, participants utilise most of the credit granted to them.

Result 3: Survivors' loss allocations are generally small and borne by participants that are most able to withstand them.

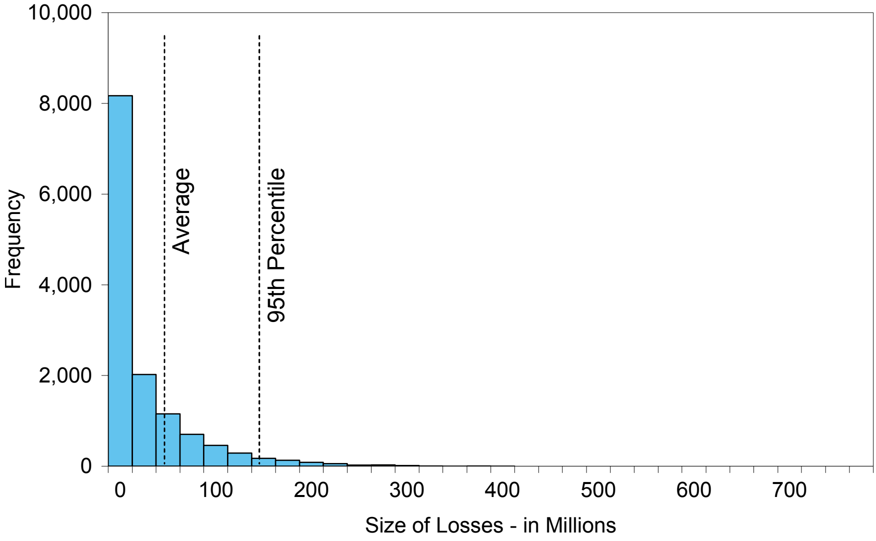
Recall from equation (9.8) that, following a default, each survivor is allocated a share of the defaulter's shortfall in proportion to the size of the BCL that it has granted to the defaulter. Figure 9.2 illustrates the distribution of survivors' losses. As with shortfalls, losses are, in general, small but variable.[21] Specifically,

- The average loss allocation to any participant over the sample period is CAD 16.2 million, with a standard deviation of CAD 38.1 million.
- The average worst loss that participants are exposed to on any day is 15.6 times larger than the average and amounts to CAD 252.8 million. Therefore, the day that a default occurs could affect the size of participants' losses.
- Large participants' loss allocations are, on average, 3.7 times those of small participants. Large losses are thus borne by large participants that are better able to bear them.
- The largest loss allocation any participant receives on any day is CAD 753.7 million. However, 95% of losses are CAD 136 million or lower.

---

[21] Recall that a defaulter's shortfall is the amount by which the defaulter's net debit position exceeds its collateral pledged to the system, so survivors' losses take defaulters' collateral into account.

Figure 9.2 **Size distribution of all participants' loss allocations**



Result 4: Small participants take on the greatest losses compared with asset size.

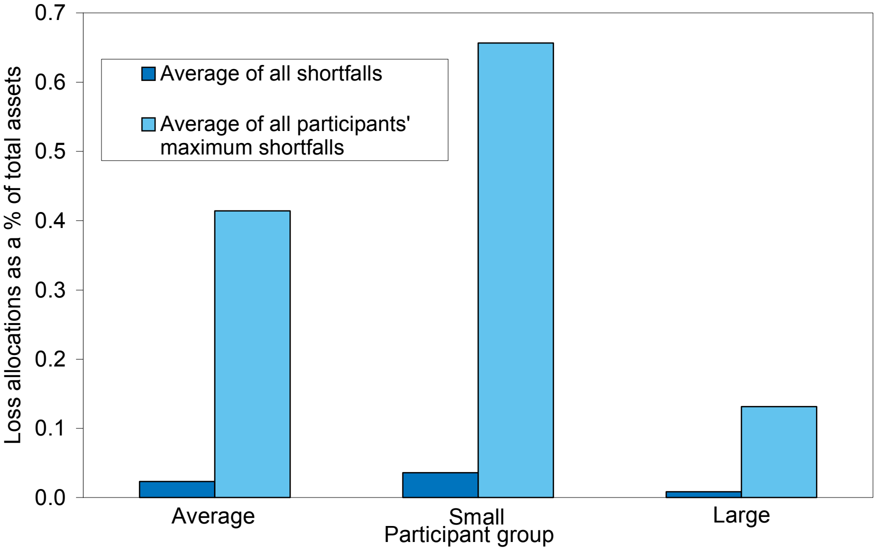
To scale loss allocations for each participant, I compare losses with each participant's total assets and refer to loss allocations divided by total assets as loss-to-asset ratios. For participants overall, loss-to-asset ratios are small. The average loss-to-asset ratio for all participants is 0.02%. When the largest loss allocation that each participant incurs on any day is considered, the loss-to-asset ratio increases to only 0.4%.

Figure 9.3 illustrates average and maximum loss-to-asset ratios for all participants, and also when grouped as large or small participants.[22] Small participants withstand losses that are approximately four times larger as a proportion of assets than large participants, meaning that small participants take on relatively more risk in the system. The loss ratios for all small participants but one, however, are very small.

---

[22] Note that because the Bank of America is a branch, it is considered a large participant, since the assets of the Bank of America (not the Canadian branch) are used to benchmark its loss allocation.

Figure 9.3 **Participants' losses as a percentage of total assets**



Result 5: Losses compared with capital are generally small but lead to noticeable increases in leverage in some cases. Nevertheless, participants are, in all cases, robust to defaults.

Loss allocations are measured against the highest quality Tier 1 capital because it is the most conservative estimate of the resources that banks have to absorb losses. The results illustrate the capital losses that would result from participants' loss allocations, and whether survivors can withstand their losses.

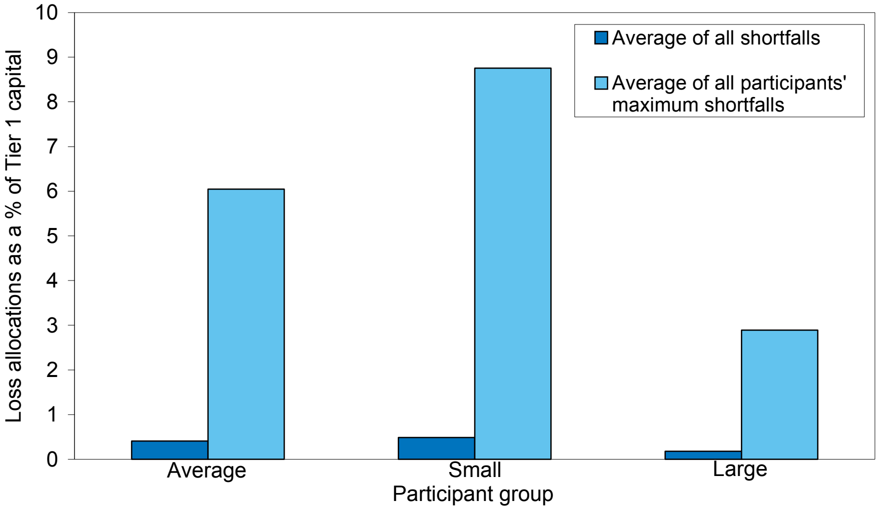
Figure 9.4 illustrates loss-to-capital ratios for all participants, and also when grouped as large and small participants.[23] Participants' average loss-to-capital ratios are very small: losses as a percentage of capital amount to just 0.35% on average. On the worst days, however, participants' loss-to-capital ratios are 17 times larger. Small participants' average loss-to-capital ratios exceed those of large participants by approximately three times. Thus, two themes that have occurred throughout the results are repeated: (i) the impact of a default on a worst day greatly exceeds that of an average day, and (ii) small participants are more affected by defaults than large participants.

---

[23] The Bank of America National Association's capital is used and that bank is considered to be a large participant.

In the worst case, losses can be as high as one-third of capital. Even in the worst case, however, the participant's capital remains better than that required by its supervisor. Therefore, even the most significant loss does not cause any participant to subsequently fail.

Figure 9.4 **Participants' losses as a percentage of capital**



Result 6: If participants were to use all their pledged collateral to cover their net debit position, shortfalls and losses would be both much smaller and much less frequent.

Recall that shortfalls and losses have thus far been calculated based on the collateral that participants have apportioned to Tranches 1 and 2 in LVTS. Apportioned collateral represents participants' minimum levels of pledged collateral required for LVTS. On average, however, participants pledge approximately three times their apportioned collateral to the Bank of Canada and this excess collateral could be apportioned (or put into use) at any time.[24]

When based on total collateral pledged, the instances and values of shortfalls decrease significantly. In fact, five participants do not incur a shortfall on any day during the data sample. Losses incurred by

---

[24] See McPhail and Vakos (2003) for a model of collateral holdings in LVTS that largely explains why participants hold excess collateral. Participants' total collateral needs are a function of the opportunity and transactions costs of collateral, the variance in payments flows and the cost of payments delay.

survivors and loss ratios also decrease by between 50% and 90%. The implication is that participants' losses become almost negligible as a percentage of their assets. Even the participant that consistently incurs the greatest loss ratios sees its largest loss incurred reduced to less than 1% of its total assets (Table 9.2).

Table 9.2 **Potential shortfalls and losses based on total collateral**

	Result	Reduction compared with base case (apportioned collateral) (%)
Shortfall on percentage of days	8.3%	71
Average shortfall	CAD 20.4 million	90.3
Average of participants' maximum shortfall	CAD 446.0 million	49.0
Average loss	CAD 1.86 million	88.5
Maximum loss	CAD 35.44 million	53.6
Average loss-to-asset ratio	0.002%	90
Maximum loss-to-asset ratio	0.05%	54.5

9.6 Factors affecting shortfalls and losses

The shortfalls and losses to survivors found in this study are based on the assumptions that each participant is closed at the time of its largest net debit position incurred, given actual LVTS data and that the default is unanticipated. This section considers the effects of changing particular assumptions central to the analysis.

(i) Closure occurs during the LVTS day

I have assumed that a participant is closed by its regulator during the LVTS day, and my assumptions make it possible to easily generate unanticipated defaults and have them occur at the worst moment in the day. In all likelihood, a regulator would avoid shutting down a participant during the Canadian business day (and during the LVTS day). If a participant were closed outside of LVTS hours, the payments system would not be directly affected.

(ii) Shortfalls based on positions actually incurred

As section 9.5.2 explains, participants incur shortfalls that are small compared with the maximum shortfalls allowed.[25] If a participant were to experience large payment outflows prior to a default (if, for instance, a bank's failure were widely anticipated and a bank run resulted), the participant might incur a shortfall that was close or equal to its maximum allowed shortfall. Increasing participants' shortfalls to the maximum allowed, other things equal, would create losses for survivors that are much larger than the ones found here.

(iii) BCLs granted to the defaulter

The assumption that a default is unanticipated means that BCLs are likely larger than they would be in the case of an anticipated default. Participants have an incentive to reduce BCLs to a participant they believe may default in order to minimise their exposure to the defaulter. Other things equal, smaller BCLs granted to the defaulter would result in smaller losses for survivors than we find here.

(iv) Recovery rates are not taken into account

In the event that a participant incurred a loss resulting from the default of another participant, it would become an unsecured creditor to the estate of the failed institution. It is likely that the defaulter would recover some portion of its loss. Other studies point to recovery rates of 40% and 95%. See Furfine (2003), James (1991) and Kaufman (1994). In Canada, recovery rates for bank failures that occurred between 1967 and March 2001 are estimated at 70–80%.[26]

I have chosen not to reduce participants' ASOs by expected recovery rates for two reasons. First, participants must meet their entire ASO on the day a participant defaults. Thus, using participants' entire ASOs to estimate losses illustrates the upfront and maximum obligation that participants will incur before they recover some portion of their funds later. Second, I can be very conservative and not account for recovery from the estate of the failed institution because I do not observe any knock-on defaults.

---

[25] See equation (9.7) to understand the maximum shortfalls that participants can incur.

[26] From the Canadian Deposit Insurance Corporation annual reports and Bank of Canada staff calculations.

9.7 Conclusion

LVTS incorporates risk controls and a residual guarantee by the Bank of Canada that make it robust to multiple defaults. It employs risk-sharing whereby survivors may be allocated a share of the defaulter's losses in the event of a default. The system's rules give participants both the ability and the incentives to control their exposures to other participants so as to keep potential losses manageable. In this paper, I have created the largest possible unanticipated defaults based on a sample of actual LVTS activity and estimated whether participants are adequately controlling their risk to be able to withstand the default of another participant.

I have found that, in general, participants are easily able to withstand their loss allocations. This partly results from the fact that, on average, participants, each of which I consider in turn to be a defaulter, create net debit positions intraday that are much smaller than the maximums possible. In both absolute and relative terms, participants' loss allocations are generally small, but when I compare losses with assets and capital, small participants take on relatively much more risk than large participants.[27] Nevertheless, I find that all participants are robust to the defaults generated here.

I have also calculated results based on defaulters covering their positions with all their LVTS collateral, including the significant amount of excess collateral most keep in reserve for LVTS purposes. The frequency and size of shortfalls and survivors' losses decrease by between 50% and 90%.

I believe that the losses found in this study are probably larger than would occur if a participant were actually to default. First, I have used the largest shortfalls I can create based on the data to maximise survivors' losses. Second, I have assumed that the default is unanticipated. This prevents participants from reducing or eliminating BCLs to the defaulter to avoid sharing losses. Finally, I have assumed that survivors do not recover any of their losses. Although the theoretical shortfalls generated in this study are small compared with the maximums that defaulters could incur, I believe that the other three factors, and especially the second, greatly outweigh this fact to create losses that are much larger than what one would expect to observe in reality.

---

[27] This is because small participants on average grant larger bilateral credit limits relative to their size than do large participants.

There appear to be two important questions for further study. First, why are participants' net debit positions so small compared with the maximums allowed in LVTS? Second, what would be the effect of an anticipated default in LVTS? An anticipated default would likely affect both BCL-setting behaviour and participants' positions. I believe that the impact of an anticipated default would likely be smaller than those considered here, but this requires further analysis.

References

Alberta Treasury Branches Financial (2005) **2004/2005 Annual Report.** http://www.atb.com/dev/aboutatb/documents/atb_2005_annual_report.pdf.

Angelini, P – Maresca, G – Russo, D (1996) **Systemic Risk in the Netting System.** Journal of Banking and Finance 20, 853–898.

Bank for International Settlements (2003) **Glossary of Terms Used in Payment and Settlement Systems.** http://www.bis.org/publ/cpss00b.htm.

Bank of America National Association (2005) **2004 Annual Report.** http://www.bankofamerica.com/annualreport/2004/

Credit Union Central of Canada (2005) **Annual Report 2004.** http://www.cucentral.ca/Public/insidecdn/financialdata/2004AnnualReport/index2.htm.

Desjardins Group (2005) **Annual Report of the Desjardins Group.** http://www.desjardins.com/en/a_propos/publications/rapports_financiers/rapport_annuel/index.jsp.

Engert, W (1992) **An Introduction to Multilateral Foreign Exchange Netting.** Bank of Canada Working Paper No. 1992–5.

Engert, W (1993) **Certainty of Settlement and Loss Allocation with a Minimum of Collateral.** Bank of Canada Working Paper No. 1993–14.

Furfine, C H (2003) **Interbank Exposures: Quantifying the Risk of Contagion.** Journal of Money, Credit and Banking 35, 111–128.

Galos, P – Soramäki, K (2005) **Systemic Risk in Alternative Payment System Designs.** European Central Bank Working Paper No. 508.

Humphrey, D B (1986) **Payments Finality and Risk of Settlement Failure.** In Saunders, A and White, L (eds) Technology and the Regulation of Financial Markets: Securities, Futures, and Banking, 97–120. Lexington Books. Lexington, MA.

James, C (1991) **The Losses Realized in Bank Failures.** Journal of Finance 46, 1223–1242.

Kaufman, G G (1994) **Bank Contagion: A Review of the Theory and Evidence.** Journal of Financial Services Research 8, 123–150.

McPhail, K – Vakos, A (2003) **Excess Collateral in the LVTS: How Much is Too Much?** Bank of Canada Working Paper No. 2003–36.

Northcott, C A (2002) **Estimating Settlement Risk and the Potential for Contagion in Canada's Automated Clearing and Settlement System.** Bank of Canada Working Paper No. 2002–41.

Office of the Superintendent of Financial Institutions (OSFI) (2005a) **Financial Data - Banks.** http://www.osfi-bsif.gc.ca/osfi/index_e.aspx?ArticleID=554.

Office of the Superintendent of Financial Institutions (OSFI) (2005b) **Financial Data – Foreign Bank Branches.** http://www.osfi-bsif.gc.ca/osfi/index_e.aspx?ArticleID=556.

Appendix 1

Shortfalls and coverage by defaulters

Coverage of the largest net debtor's position

As noted in the main text, the risk controls ensure that there will be sufficient collateral to cover the default of the largest net debtor. Recall that all participants that grant BCLs to other participants are required to apportion collateral to cover the largest BCL that they grant multiplied by the systemwide percentage and, in the event of one or more defaults, will have to contribute up to that amount. Thus the collateral apportioned by all participants, other than the defaulter (participant i), to cover defaults is as follows

$$\sum_{j=1}^{N-1} C2_j = \theta \cdot \sum_{j=1}^{N-1} \max_i(BCL_{ji}) \qquad (A1.1)$$

Using participant i's maximum own-collateral shortfall from the main text, the maximum loss accruing to surviving participants is as follows

$$\sum_{j=1}^{N-1} \max L_j = \theta \cdot \left(\sum_{j=1}^{N-1} BCL_{ji} - \max_j(BCL_{ij}) \right) \qquad (A1.2)$$

where $\sum_{j=1}^{N-1} \max L_j =$ the maximum losses from participant i's default that other participants j could incur.

Therefore, the collateral apportioned by survivors always exceeds the survivors' maximum possible losses

$$\sum_{j=1}^{N-1} \max_i(BCL_{ji}) > \left(\sum_{j=1}^{N-1_{-i}} BCL_{ji} - \max_j(BCL_{ij}) \right) \qquad (A1.3)$$

which must hold because $\sum_{j=1}^{N-1} \max_i(BCL_{ji}) \geq \sum_{j=1}^{N-1} BCL_{ji}$ and $\max_j(BCL_{ij}) \geq 0$.

The default of two participants

If more than one participant defaults on the same day, the maximum that each surviving participant, j, will have to contribute to cover the losses of all defaulters on a single day is its maximum additional settlement obligation ($\max ASO_j$), which is set equal to the maximum BCL it has granted to any other participant, multiplied by the systemwide percentage. Recall that participants apportion T2 collateral equal to this value, so $\max ASO_j = \theta \cdot \max_i(BCL_{ji}) = C2_j$. Participants' ASOs vis-a-vis each defaulter are calculated, and if any participant's combined ASOs resulting from the multiple defaults on a single day exceed its maximum ASO, its actual ASO will be set equal to its maximum ASO.

Consider a case where participants i and k default on the same day. Participant j's actual ASO is as follows

$$ASO_j = \min\left[\max ASO_j, \left(OCS_i \cdot \frac{BCL_{ji}}{\sum_{j=1}^{N-1} BCL_{ji}} + OCS_k \cdot \frac{BCL_{jk}}{\sum_{j=1}^{N-1} BCL_{jk}}\right)\right] \quad (A1.4)$$

Therefore, participant j's ASO is the minimum of its maximum ASO and the sum of its loss allocations to the two defaulters. In the latter case, the Bank of Canada will contribute the difference.

In this case of two defaulters on a single day, it is possible that the second term in equation (A1.4), the survivor's calculated share of the losses, exceeds the first term: the collateral of participant j. Whether each survivor's calculated ASOs are met (that is, whether survivors cover all losses) depends[28]:

- positively on the size of the largest BCL it has granted to any participant, assuming that the largest BCL is not granted to either defaulter;
- negatively on each defaulter's own collateral shortfall; and
- negatively on the ratio of the BCL that the survivor has granted to each defaulter compared with its maximum BCL granted,

---

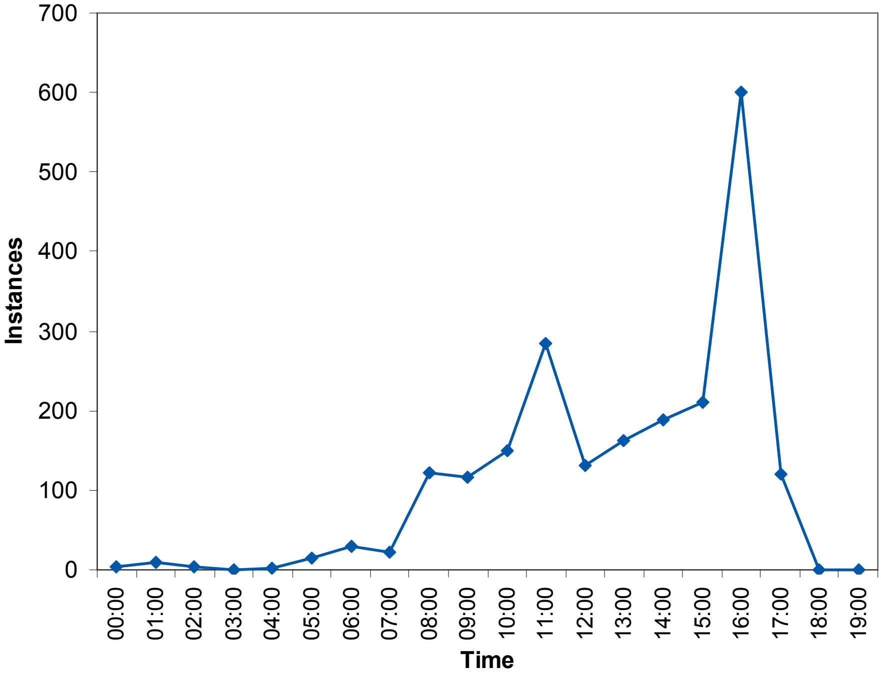
[28] Recall that the Bank of Canada will have an ASO equal to 5% of each defaulter's losses in this case, because it has granted a BCL to each participant of 5% of the sum of BCLs received from other participants.

assuming that its maximum BCL is granted to a surviving participant.

Therefore, in the case of multiple defaulters on a single day, the Bank of Canada may have to contribute.

Appendix 2

Figure A2.1 **Time of participants' maximum net debit positions**



The most common time for participants to incur their maximum net debit position is between 4 pm and 5 pm, which corresponds to the settlement of Canada's securities clearing and settlement system. The next most common time is between 11 am and 12 pm, which corresponds to the settlement of Canada's retail payments system.

Chapter 10

Simulation of operational failures in equities settlement

Matti Hellqvist – Heli Snellman

10 Simulation of operational failures in equities settlement

Abstract

This paper presents a simulation model of Finnish securities settlement system for equities. With the model, scenario analysis and stress testing are performed. The purpose of the study is to analyse robustness and efficiency of this equities settlement system and systemic risk caused by failures in securities settlement.

The analysis is based on real data of settlement transactions and flow of post trade process. The main scenarios of the analysis were operational failures of different players or components of the post trade infrastructure: the clearing participants, the main settlement algorithm and the liquidity bridge between the large value payment system and securities settlement system.

Results of the simulations show that operational failure of individual clearing participants should last more than one settlement day to cause significant impact in the settlement system under study. No contagion from failures in securities settlement system to operation of large value payment system was observed. The main settlement algorithm mitigated efficiently the impact of the scenarios in the settlement system.

All the observations are bound to the data set used in this study. The constructed model allows only very limited reactions to the intermediaries in the simulated scenarios.

10.1 Introduction

One of the statutory tasks of many central banks is to ‘promote the smooth operation of payment systems’. Oversight of financial market infrastructure is derived from this responsibility. The task of oversight is to monitor the overall or systemic risks and performance of the financial market infrastructure and to contribute to developments to foster increased stability and efficiency of the systems. Although the quoted wording of the statutes does not directly refer to securities settlement systems, they are included in the scope of oversight because of the ample liquidity flows between payment and securities settlement systems. Also, the handling of collateral for central bank credit connects some securities settlement systems to the transmission channel of monetary policy and thus increases their importance to central banks and financial markets as a whole.

This paper presents a scenario analysis and stress-testing of HEXClear. HEXClear is the securities settlement system for equities operated by the Finnish Central Securities Depository (APK). The main scenarios of this study describe (1) failure of the entire settlement process to repatriate funds, (2) failure of individual clearing participants in the pre-settlement process, (3) failure of certain ICT connections shared by several clearing participants and (4) failure of the most commonly used settlement algorithm.

The purpose of the analysis is to test the robustness of post-trade processing in HEXClear and to quantify the impact of assumed scenarios within the settlement system and the contagion effects to other connected financial market infrastructure. The impact is measured by the number and value of unsettled transactions and the magnitude of liquidity shortages faced. Ultimately, the possibility of systemic risk due to failures in HEXClear is analysed. The efficiency of the settlement process is also studied by comparing the liquidity usage of normal and back-up settlement process. The capability of the settlement process to mitigate the impact of simulated failure scenarios and to reach a relatively high settlement ratio in stress situations is also an indicator of the efficiency of the settlement.

The analysis is based on real data of settlement transactions and book-entry account balances from a one-month observation period during June and July 2005. The simulations are performed with a model of the HEXClear system in the Bank of Finland payment and settlement system simulator (BoF-PSS2). Simulated operational failure scenarios are constructed on the basis of an analysis of information flows from the post-trade process and data. Thus the majority of the simulations in this study do not feature any reactions or changes in the behaviour of the clearing participants. In some scenarios limited variations from historical actions are allowed in the form of the possibility to import more cash to the settlement process in a failure situation. This enables the measurement of additional cash needs caused by the assumed failure and of the impact of such reactions at system level.

This study contributes to the scarce literature of simulation studies and risk analyses of securities settlement systems and securities settlement architectures. Simulation was already mentioned in BIS (1992) as the generally applicable method for assessing ’arrangements for ensuring timely completion of settlements’. It is known that simulations have been used in the design and test phases of several securities settlement systems.[1] Such simulations are, however, seldom published and normally focus on the applicability of the proposed settlement process instead of systemic risk or efficiency issues.

Hellqvist and Koskinen (2005) have reported results from a study similar to the current one, which included stress-testing and simulations of the Ramses system, ie the Finnish securities settlement system for money market transactions. The study presents empirical distributions for the liquidity impact of operational failures of clearing participants and the scale of cascading settlement failures during settlement. According to the results of that study, the Ramses system seems not to pose a significant systemic risk to Finnish financial markets, even in the most severe scenarios.

Besides that, the only published simulation-based studies of securities settlement systems of which the authors are aware are by Iori (2004) and Devriese and Mitchell (2006). Both of these are based on artificial data. Iori compares gross and net securities settlement structures and studies the system-wide implications of exogenous random delays of individual transactions. The number of settlement cycles and several market parameters are also varied in the model. The study reported no clear ranking of settlement system architectures although gross system structure seemed to be more stable than net structure. Devriese and Mitchell construct an artificial securities market and settlement system. They analyse the dynamic impact on settlement of default of the largest participant in the artificial market. According to the authors, central bank liquidity support cannot abolish the settlement failures caused by significant market disruptions because there are both cash and securities legs in the securities transactions, and liquidity support can only affect the cash leg.

The current study uses a similar approach to that of Devriese and Mitchell, ie failure – although operational – is aimed at institutional participants involved in the post-trade process. Iori, instead, models the delays of individual trades due to human errors. Thus the focus of this study is on abnormal situations, while Iori models the reasons behind sub-optimal settlement ratios in normal day-to-day settlement. Compared to both referred studies, the current study has the advantage of real data and realistic settlement logic. It is also based on a more realistic picture of market structure and market practices. The

---

[1] See eg Nyholm (2004), who also gives the description for the implementation of HEXClear used in this study, and Riksbank (2004), who mention a simulation made by VPC, the Swedish CSD, in their assessment of the then new securities settlement system in Sweden.

resulting settlement ratios observed in this study can be compared to the theoretical values presented by Devriese and Mitchell.

Some selected results from this analysis have been presented earlier in the Bank of Finland Financial Stability Review in 2005 and 2006. Compared to those articles, this paper presents the analysis and also the model used in more detail. In comparison with Hellqvist and Koskinen (2005), the approach is similar but the current study analyses a different system and uses different data and a more realistic settlement process.

The structure of the report is as follows. Section 10.2 describes the data used in the analysis and market structure of post-trade infrastructure in Finland. In Section 10.3 we present the HEXClear system, the actual settlement algorithm and the constructed simulation model. Section 10.4 presents the scenarios used in simulations and discusses the results of the simulations. Section 10.5 concludes the paper.

10.2 Data and market structure

Data used in this study were collected from the HEXClear system in summer 2005. The detailed data set consists of all transactions on 19 settlement days in June and July, including all the essential information related to transactions that were included in the settlement process of the HEXClear system during the observation period. Only transactions which had Waiting for Settlement (WFS) status at some moment during the observation period were included in the data set.

The data set includes information on all trade enrichments or allocation data that define the actual bookings of settlement transactions in the book-entry accounts. It also contains the history of each transaction and enrichment and information on balances and registrations in the Central Register. In addition to modelling on the BoF-PSS simulator, the data set was directly analysed statistically. These results are presented in this section to broaden the description of the market structure.

10.2.1 Player roles in the post-trade processes

The Finnish post-trade infrastructure is populated by commercial banks, brokerage companies, the Central Securities Depository (CSD)[2], the central bank and providers of outsourced post-trade processing services. The tasks of the post-trade process are, however, divided into a more detailed structure of player roles, which are presented below.

The final investors, which can be individuals or institutions, access the equities market via brokerage companies or broker functions of commercial banks. These work as intermediaries between investors on the exchange. Brokers are also used in off-exchange trades, ie in the OTC market. During the observation period there were 57 brokers operating in the Helsinki Stock Exchange.

When a trade is concluded it is transferred to the settlement system. The clearing party is an institution to which the CSD has granted the right to participate in the clearing and settlement system. Thus they play an intermediary role between brokers and the CSD operating the clearing and settlement system. Each broker needs to have a default clearing party, which is responsible for the settlement of the broker's trades. Clearing parties or clearing participants are the most important players analysed in this paper. During the observation period there were 25 of them in the HEXClear system.

The third players in the post-trade process are the account operators, who have the right to make registrations on the book entry accounts of the Central Register operated by APK. Each clearing participant needs to be either an account operator or an agent of another account operator. Being an account operator is an intermediary role between the Central Register and the brokers, custodian banks or clearing participants or, alternatively, an enlargement to the role of clearing participant. In the data used in this study there were 10 account operators, including APK itself.

There are also custodian banks. They offer custody and account services for foreign customers, work as account operators towards the Central Registry and, based on this, produce for their clients value-added services such as credit limits, securities lending or processing of corporate actions. The custodians are typically large commercial banks, which serve foreign brokers or investors under a separate clearing party name for their custody functions. One reason for this separation of operations to specialised clearing parties is Finnish

---

[2] The Finnish CSD uses the acronym APK.

legislation, which requires a direct holding account structure for Finnish investors. For foreign investors nominee accounts can also be used, and even customer specific nominee accounts are known to exist.

The CSD is involved in the structure by maintaining the Central Register of book-entry accounts and the clearing and settlement system itself. There is currently no central counterparty clearing (CCP) available for equities settlement. The connections between the CSD and the central bank – the cash model – are discussed when the HEXClear system is presented in more detail in Section 10.4.

The definitions of player roles are not the same as the outlines of individual companies. Instead, an institution fulfilling the requirements is granted the licence to operate in a certain role. As a result many institutions serve in multiple roles – for example, some individual commercial banks have all of the roles described, except from being the CSD or central bank.

10.2.2 Settlement transactions

The number of settlement transactions in the observation period was nearly 906,000. The share of exchange transactions, so-called 1-transactions, was 83% of this total volume. When the number of securities transferred in settlement transactions is considered instead of the number of transactions, exchange trades constitute only 29% of all transactions. The rest are off-exchange transactions or so-called 5-transactions. This includes clearing transfers and off-exchange trades (OTC-trades). Many 5-transactions are actually related to exchange trades. Examples of this include an investor using a brokerage company which is not their account operator or custodian. In such a situation the broker can execute, for example, a large sell order from the investor in a number of individual smaller exchange trades (1-transaction) and receive the total amount of securities with a single off-exchange clearing transfer (5-transaction) from the final investor's custodian.

The settlement of exchange trades is normally performed three local banking days after the actual trade (T+3). With off-exchange trades the participants involved can agree the settlement cycle themselves. The settlement system also allows settlement during the trade date, ie with a T+0 cycle. This possibility increases settlement efficiency and decreases replacement cost risk. It is mainly used for off-exchange transactions. In terms of trade volume, 2.7% of these were settled during the trade date in 2005.

Not all the transactions are settled on the agreed date. The data from the observation period show that altogether 14,834 transactions, ie 1.6% of all transactions, were settled later than originally agreed between seller and buyer. This is more common for off-exchange transactions: 8.4% of them are settled later than originally defined while only 0.7% of exchange trades are similarly delayed. A significant share of off-exchange transactions (2.3% or 3,400 transactions) is also settled before the agreed date. This reflects flexibility in the use of off-exchange transactions.

As to the most liquid equities, the relative share of delayed transactions was smaller than for the whole data set. These include Nokia, forest industry companies and other equities in the OMX25 list. Such equities are more often traded by foreign counterparties, which could theoretically slow down processing in case of a failure or delay. One explanation for the smaller share of delayed trades among liquid shares could be the availability of securities lending services for the most liquid equities. This was, however, not analysed, since not all lending transactions were identifiable in the data.

10.2.3 Information flows in settlement process

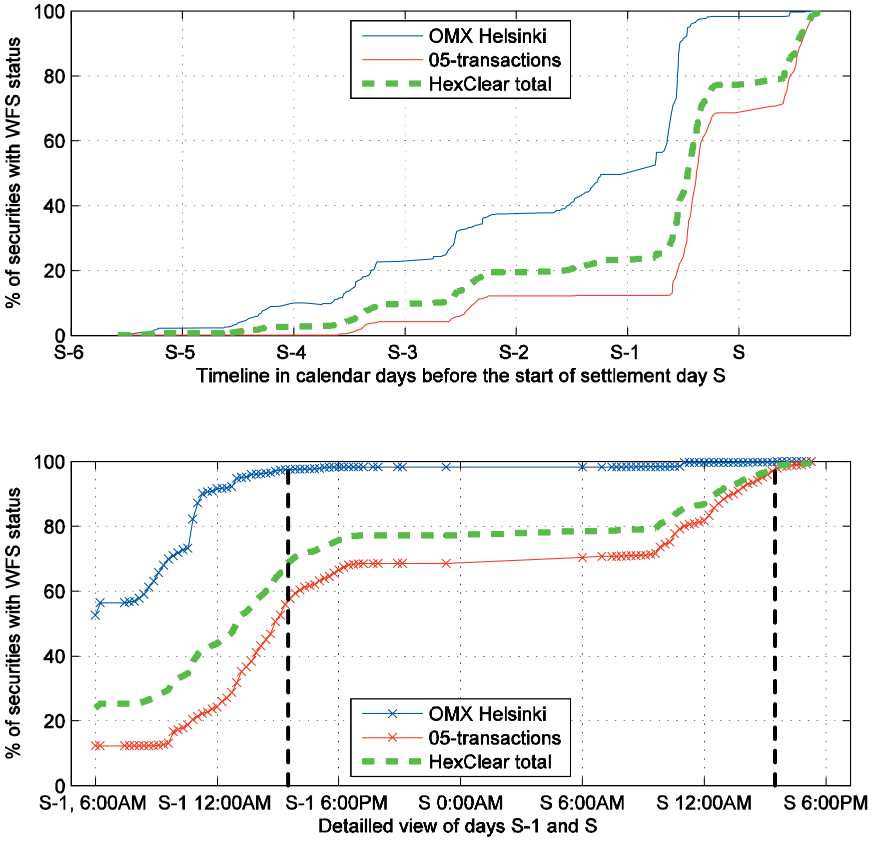
Information flows in post-trade processes are critical to the smooth clearing and settlement of transactions. Impacts of operational failures or other scenarios analysed in this study typically emerge via disturbances in data flows. Thus understanding the timing and structure of information flows is mandatory for system stress-testing.

Equity trades executed in the Helsinki Stock Exchange's SAXESS trading system are automatically transferred to APK's HEXClear system for clearing and settlement. These trades are matched in the trading platform. Transactions other than exchange trades are entered into HEXClear by the clearing participants and matched there. All transactions are settled in the same settlement process.

Settlement of transactions in HEXClear normally takes place three banking days after the trade. All the information needed for settlement is not, however, available on the trade date in the trading system. Allocation data and other trade enrichments are submitted by the clearing participants to the settlement system later on. Transactions which have all the necessary data available are labelled as 'waiting for settlement' (WFS). Figure 10.1 presents the share of securities in transactions which have WFS status as the start of the agreed settlement day approaches. The figure is based on the number of securities because this way it shows the timing of transactions weighted with the transaction size. Free-of-payment transactions are also included in this approach. The timeline in the figure is calendar days and comprises weekends and bank holidays. Thus, in case of an operational failure, all time available for solving the failure is included in the timeline of this figure, not only banking days.

The figures show that the majority of exchange trades reach WFS status before the settlement day begins and 50% are already finalised one day before settlement day. Of the other settlement transactions, just 12% have WFS status one day before and 68% have reached it when the settlement day begins. The remaining 32% of 5-transactions only become ready for settlement during the settlement day.

Figure 10.1 **Cumulative share of book-entries in trades with waiting for settlement (WFS) status as a function of time difference to start of settlement day. Division by marketplace. Days from S-6 are shown in the upper figure and a closer look at days S-1 and S is provided in the lower figure.**



The different usages of transactions depending on their marketplace are one explanation for the observed difference in timing of the final confirmation of transactions. For example, other settlement transactions are more often introduced as T+0 transactions, such as securities loans or other transactions derived from the needs of the ongoing settlement.

The two black vertical lines in the lower figure represent cut-off times given in the rules of HEXClear. The 5-transactions, where book-entries are delivered from a custody situation to another clearing party and where this delivery is linked with some exchange trade, have to be confirmed in HEXClear before 3.30 pm during day S-1.[3] This explains the first jump in the share of 5-transactions with WFS status. The second cut-off time is 3.30 pm during the settlement day, which is the last allowed time for confirmation of trades to custody clearing parties for settlement during the ongoing day. (It also applies to the introduction of T+0 transactions).

The blue line representing exchange trades increases simultaneously with the increase in 5-transactions during day S-1. This indicates that many exchange transactions are only finally modified to WFS status after the underlying clearing transfers have reached WFS status. This reflects the risk management procedures of clearing participants but contradicts the rules of APK, which require that 'a clearing party must, as soon as possible and by no later than after the trade has been confirmed, register allocation data relating to the trade in the HEXClear system'.[4]

The figure only includes trades and other settlement transactions and excludes account transfers performed in the Central Register and other similar events which affect the book-entry account balances. If these were included, they would increase the proportion of transactions matched during settlement day S. This has been taken into account in the simulations and scenario data sets.

10.2.4 Market structure

The post-trade processing of equities in Finland is rather concentrated. This tends to increase the potential impact of operational failures of the major participants, which are analysed in the simulations. The clearing participants with the eight largest market shares of clearing and settlement activities during the observation period[5] are listed below in Table 10.1. The table is based on transaction volumes. An individual company may operate in the settlement system with several clearing party names and the number of clearing parties for each of the participants is also listed.

---

[3] Decisions related to rules of APK, Registration and clearing schedules.

[4] Rules of APK, 4.2.12.

[5] For reference values over longer period see Bank of Finland bulletin, Financial stability 2005, p. 71.

Table 10.1 **Participant shares of settlement transaction volumes and the number of clearing party names of individual companies during the observation period**

Company	% of total volume	Number of clearing party names
Nordea Pankki Suomi Oyj	46%	2
Svenska Handelsbanken AB, Finnish subsidiary	10%	2
EQ Pankki Oy	6%	2
Skandinaviska Enskilda Banken AB, subsidiary in Helsinki	6%	2
Sampo Pankki Oyj	4%	1
Enskilda Securities AB, subsidiary in Helsinki	4%	1
Opstock Oy	4%	1
Fischer Partners Fondkommission AB	4%	1
Others	16%	13

The concentration of the market is different in the settlement of exchange trades and off-exchange trades. On the exchange side, the clearing participants of companies with the three largest shares have together 59% of the total volumes of the whole observation period. Day to day variations in this share were rather small, ranging from 56% to 63%, and with sample standard deviation of 3%.

On the off-exchange side the similar share of companies with three largest shares of the volume is almost 90%. Variation is smaller, ranging from 87% to 91% and with a standard deviation of 1,1%. The group of three largest is different in off-exchange than in exchange transactions.

As another measure for concentration a Herfindahl[6] index was calculated for the whole market. For exchange transaction volumes the index value was on average 0,14 (range 0,11–0,17). This can be compared to a market with an equal division of market shares between seven participants. For off-exchange transaction volumes the Herfindahl index value was higher, on average 0,42 (range 0,40–0.47). This compares almost to a duopoly situation, even if the index was calculated based on market shares of clearing participants instead of the companies shown in Table 10.1.

---

[6] Sum of squared market shares of all participants. Here market shares of clearing parties were used.

The Herfindahl index values of HEXClear can be compared to similar values from the Ramses system, the APK settlement system for money market transactions.[7] There the index values were on average 0,26. Thus the Finnish wholesale securities settlement is clearly more concentrated than the settlement of exchange trades but not as concentrated as off-exchange transactions in HEXClear.

10.3 HEXClear system

The HEXClear system performs the clearing and settlement of equities cash market transactions: exchange trades from OMX Helsinki, off exchange trades of equities made by Finnish counterparties and other settlement transactions. Trades are settled in the HEXClear system either by an optimisation process or in a continuous trade-by-trade settlement. The optimisation process aims at maximising the number of equities in simultaneously settled transactions and it takes place at specified moments in the form of batch runs during a settlement day. In other words, optimisation is used to establish whether a number of individual transactions can be settled together, even though it would not be possible to settle them one-by-one. The RTGS process settles the trades on a gross basis throughout the settlement day. In both processes the settlement is performed with a DVP-1 model,[8] and the settlement is final immediately after booking of securities and transfer of cash is executed.

The HEXClear system provides preliminary information about the expected outcome of settlement during S-1 day and in the morning of the settlement day. These preliminary information batches provide forecasts for the amount of cash required in optimisation and ease the liquidity management of the participants. The final preliminary information batch is based on transactions, which are ready for settlement in the morning of the settlement day. These transactions are earmarked for inclusion in the first optimisation and clearing parties are told of the resulting cash requirements. The first optimisation batch, at 10 am Finnish time, includes the earmarked transactions and the cash submitted by the participants and it normally settles a very large majority of daily totals. During the observation period the share

---

[7] See Hellqvist – Koskinen (2005).

[8] BIS (1992), Delivery versus payment in securities settlement systems.

of first optimisation was on average 97,3% of total daily transaction volume and 79,5 % of the total number of securities settled.

10.3.1 HEXClear cash model

Money transfers of the transactions settled in HEXClear system are executed with cash deposited in the APK account in the Bank of Finland's BoF-RTGS system. The cash position of this account is mirrored into sub-accounts of individual clearing participants in HEXClear. It is stated[9] in Finnish law that these funds are owned by participants and can be used only for settlement of transactions. Thus the overall balance is still the liability of the central bank – central bank money in custody. Additionally, the Bank of Finland has not outsourced central bank accounts to APK and there are no credits granted in the HEXClear system either by APK or the settlement process.

The different central bank money models of securities settlement have been described in a working paper by ECB[10] and illustrated below in Figure 10.2. The advantages of the autonomous cash model of the HEXClear system are operational robustness, the possibility of using DVP model-1 settlement even for retail transactions without flooding the BoF-RTGS with transaction volumes, the fact that no legal outsourcing of central bank accounts or operations are needed and settlement with central bank money. APK's Ramses system which uses a similar model as HEXClear has been approved by the Eurosystem for collateral operations.

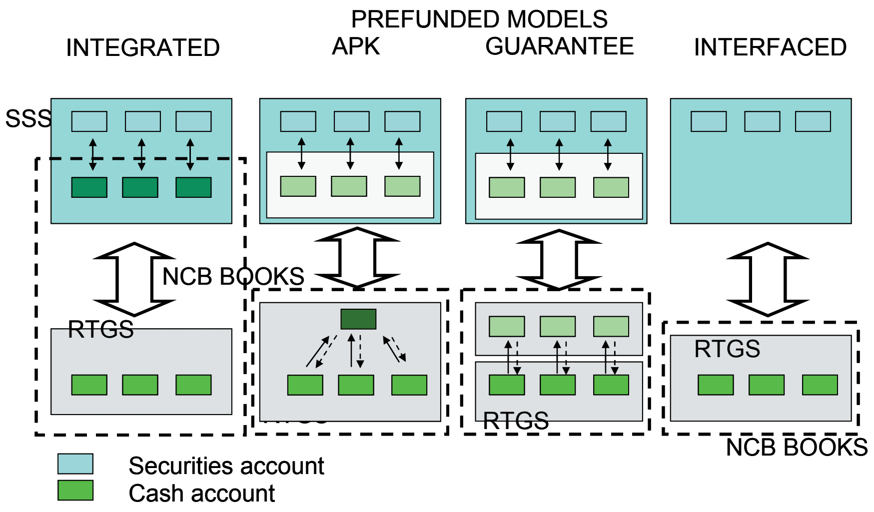
The drawback of the structure is the shattering of liquidity to separate systems and sub-accounts. The same problem applies to all ancillary system structures of large value payment systems. It can be mitigated by efficient and predictable liquidity usage and flexible possibility to transfer liquidity in and out of the system.

---

[9] Securities market act, chapter 4a. Section 9: Safeguarding the position of clearing party.

[10] ECB working papers on various models can be found at http://www.ecb.int/pub/pdf/other/useofcbmoneyforssten.pdf.

Figure 10.2 **CSD central bank money settlement models**



10.3.2 Settlement algorithm

The simulation model used to mimic HEXClear was based on a description presented in Nyholm (2004). This was known not to be the final blueprint of the algorithm, but sufficiently close to it. The key features of the model have been discussed with APK experts to ascertain as close a match as currently possible with the structure of HEXClear during the observation period. Later on there were changes and enhancements in the real HEXClear environment, which are not included or reported in this study.[11]

HEXClear optimisation

The implementation of HEXClear optimisation in the simulation seeks to maximise the number of equities in trades which are selected to be settled, ie included in the solution of the optimisation. The selection of trades to be settled is a knapsack problem and it thus belongs to class of NP-problems. For these there are no known algorithms giving exact an optimal solution in guaranteed polynomial time.[12] To limit the

---

[11] These include changes in the daily schedule of settlement and connections from participants to HEXClear. For up to date information see www.apk.fi.

[12] See eg Papadimitriou (1994), p. 202.

computational time the algorithm uses an approximate heuristic approach.

In the description of the algorithm below, the **net position** of an account is a computational variable. It presents the balance of an individual cash or book entry account which would result if all the bookings included in the solution were executed. The main logic of the algorithm implemented in BoF-PSS2 is presented in Figure 10.3 below. Two subroutines are presented in more detail in the writing.

The **knapsack heuristic** algorithm is used to restore the net position of the selected account into positive value. This is done by excluding some transactions from the solution which debit the current resolved account. To select the excluded transactions, debiting transactions are first divided into free base and cost base. The **free base** includes such transactions which do not cause or increase a negative net position on any other account. Transactions in the **cost base** are those which form a chain of settlement transactions; if such a transaction is excluded there are other consequent transactions that need to be excluded from the solution as well.

The implemented knapsack algorithm tries first to exclude transactions from the free base. If this is not sufficient, all transactions in the free base are excluded and selecting among the cost base is done. Selection is performed in either case with greedy search[13] according to the price/weight ratio of transactions. The price of a transaction is the share of negative net position, which it can remove from the account under study if the transaction is excluded. Weight is the number of equities in the transaction.

For transactions in the cost base Nyholm describes a recursive cost calculation method, which elaborates the chain of transactions and eventually aggregates the total number of equities in transactions to be excluded from each chain. The selection of transactions to be excluded in the cost calculation considers at each recursive level of the chain only the own weight of each transaction. The new negative net positions which may emerge deeper in the transaction chain are dealt with separately, which finally gives the aggregated cost of the whole chain of transactions.

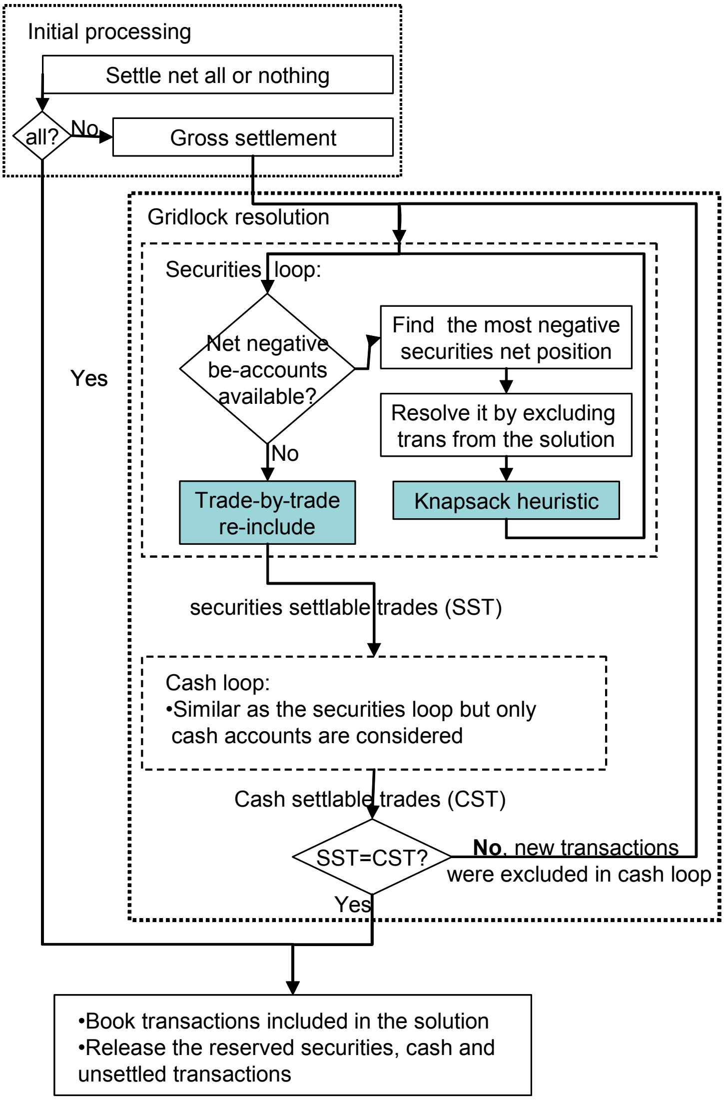
The algorithm implemented in BoF-PSS differs somewhat from this. It has a parameter for the maximum depth of cost calculation in the transaction chains. Up to this level costs are also calculated recursively for each decision. To limit the computational time of processing, the level of recursion was limited to one in actual

---

[13] See eg Martello and Toth (1990).

simulations, ie the weighting of cost base and free base transactions was identical.

Figure 10.3 **Overall structure of the optimisation process. Shaded steps of the process are explained in detail in the text.**



The **trade-by-trade re-include** routine is executed at the end of each securities and cash optimisation loop. It works through all individual transactions dropped out during the previous cycle and returns them to the solution if it is possible without causing negative net positions. This decreases the risk of dropping out transactions unnecessarily.

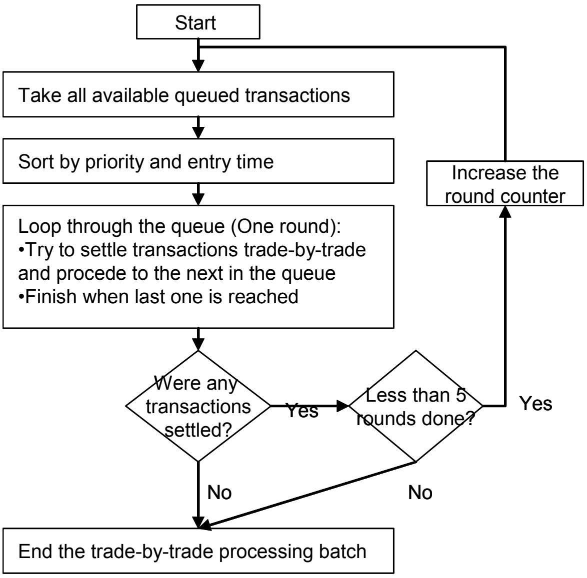
In real HEXClear, the same optimisation process is also used in preliminary information batches without cash constraints. Thus the result from preliminary information batches is twofold: first it gives the group of transactions that, which is settle-able and holds largest possible total number of securities. Secondly, it indicates the amount of liquidity required for settlement of this group. In the simulations no preliminary information batches were executed since their real life outcomes were known.

The trade-by-trade RTGS process

Trade-by-trade processing begins after the first optimisation cycle of the settlement day. In the real system it can also be run simultaneously with optimisations later in the day. To avoid conflicts in such cases, transactions included in the optimisation are marked and the gross amount of securities and cash is reserved for the duration of each optimisation.

Although the trade-by-trade process is called a real-time process, it is not actually executed continuously. It is run in batches and after one batch is finished the next starts five minutes later. The steps of the process are shown in Figure 10.4 below.

Figure 10.4 **Trade-by-trade process logic**



The sleeping time between the batches and the maximum number of rounds for workking through the queue in an individual batch are both design parameters in the real system. These are also variable parameters in the BoF-PSS2 implementation. If a similar system was used for money market transactions and collateral management of central banks, such delays and a lack of truly continuous settlement could be problematic for central bank purposes. It must be noticed, however, that some possibly more significant changes in the system setup would have to take place if it were used in operations related to central bank monetary policy, eg sychronisation of opening hours and days between the HEXClear and TARGET systems.

User modules in BoF-PSS2

In the simulator, HEXClear optimisation was implemented as a multilateral net settlement algorithm module (MNS) and RTGS queue release as a partial net settlement module (PNS). This allowed a timed start of the latter and the inclusion of both of these in same system setup.

The allocation data of HEXClear and the possibility that a single trade could have an arbitrary number of actual bookings had to be taken into consideration in the modelling. This was tackled by developing so-called group-code algorithms, which generalise the

DVP-link code structure of the simulator. These are presented in more detail in BoF-PSS2 manuals.[14]

Simplifications made in the modelling

Some simplifications were made in the implemented HEXClear model compared to the real system. This section presents the known differences. All these remaining simplifications are considered to have a very small impact on the results.

The restrictions and reservations that can be imposed on bookings or account balances were not implemented since their number in the real data was comparatively small. These include sell reservations, clearing reservations and restrictions on disposal of a purchase.[15] In terms of volume, only 0.34% of all allocation data involved some of these.

The optimisation features a heuristic Greedy knapsack-algorithm, not an exact algorithm. According to the Nyholm paper, smaller optimisation sub-problem instances are solved in HEXClear with exact code. Also the cost calculation in the knapsack was more restricted in the model, as discussed above.

The account balance of one book-entry account was manually altered on the basis of trade data and public information after it was noticed that some big trades did not settle in the benchmark simulation of a normal situation. We are unable to tell whether this is a sign of some larger systematic error in the data used or in the process where this data was transformed into simulator input.

Account transfers made in the central registry were not DVP-linked in the simulations due to a lack of required data fields. Because of this, the credit side can at times be executed without debiting real accounts.

In a normal situation, cash sub-accounts in HEXClear start each day with a zero balance. In the simulation these accounts were given a small starting balance (€100) to avoid settlement failures caused by rounded cash transaction values in the data.

---

[14] See BoF-PSS2 User manual, Section 4.2.2.

[15] See APK rules, General Definitions.

10.4 Scenarios and simulation results

Hypothetical scenarios were created on the basis of analysis of settlement system implementation and information flows of that time.[16] All scenarios were considered to be plausible under some extreme conditions and based on the known history of similar minor-scale disturbances.

The main scenarios were as follows:[17]

1. Securities settlement system as a **liquidity sink**: All payments to HEXClear are executed but no cash can be repatriated. The failure situation lasts one whole settlement day. This scenario measures the extreme level of contagion from HEXClear to the operations in BoF-RTGS or eventually TARGET.
2. Internal **operational failure of one individual clearing participant** for one whole day. This scenario describes a situation where all actions performed in the HEXClear user interface by the failing clearing participant are cancelled. Two durations of the assumed failure were tested: one and two whole settlement days.
3. **Failure of data connections between APK and VPC.** As a result all clearing parties which connect to HEXClear via VPC would be inoperable as in scenario 2.
4. Optimisation in settlement process is inoperable and the **backup trade-by-trade RTGS** process is used to settle all trades.

---

[16] After the observation period of this study, eg the connections from clearing participants to APK have been separated from internal VPC-APK data link.

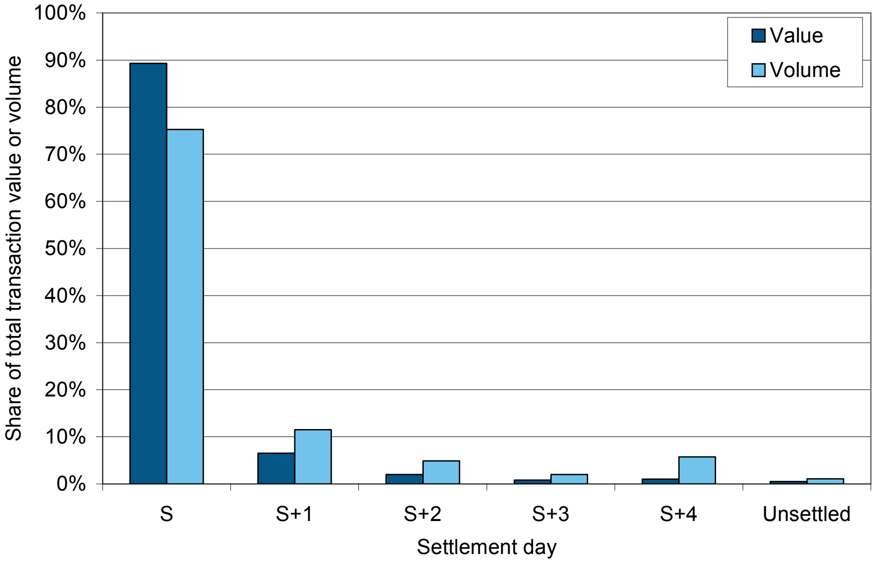
[17] Some other scenarios were also considered but not implemented during the current study. Scenario 2 could be altered to the failure of an individual broker or account operator rather than a clearing party. The specific points of impact of such a failure should be studied properly. Also, failure of one commercial bank in all the roles it has could be analysed (account operator, and several clearing parties or brokers in one conglomerate) or at least failure of all clearing party codes of one custodian. The failure of one commercial bank could also be analysed, focusing on the financing function which bigger banks provide to brokers outside the SSS. Similarly, as in the CSD's network failure scenario, multiple participants operating the post-trade infrastructure could be stressed simultaneously. For plausible scenarios, description of the whole technical structure supporting the information flows should be available. An example would be a case where several participants share a common connection provider and thus a common vulnerability to ICT problems.

10.4.1 Simulation of a normal situation

The simulated scenarios were analysed by comparing them with a benchmark simulation which mimics the real history of securities settlement without any form of stress factor. The benchmark data for one day included all the transactions that were included in the settlement process during that day, ie transactions with a 'waiting for settlement' (WFS) status.

Because of the simplifications in the model, the precision of the benchmark simulation formed in this study was not perfect. This was shown by the number of trades that did not settle during the correct intended day. Instead, some transactions were settled only during the following day or consequent days of the undisturbed simulation and a small number was always left unsettled. This is illustrated below (Figure 10.5) by presenting the final settlement day of the trades that were submitted during one example day from the benchmark simulation.

Figure 10.5 **When did the transactions submitted during one example day (13 June 2005) of benchmark simulation actually settle in terms of value or volume?**



In this example case the final settlement rate of the benchmark day is 99% (or 98%) in terms of volume (value), but a considerable proportion of this is achieved only during consequent days when a new set of transactions is included in the settlement. The simulations were limited to calendar weeks. Thus for some days there were more consequent days and, as a result, the settlement ratio of individual days in the benchmarks varied between 99% and 85% in terms of volume.

It is worth noting that not all trades involved in the settlement process in a given day settle during that same day in the real system either. The number of postponed trades in the benchmark simulation was still rather high in comparison to the performance of the real system.[18] Finding the reason for this difference was left for future studies. In the current study it is assumed that when the result of an imperfect benchmark is compared to scenario outcomes produced with the same logic, the difference gives a realistic picture of HEXClear performance in the overall level.

10.4.2 Liquidity sink scenario

In the stress-testing of HEXClear, the BoF-RTGS system was included in the model as a separate and independent system. Also, all the payments processed in BoF-RTGS were included in the simulations. The purpose was to enable measuring of contagion of liquidity shortages from one system into another at a realistic level: in the form of number or value of delayed or unprocessed payments.

The possible worst case of liquidity shortages in BoF-RTGS can be analysed from simple BoF-RTGS transactions. Assuming that all cash transfers to HEXClear are executed but none of the cash repatriations can be made for a one whole day (scenario 1), the maximum level of liquidity sink effect is reached – which could be caused by problems in HEXClear.

From the results of such a simulation it was noticed that no single transaction in BoF-RTGS was left unprocessed because of the HEXClear liquidity sink during the one-month period under study. During one day one single payment was delayed for 19 minutes.

One reason why payment delays or other impacts on the BoF-RTGS level were not observed is the daily schedule of payments to and from HEXClear. Most of the time clearing participants transferred the required cash to HEXClear in the morning before the first

---

[18] During a normal day in the real system the share of transactions transferred to next day is 0.2–0.3% of daily volumes.

optimisation and repatriated the final amount of cash only at the end of day. This was analysed earlier in the Bank of Finland financial stability report.[19] Thus, their general liquidity management during the day becomes separate from the cash sent to HEXClear and is not influenced by incidents in HEXClear as long as HEXClear does not require an addition of cash.

Based on this analysis a failure in HEXClear is highly unlikely to cause any systemic liquidity problems in BoF-RTGS.

10.4.3 Failure of a clearing participant in pre-settlement process

Failure of a clearing participant in HEXClear is modelled as a problem in the internal systems of the corresponding individual clearing participant. This would cause an inability to use the HEXClear user interface and cut out the information flow defining the allocation data of settlement transactions. For the sake of simplicity, only scenarios with failures lasting full calendar days were considered.

In this scenario it is assumed that fund transfers from BoF-RTGS to HEXClear are not hindered. This is a reasonable assumption because there are usually only very few of these payment instructions and in case of failure these could be handled as manual transactions or via other contingency arrangements, although there might be slight time delays in such a process.

The data set for the scenario was created by filtering out from the original benchmark data such transactions in which the failing participant was active and where the status of the transaction had changed to ‘waiting for settlement’ (WFS) during the assumed failure period. This change of status indicated a flow of some essential information that can hinder the settlement if the information is delayed or not processed at all in a failure situation.

Based on the analysis of timing of information flows related to the settlement process presented in Section 10.3.3, two separate durations for the failure situation were analysed: one and two whole banking days. Simulation results from these sub-scenarios are presented separately below.

---

[19] Bank of Finland Bulletin, Financial stability special issue 2004, p. 80.

Operational failure lasting one day

If the assumed operational failure lasted only one settlement day, the impact on settlement would be caused by cancelled intraday transactions and the consequences of their removal. The share of such transactions was shown to be 32% of the 5-transactions and 22% of all transactions in terms of the number of book-entries involved. In this scenario cash obligations and the resulting incoming funds are assumed to be unchanged.

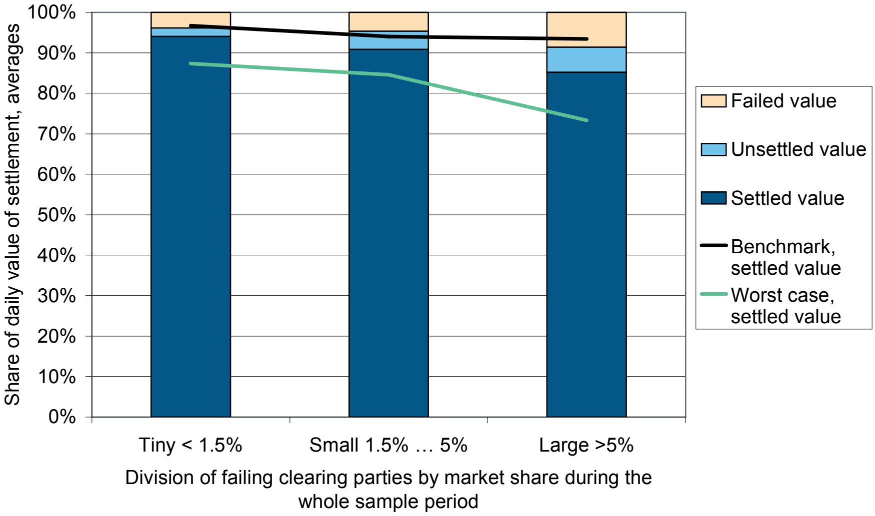
The impact of a failure of one clearing participant was tested with 55 independent scenarios. Each clearing party was stressed with data from one day and the ten biggest ones with three additional days. To illustrate the results, the failing clearing participants were divided into three groups based on their market shares[20] of the number of securities settled during the entire sample period: tiny ones having less than 1.5% each of total number of securities settled, large ones having more than 5% each and small ones accounting for the rest. Each group contains at least 4 different clearing participants and 16 or more independent scenarios.

The average outcome from the 55 simulated scenarios was shown in Figure 6 for each group of clearing parties. The failed transactions represent the initial impact of the assumed failure, ie filtered transactions. Unsettled transactions were included in the simulated settlement process but could not settle due to cash or securities shortages. The rest were settled normally. The figure represents the value of simulated transactions. This means it includes all the bookings with their face value and thus the number of securities and value of cash transfers in euros are aggregated into the same measure.

---

[20] This differs from the market shares presented earlier in Section 10.3.4, which were based on transaction volumes for each bank. One bank may operate in the system with several clearing party names.

Figure 10.6 **Impact of operational failure of one clearing party lasting one day on the value of unsettled transactions**



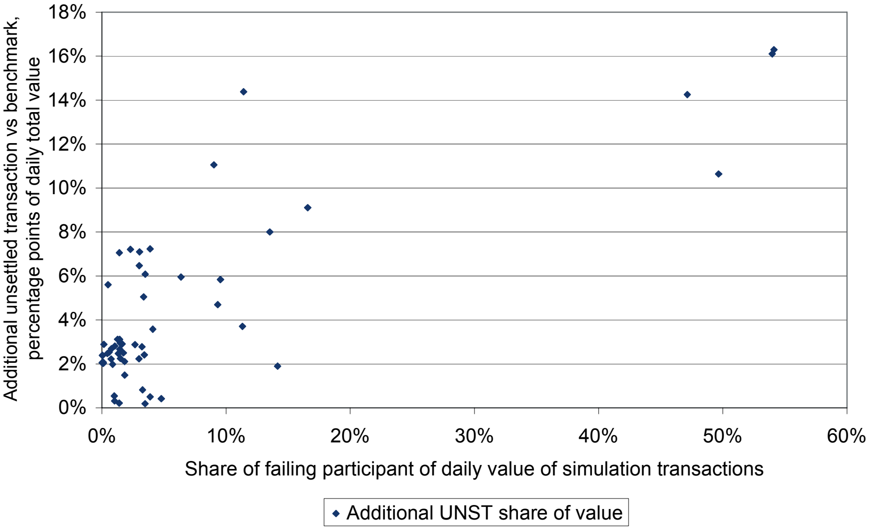
No extreme system-wide implications were observed in any of the simulated scenarios with failure lasting one day. In the worst individual cases, the share of settled transactions dropped to 85–88% for small participants and to 75% for large ones.

The comparison of the results with the benchmark is not very straightforward; it is not necessarily the same transactions that become unsettled in failure scenarios. The average overall share of settled transactions in the benchmark simulation is shown with a black line in the figure. The difference between this line and the blue bar gives the average decrease in the value of settled transactions.

At the level of individual scenarios, the decrease was at most 16% of the total daily value of simulated transactions. In transaction volumes, the impact was even more modest. The biggest increase in number of unsettled transactions was 3.2%.

The correlation of the size of the failing participant and the additional share of unsettled value in failure scenarios is presented below in Figure 10.7. Here the size of the failing participant stands for daily share of the value of simulated transactions of the participant failing in each particular scenario. Thus it again includes both cash and securities bookings. The correlation between the variables is evident, with the correlation coefficient at 0.76, but the variance is even more significant. Even within this limited set of observations, the impact of scenarios where participants of same size are failing varied, eg from almost 0 to 7% of total volumes for participants with roughly 5% of the daily share of value of bookings, and between 2% and 14.5% for participants with a roughly 10% share of the bookings. The reason for this instability in the outcome of the settlement should be studied further. It is unclear whether the true optimal solution for the knapsack selection problem of trades to be settled is really very sensitive to initial conditions or whether the heuristic algorithm implemented and used in the simulations is sensitive to initial conditions and sometimes incapable of reaching optimal or near-optimal solutions.

Figure 10.7 **Correlation of the daily share of a failing participant of simulation transaction values and the increase in value of unsettled transactions. Operational failure of one clearing participant lasting one day.**



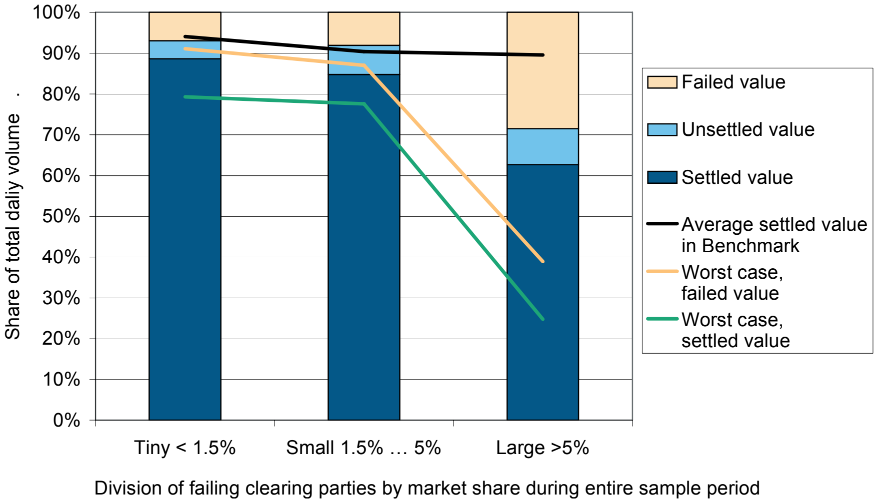
Operational failure lasting two days

When a failure situation is extended to two bank days, the setup differs from the previous one. This is because the settlement process in HEXClear can adapt to the failure and calculate the cash obligations for clearing participants in the preliminary information batches for the altered set of settlement trades. Cash shortages should therefore not emerge, but cancelled trades can cause securities shortages and consequent failures in settlement. The increase in liquidity need was modelled by giving sufficient large balances for the cash accounts to make sure no cash shortages would arise in settlement.

The settled value of payments in this scenario is presented below in Figure 10.8. Critical observation of the results indicates that an internal failure of some individual clearing participants can cause failure of transactions totalling 30% or even 60% of total value if that failure continues to the next settlement day. On the basis on our exercise, internal systems of the largest clearing participants should be considered critical to the operations of HEXClear as a whole.

It can be noticed that the optimisation of HEXClear efficiently mitigates the additional impact of failure in this kind of failure situation. The value of transactions remaining unsettled in the process did not increase in proportion to the failing participant's market share and value of failed transactions.

Figure 10.8 **Operational failure of one clearing participant. Impact on settled value after adapting cash positions to changes in settled transactions.**



When transaction volumes were analysed instead of values, the same structure was observed again. However, the contrast between the impact of large clearing parties and small or tiny ones was sharper. The share of the failed volume in one individual scenario was 53% of total volume while the median volume of failed transactions from all observations was only 2%.

If settlement on the second day was executed on the basis of preliminary information of trades used in the normal unstressed case, ie without adapting the cash positions, shortages of cash would obviously be faced in the stressed settlement. The additional share of unsettled transactions caused by this assumption varied between 0.1% and 2.4% of the total value of transactions. In terms of volume, the number of settled transactions was sometimes even higher with original cash positions. This can be explained by the objective of the optimisation, which seeks to maximise number of equities settled (ie the value of settlement) not the number of trades settled (ie the volume).

10.4.4 Failures in data connections between the CSDs

The scenario of failure in CSD's data connections is similar to the one replicating an operational failure of an individual participant in the pre-settlement process presented above. The difference is that several clearing participants are excluded from the settlement simultaneously. This is possible since during the sample period several clearing participants were known to connect to APK systems via an internal link between VPC and APK. Thus, this connection in itself involved a bigger node risk in relation to the HEXClear process than the internal systems of individual participants.

The combined market share of clearing participants using CSD's connection alternative was 13% of the total value of settlement transactions in the sample period. The scenario was tested in 14 simulations: one- and two-day operational failure starting on seven separate days. The results are summarised below in Table 10.2.

Table 10.2 **CSD's data connection failure, average scenario results**

Values	Filtered	Unsettled	Settled	Change in share of settled compared to benchmark
One-day failure	6.6%	6.6%	86.8%	-4.97%
Two-day-failure	23.0%	7.4%	69.6%	-19.67%

Values	Filtered	Unsettled	Settled	Change in share of settled compared to benchmark
One-day failure	0.4%	3.5%	96.1%	-0.70%
Two-day-failure	14.8%	5.4%	79.8%	-16.27%

When the results are compared to the average failure of individual large participants presented in Sections above, the impact of CSD's network failure was smaller. Thus settlement transactions from participants using this connection alternative seem to be less connected with the rest of the settled transaction volume.

10.4.5 Settlement with the RTGS trade-by-trade process

In the daily schedule of HEXClear, the RTGS process begins after the first optimisation. It is supposed to serve as a swift intraday settlement of T+0 trades and settlement transactions. It can also be considered as a backup process that can be used instead of optimisation if necessary. The impact of using the RTGS process as a contingency arrangement was analysed.

Liquidity need of RTGS process

The HEXClear system's optimisation process saves liquidity because without optimisation the netting effect of the simultaneous settlement of transactions would be lost. The additional liquidity needed in trade-by-trade settlement was quantified and compared to the amount of liquidity actually tied to settlement. The compared value from trade-by-trade settlement was the maximum total daily amount of liquidity tied to settlement at a given moment.

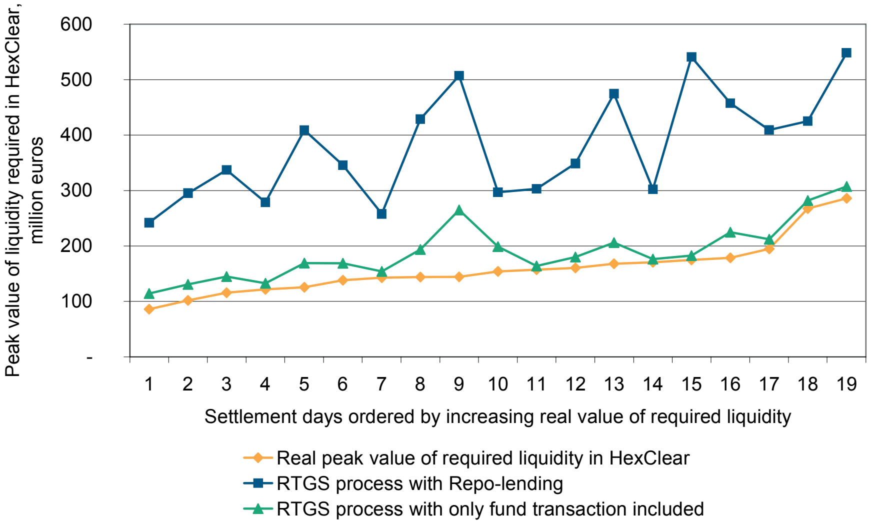
The order in which transactions were settled corresponded to a situation in which the transactions would be submitted to APK's trade-by-trade process if the optimisation process is not available at all. In reality, the order of settlement may differ slightly due to parallel processing available in APK, which the simulator cannot replicate. In trade-by-trade settlement the order of transactions is essential because it may have a key impact on both the liquidity needs and on the outcome of the settlement process.

The removal of the netting impact of the optimisation also affects securities accounts. To cover these shortages the clearing parties in the simulation were given the opportunity to borrow securities from efficient repo markets at the market value of the securities. In the Finnish direct holding structure, shortages can become significant because trades are settled directly on the accounts of the individual investors. In reality, there is a variety of securities lending products available. Moreover, the availability of individual lending facilities differs by type of security and size of the investor. An efficient securities lending system promotes smooth post-trade processing. However, a shortage of securities cannot be covered free of charge, and thus the use of the system usually has at least a marginal cost effect.

The results show that an efficient clearing and settlement method may require close to €400 million less liquidity per day. This is a relatively large amount since the average of total liquidity tied to HEXClear was €160 million during the sample period. In trade-by-trade settlement, the order of transactions may momentarily cause a major shortage of securities, which explains large individual variations in the efficiency of clearing and settlement methods from one day to another.

When the securities shortages were not taken into consideration, the difference between the amount of liquidity required by various settlement methods was generally smaller. Daily liquidity savings amounted to 21% on average, with the savings rising to 83% on the most extreme settlement day.

Figure 10.9 **Liquidity requirement of the RTGS process compared to HEXClear**



The results were achieved by comparing the actual amount of liquidity tied to equities clearing and settlement to a simulated situation where trades are settled trade by trade. In practice the liquidity requirement in the trade-by-trade process would be even higher because cash would probably be transferred to the settlement system in bulk amounts to allow a smooth trade-by-trade settlement process. Another alternative would be to make the bookings of fund transfers directly on BoF-RTGS accounts of the clearing parties, as in the interfaced model.

Settlement delays and cascading settlement chains in the RTGS process

The RTGS process is actually currently executed in batches. In one batch, only five consequent settlement rounds are executed, as was discussed in section 10.3.2. In the settlement transactions typically form chains where one trade can only be settled if another is settled first. If the entire volume of transactions is processed in a gross process, five rounds can be inadequate to work through all such chains and may cause unnecessary delays when the process is sleeping.

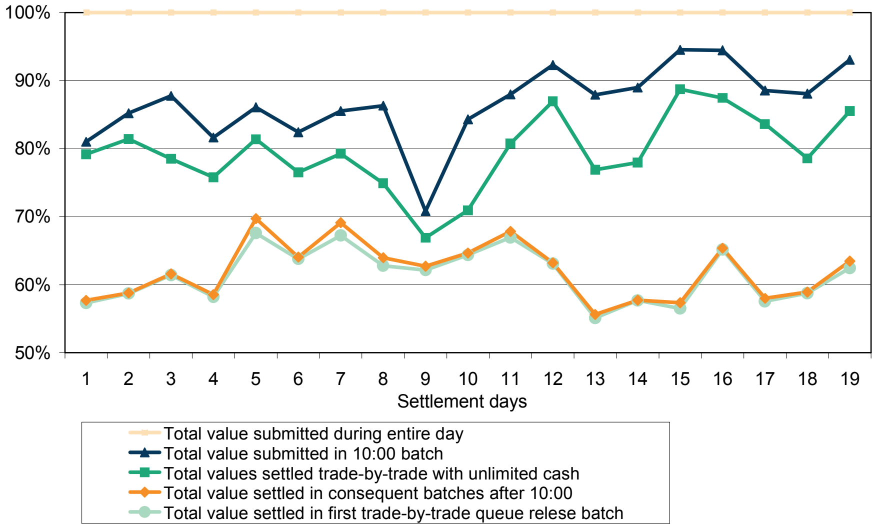
The length of the settlement chains was studied by forcing all trades to the RTGS process. The set of trades included in this analysis was those trades that were submitted and ready to be settled at 10.00 am on each day of the sample period. These trades would have entered the first optimisation in a normal situation. After simulating the settlement with the RTGS process, the number of consequent batches and number and value of transactions settled on each batch was observed.

As an initial assumption the cash amount available was the same as in the case of normal settlement with optimisation. From the results it can be observed that gridlocks will emerge with this liquidity level in trade-by-trade settlement; on average only 72% of the value and 89% of the transactions included in the first batch were settled during the trade-by-trade queue release process. It can also be noticed that a large majority of those trades which can be settled will settle immediately in the first batch, ie during first five queue-release rounds. Only 0.94% of the value and 0.70% of the volume of trades was left for the consequent batches.

The average number of queue-release batches needed was 3.84. When the 5-minute sleep period is taken into account, the average number of batches would cause more than 14 minutes of additional delay for the few transactions settled in the last round. The average number of individual queue-release rounds needed was 19.21, which can be interpreted as the average length of the longest cascading settlement chain in the settlement transactions. However, the expected value for sleep delay of individual transactions was only 2.7 seconds, with the condition that it will eventually settle in the trade-by-trade process. Thus the number of rounds and possible delay caused is marginal compared to trades that are left unsettled in trade-by-trade processing.

For comparison, a case with an abundant cash amount was analysed. In this setup only those settlement chains are included, which are caused by securities leg of the settlement. The average number of batches needed (3.79) was only slightly smaller in this comparison. The number of gridlocks was, however, decreased significantly since 91.8% of the value and 97.8% of the volume was eventually settled with the trade-by-trade process.

Figure 10.10 **Share of daily trade values settled in different trade-by-trade scenarios**



The scenario results of the daily share of volume settled in different scenarios are visualised in Figure 10.10. The thick dark blue line represents the outcome of the scenario with the initial assumption, ie with original cash positions. The small difference between this line and the lowermost thin red line represents the share of transactions which get delayed due to sleeping times in the RTGS process. The difference between the dark blue line and magenta line show the size of gridlock – the value of transactions unsettleable for the RTGS process with original cash positions. Removing cash constraints shrinks the total value of the gridlocked transactions. This is shown by the thin cyan line representing the total value of transactions settled with the RTGS process given unlimited cash.

Increasing the share of transactions included in trade-by-trade process seemes to increase the share of settled only when cash constraints are removed. Also worth noticing are the big differences between days. A similar figure for trade volumes would show smaller variations. There the lowest value of settlement ratio with the original amount of cash would be 82% and with unlimited cash 93%.

10.5 Conclusions

This article focuses on stress-testing of APK's HEXClear system with simulations and presents the model constructed for the scenarios.

The data from one month in 2005 was studied directly to establish a good picture of the market structure of securities clearing and settlement. A high concentration of post-trade processing was observed in the system. The concentration was strongest in the settlement of off-exchange transactions, while the settlement of exchange trades is more evenly distributed between the clearing parties. An examination the information flows of the settlement process showed that 53% of securities belong to transactions that receive the final necessary information for settlement during S-1, ie the day before settlement day. Similarly, 33% receive final information only during the actual settlement day. Most transactions finalised at a late stage are off-exchange transactions. High concentration rates and timing of information flows close to the settlement day increase the potential impact of operational failure scenarios on the settlement system.

The robustness of the settlement model was tested with four simulated main scenarios. It was observed that, due to the cash model of HEXClear and current market practices, it is highly unlikely that failures in securities settlement could disturb the smooth functioning of other systems in the Finnish market infrastructure. Within HEXClear, the failure situation of individual participants should endure for more than one settlement day to become critical. An internal technical failure of some individual clearing participants lasting two days was found to be capable of causing the failure of transactions of up to 60% or 70% of the total value of settlement. Also only the failure of major clearing participants showed a striking impact on settlement. Another scenario showed that operational failures in the ICT link between VPC and APK would have caused considerable stress on settlement in HEXClear.

In all of the simulated cases the largest impact was caused by transactions filtered out due to a market participant’s assumed internal operational problem. Most transactions included in the settlement were eventually settled. This underlines the ability of HEXClear to efficiently mitigate failure situations via the optimisation settlement process.

Replacing optimisation with an RTGS process would significantly increase liquidity need. It was proven that most transactions would still settle without longer delays unless they are involved in a gridlock on the securities leg side. We also made some observations regarding enhancements that central bank user needs would require if HEXClear was the only SSS used in Finland.

Due to developments made in the current study, BoF-PSS2 now includes features that allow modelling of retail securities settlement systems, such as HEXClear. This capability and the model of current study could be utilised, for example, if the potential effects of the Nordic Single settlement system planned by the Nordic Central Securities Depository[21] were to be analysed.

The question of how close to the exact optimal solution of the knapsack problem the heuristic optimisation settlement process implemented in BoF-PSS2 can come with a severely stressed data set is left for future studies. Our evaluation of HEXClear, although profound, cannot be fully conclusive since the benchmark simulation proves that we have not been able to completely replicate the HEXClear algorithm at this stage. Similarly the usability of the RTGS process as a contingency depends on the capability of that process to handle large transaction volumes in limited time. The real algorithms should thus be tested in a load situation in APK's test environment to achieve completely realistic results. Also in that approach the time complexity of the implementation could be measured.

All the scenarios analysed in this study were bound to the data from the observation period. The real system allows participants to adapt and react to failure situations, but no such changes were assumed or included in the simulations. Behavioural changes on the part of clearing participants made during the normal process or as reactions to abnormal situations may have feedback effects on the systemic level and change the overall results of this study.

The analysis presented only focuses on HEXClear but can be interpreted more generally as an analysis of the Finnish market structure and market practices. It is intended as a contribution to the oversight and development work of securities settlement systems by illustrating the dependencies, vulnerabilities and scale of events that may take place in such infrastructure.

---

[21] The Nordic Central Securities Depository (NCSD) was created through the consolidation of Swedish VPC and Finnish APK at the end of 2004. VPC bought 100% of the shares of APK, and this new company operates under the trademark of NCSD.

References

Bank of Finland (2006) **Bank of Finland Bulletin, Financial stability special issue.** 2006.

Bank of Finland (2005) **Bank of Finland Bulletin, Financial stability special issue.** 2005.

Bank of Finland (2006) **BoF-PSS2 user manual version 2.2.0.** 2006

Bank of International Settlements (1992) **Delivery versus payment in securities settlement systems.** Report by CPSS.

ECB (2004) **The use of central bank money for settling securities transactions.** May 2004. http://www.ecb.int/pub/pdf/other/useofcbmoneyforssten.pdf.

Devriese, J – Mitchell, J (2006) **Liquidity risk in securities settlement.** Journal of Banking & Finance, Vol. 30, Issue. 6, 1807–1834.

Hellqvist, M – Koskinen, J (2005) **Stress testing securities clearing and settlement systems using simulations.** In Leinonen, H (ed.) Liquidity, risks and speed in payment and settlement systems – a simulation approach. Bank of Finland Studies E:31.

Iori, G (2004) **An analysis of systemic risk in alternative securities settlement architectures.** European Central Bank Working Paper No. 404.

Martello, S – Toth, P (1990) **Knapsack problems – algorithms and computer implementations.** John Wiley & Sons, New York.

Nyholm, T (2004) **Optimization Algorithm for Securities Settlement.** Master's Thesis, Helsinki. University of Technology, System Analysis Laboratory. (In Finnish.)

Papadimitriou, C (1994) **Computational complexity.** 1st edition. Addison-Wesley Publishing Company, New York.

Sveriges Riksbank (2004) **Assessment of securities settlement systems in Sweden.** November 2003 with revisions to Standard 16 from June 2004. http://www.riksbank.se/upload/Dokument_riksbank/Kat_AFS/assessment_securities_settlement.pdf

Rules of APK and the decisions of related to the rules. http://www.apk.fi/473_ENG_ST.html.

Bank of Finland Publications

Scientific monographs

Series E (ISSN 1238-1691, print) (ISSN 1456-5951, online)

(Series E replaces the Bank of Finland's research publications series B, C and D.)

E:1 Jukka Vesala **Testing for Competition in Banking: Behavioral Evidence from Finland.** 1995. 206 p. ISBN 951-686-447-3.

E:2 Juha Tarkka **Approaches to Deposit Pricing: A Study in the Determination of Deposit Interest and Bank Service Charges.** 1995. 166 p. ISBN 951-686-457-0.

E:3 Timo Tyrväinen **Wage Determination, Taxes, and Employment: Evidence from Finland.** 1995. 212 p. ISBN 951-686-459-7.

E:4 Sinimaaria Ranki **Realignment Expectations in the ERM: Causes and Measurement.** 1996. 164 p. ISBN 951-686-507-0.

E:5 Juhana Hukkinen **Kilpailukyky, ulkomaankaupan rakenne ja taloudellinen kasvu** (Competitiveness, structure of foreign trade and economic growth). 1996. 134 p. ISBN 951-686-512-7.

E:6 Eelis Hein **Deposit Insurance: Pricing and Incentives.** 1996. 120 p. ISBN 951-686-517-8.

E:7 Vesa Vihriälä **Banks and the Finnish Credit Cycle 1986–1995.** 1997. 200 p. ISBN 951-686-537-2.

E:8 Anne Brunila **Fiscal Policy and Private Consumption-Saving Decisions: European Evidence.** 1997. 147 p. ISBN 951-686-558-5. (Published also as A-131, Helsinki School of Economics and Business Administration, ISBN 951-791-225-0, ISSN 1237-556X)

E:9 Sinimaaria Ranki **Exchange Rates in European Monetary Integration**. 1998. 221 p. ISBN 951-686-564-X.

E:10 Kimmo Virolainen **Tax Incentives and Corporate Borrowing: Evidence from Finnish Company Panel Data.** 1998. 151 p. ISBN 951-686-573-9. (Published also as A-137, Helsinki School of Economics and Business Administration, ISBN 951-791-290-0, ISSN 1237-556X)

E:11 Monica Ahlstedt **Analysis of Financial Risks in a GARCH Framework**. 1998. 181 p. ISBN 951-686-575-5.

E:12 Olli Castrén **Fiscal-Monetary Policy Coordination and Central Bank Independence**. 1998. 153 p. ISBN 951-686-580-1.

E:13 Antti Ripatti **Demand for Money in Inflation-Targeting Monetary Policy**. 1998. 136 p. ISBN 951-686-581-X.

E:14 Risto Koponen – Kimmo Soramäki **Intraday Liquidity Needs in a Modern Interbank Payment System. A Simulation Approach.** 1998. 135 p. ISBN 951-686-601-8.

E:15 Liisa Halme **Pankkisääntely ja valvonta. Oikeuspoliittinen tutkimus säästöpankkien riskinotosta** (Banking regulation and supervision: A legal policy study of risk taking by savings banks). 1999. XLIV + 560 p. ISBN 951-686-606-9, print; ISBN 951-686-607-7, online.

E:16 Juha Kasanen **Ilmoitusvelvollisten osakeomistus ja -kaupat Helsingin Pörssissä** (Corporate insiders shareholdings and trading on the HEX Helsinki Exchanges). 1999. 146 p. ISBN 951-686-630-1, print; ISBN 951-686-631-X, online.

E:17 Mikko Spolander **Measuring Exchange Market Pressure and Central Bank Intervention.** 1999. 118 p. ISBN 951-686-645-X, print; ISBN 951-686-646-8, online.

E:18 Karlo Kauko **The Microeconomics of Innovation: Oligopoly Theoretic Analyses with Applications to Banking and Patenting.** 2000. 193 p. ISBN 951-686-651-4, print; ISBN 951-686-652-2, online. (Published also as A-166, Helsinki School of Economics and Business Administration, ISBN 951-791-442-3, ISSN 1237-556X)

E:19 Juha Kilponen **The Political Economy of Monetary Policy and Wage Bargaining. Theory and Econometric Evidence.** 2000. 180 p. ISBN 951-686-665-4, print; ISBN 951-686-666-2, online.

E:20 Jukka Vesala **Technological Transformation and Retail Banking Competition: Implications and Measurement.** 2000. 211 p. ISBN 951-686-695-6, print; ISBN 951-686-696-4, online. (Published also as A-184, Helsinki School of Economics and Business Administration, ISBN 951-791-518-7, ISSN 1237-556X)

E:21 Jian-Guang Shen **Models of Currency Crises with Banking Sector and Imperfectly Competitive Labor Markets.** 2001. 159 p. ISBN 951-686-711-1, print; ISBN 951-686-712-X, online.

E:22 Kari Takala **Studies in Time Series Analysis of Consumption, Asset Prices and Forecasting.** 2001. 300 p. ISBN 951-686-759-6, print; ISBN 951-686-760-X, online.

E:23 Mika Kortelainen **Edge: a model of the euro area with applications to monetary policy.** 2002. 166 p. ISBN 952-462-001-4, print; ISBN 952-462-002-2, online. (Published also as A-204, Helsinki School of Economics and Business Administration, ISBN 951-791-715-5, ISSN 1237-556X)

E:24 Jukka Topi **Effects of moral hazard and monitoring on monetary policy transmission.** 2003. 148 p. ISBN 952-462-031-6, print; ISBN 952-462-032-4, online.

E:25 Hanna Freystätter **Price setting behavior in an open economy and the determination of Finnish foreign trade prices.** 2003. 84 p. ISBN 952-462-045-6, print; ISBN 952-462-046-4, online.

E:26 Tuomas Välimäki **Central bank tenders: three essays on money market liquidity auctions.** 2003. 232 p. ISBN 952-462-051-0, print; ISBN 952-462-052-9, online. (Published also as A-218, Helsinki School of Economics, Acta Universitatis Oeconomicae Helsingiensis, ISBN 951-791-762-7, ISSN 1237-556X)

E:27 Heikki Hella **On robust ESACF identification of mixed ARIMA models.** 2003. 159 p. ISBN 952-462-112-6, print; ISBN 952-462-113-4, online.

E:28 Heiko Schmiedel **Performance of international securities markets.** 2004. 275 p. ISBN 952-462-132-0, print; ISBN 952-462-133-9, online.

E:29 Tuomas Komulainen **Essays on financial crises in emerging markets.** 2004. 173 p. ISBN 952-462-140-1, print; ISBN 952-462-141-X, online.

E:30 Jukka Vauhkonen **Essays on financial contracting.** 2004. 134 p. ISBN 952-462-172-X, print; ISBN 952-462-173-8, online.

E:31 Harry Leinonen (ed.) **Liquidity, risks and speed in payment and settlement systems – a simulation approach.** 2005. Compilation. 350 p. ISBN 952-462-194-0, print; ISBN 952-462-195-9, online.

E:32 Maritta Paloviita **The role of expectations in euro area inflation dynamics.** 2005. 88 p. ISBN 952-462-208-4, print; ISBN 952-462-209-2, online.

E:33 Jukka Railavo **Essays on macroeconomic effects of fiscal policy rules.** 2005. 150 p. ISBN 952-462-249-1, print; ISBN 952-462-250-5, online.

E:34 Aaron Mehrotra **Essays on Empirical Macroeconomics.** 2006. 243 p. ISBN 952-462-290-4, print; ISBN 952-462-291-2, online.

E:35 Katja Taipalus **Bubbles in the Finnish and US equities markets.** 2006. 123 p. ISBN 952-462-306-4, print; ISBN 952-462-307-2, online.

E:36 Laura Solanko **Essays on Russia's Economic Transition.** 2006. 133 p. ISBN 952-462-316-1, print; ISBN 952-462-317-X, online.

E:37 Mika Arola **Foreign capital and Finland Central government's first period of reliance on international financial markets, 1862–1938.** 2006. 249 p. ISBN 952-462-310-2, print; ISBN 952-462-311-0, online.

E:38 Heli Snellman **Automated Teller Machine network market structure and cash usage.** 2006. 105 p. ISBN 952-462-318-8, print; ISBN 952-462-319-6, online.

E:39 Harry Leinonen (ed.) **Simulation studies of liquidity needs, risks and efficiency in payment networks.** 2007. Proceedings from the Bank of Finland Payment and Settlement System Seminars 2005–2006. 320 p. ISBN 978-952-462-360-5, print; ISBN 978-952-462-361-2, online.

ISBN 978-952-462-360-5
ISSN 1238-1691

Edita Prima Oy
Helsinki 2007