Available online at www.sciencedirect.com
ScienceDirect
Physica A 384 (2007) 693–718
PHYSICA A
www.elsevier.com/locate/physa

Congestion and cascades in payment systems

Walter E. Beyeler[a,*], Robert J. Glass[a], Morten L. Bech[b], Kimmo Soramäki[c]

[a]*Sandia National Laboratories*[1]*, Albuquerque, NM, USA*
[b]*Federal Reserve Bank of New York, USA*
[c]*Helsinki University of Technology, Finland*

Received 28 July 2006; received in revised form 23 February 2007
Available online 2 June 2007

Abstract

We develop a parsimonious model of the interbank payment system. The model incorporates an endogenous instruction arrival process, a scale-free topology of payments between banks, a fixed total liquidity which limits banks' capacity to process arriving instructions, and a global market that distributes liquidity. We find that at low liquidity the system becomes congested and payment settlement loses correlation with payment instruction arrival, becoming coupled across the network. The onset of congestion is evidently related to the relative values of three characteristic times: the time for banks' net position to return to 0, the time for a bank to exhaust its liquidity endowment, and the liquidity market relaxation time. In the congested regime settlement takes place in cascades having a characteristic length scale. A global liquidity market substantially attenuates congestion, requiring only a small fraction of the payment-induced liquidity flow to achieve strong beneficial effects.

Keywords: Network; Topology; Interbank; Payment; Money market; Sandpile model; Congestion

1. Introduction

Modern economies depend on efficient and reliable financial markets. Critical to the smooth functioning of these markets is a set of trading, payment, clearing and settlement infrastructures. Financial infrastructures are formed by a large number of technological and institutional components that interact within complex networks.

One core infrastructure is the interbank payment system which allows movement of funds between banks. Fund transfers may be related to transactions originating from money, foreign exchange or securities markets. The Fedwire Funds Service operated by the US Federal Reserve, for example, processes more than five

*Corresponding author. Tel.: +1 505 8445212; fax: +1 505 2843850.
E-mail address: webeyel@sandia.gov (W.E. Beyeler).

[1]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. The National Infrastructure Simulation and Analysis Center (NISAC) is a program under the Department of Homeland Security's (DHS) Preparedness Directorate. Sandia National Laboratories (SNL) and Los Alamos National Laboratory (LANL) are the prime contractors for NISAC under the programmatic direction of DHS's Infrastructure Protection/Risk Management Division.

0378-4371/$ - see front matter
doi:10.1016/j.physa.2007.05.061

hundred thousand payments daily with a total value exceeding $2 trillion [1]. The European TARGET system handles similar volumes in euros [2]. Fund transfers generally take place on the books of a central bank. In such systems the transfers take place in real time and funds received can immediately be used to effect further payments. This reuse allows the system to settle payments using only a small fraction of the daily turnover. For example, the daily flow of roughly $2 trillion in Fedwire is supported by a total daily account balance of approximately $15 billion [1] and an average daily overdraft of approximately $36 billion [3].

Participants have an economic incentive to minimize the funds committed to payment processing because liquidity used for settling payments imposes an opportunity cost on banks. Under-funding can also be costly, especially for bank customers and other banks in the system. Shortfalls of funds can delay a bank's payment processing, and payment systems can even enter gridlock states in which no bank can process a payment [4]. Delayed payments are unavailable to intended recipients: in this way congestion in the payment system can propagate into the economy by restricting money flow among banks and eventually among their customers.

Large-scale simulations of payment systems have been used to evaluate the possible consequences of changes in payment system rules and policies under both normal and disrupted conditions. A number of such studies are available in Leinonen [5]. These simulations have used detailed descriptions of the business rules followed by the diverse participants, including banks and system operators, to anticipate the response of specific systems to potential stresses.

In this paper we develop a parsimonious model of the interbank payment system to study congestion and the role of liquidity markets in alleviating congestion. This model focuses on the essential dynamics of payment processing in order to understand how networks of interacting agents, each following simple rules, can give rise to system-level congestion. The main features are an endogenous instruction arrival process, a scale-free topology of payments between banks as found in real interbank payment systems [6,7], a fixed total liquidity used by banks to process arriving instructions, the ability of banks to build and work off queues of instructions they cannot process, and a global market that distributes liquidity among the banks. Because we focus on the influence of liquidity and liquidity distribution mechanisms on system performance, we assume that all banks follow a cooperative strategy in submitting payments.

Financial relationships among individual decision-makers are increasingly represented using network models. One thread of research, for example [8–11], models price formation in a market made by agents responding to the behavior of their immediate neighbors in an influence network. Another thread [12–16] examines the flow of diverse goods and services between producers and consumers through a set of intermediaries, where the network links model pairwise specifications of cost and information among the individual decision-makers. In these and similar studies, the operation of the payment systems that undergird individual financial transactions is presumed. The purpose of the present study is instead to focus on payment system operations, using a stipulated forcing function to create the payment flows that express underlying economic relationships

Network models of queuing agents have been studied in many contexts, including manufacturing processes and supply chains, computer networks, and transportation infrastructures [13,16–18]. In these models, interactions between agents generally represent transfer of workload from agent to agent. In contrast, interactions between banks in payment systems transfer *capacity* rather than workload: a bank that sends many payments and sends them promptly tends to relieve rather than create congestion at receiving banks. Congestion in a payment system is both a cause and a consequence of reduced transfer capacity.

Our model has been developed in the spirit of abstract models used to study critical behavior. Bak et al. [19] discovered self-organization in systems of locally interacting elements with non-linear dynamics. In their pioneering model random stresses impinge on a lattice of coupled elements, which discharge when their state variable exceeds a threshold. Stress is dissipated at the lattice boundaries. This system is driven to a state characterized by discharge cascades at all length scales. Sachtjen et al. [20] studied the effect of several stylized topologies in systems consisting of elements with similar threshold discharge behavior, but that were stressed by random bilateral exchanges between pairs of elements connected by links. This system undergoes a transition in a tolerance parameter, below which the average cascade size becomes unbounded. It may be tuned to a critical state but, unlike Bak et al., is not driven to one.

We find that at sufficiently low liquidity our system becomes congested and payment settlement loses correlation with payment arrival. At low liquidity banks' payment processing becomes coupled creating