

SparkR Bitcoin Demo

Bernhard Haslhofer
AIT - Austrian Institute of Technology

Vienna R Meetup
2016-01-26

What is Bitcoin?

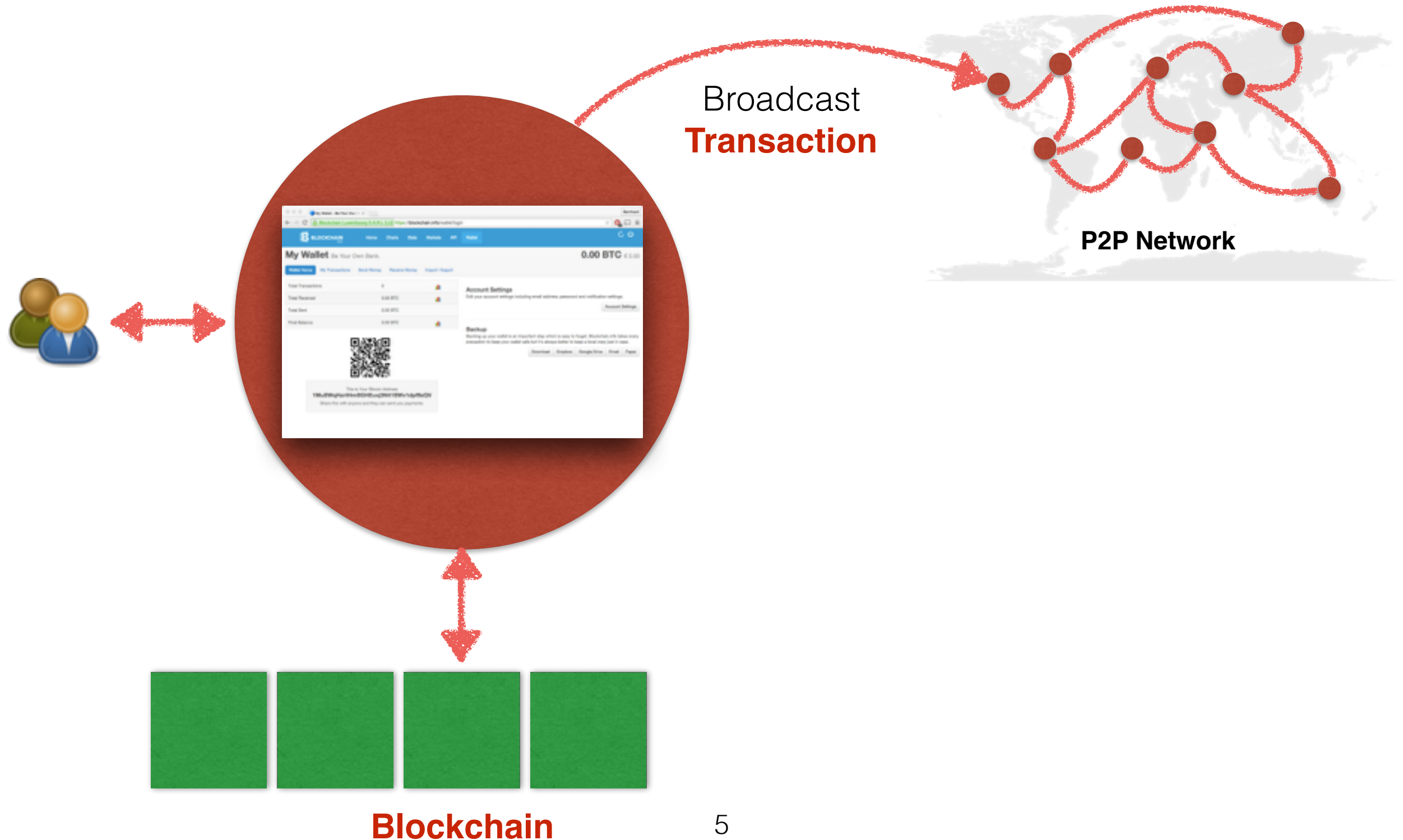
A virtual currency

- Difference to other currency systems:
 - no trusted third parties
 - no pre-assumed identities
 - no central authority
 - collective transaction management (block chain)
 - collective money issuance (mining)
- Also known as “cryptocurrency”

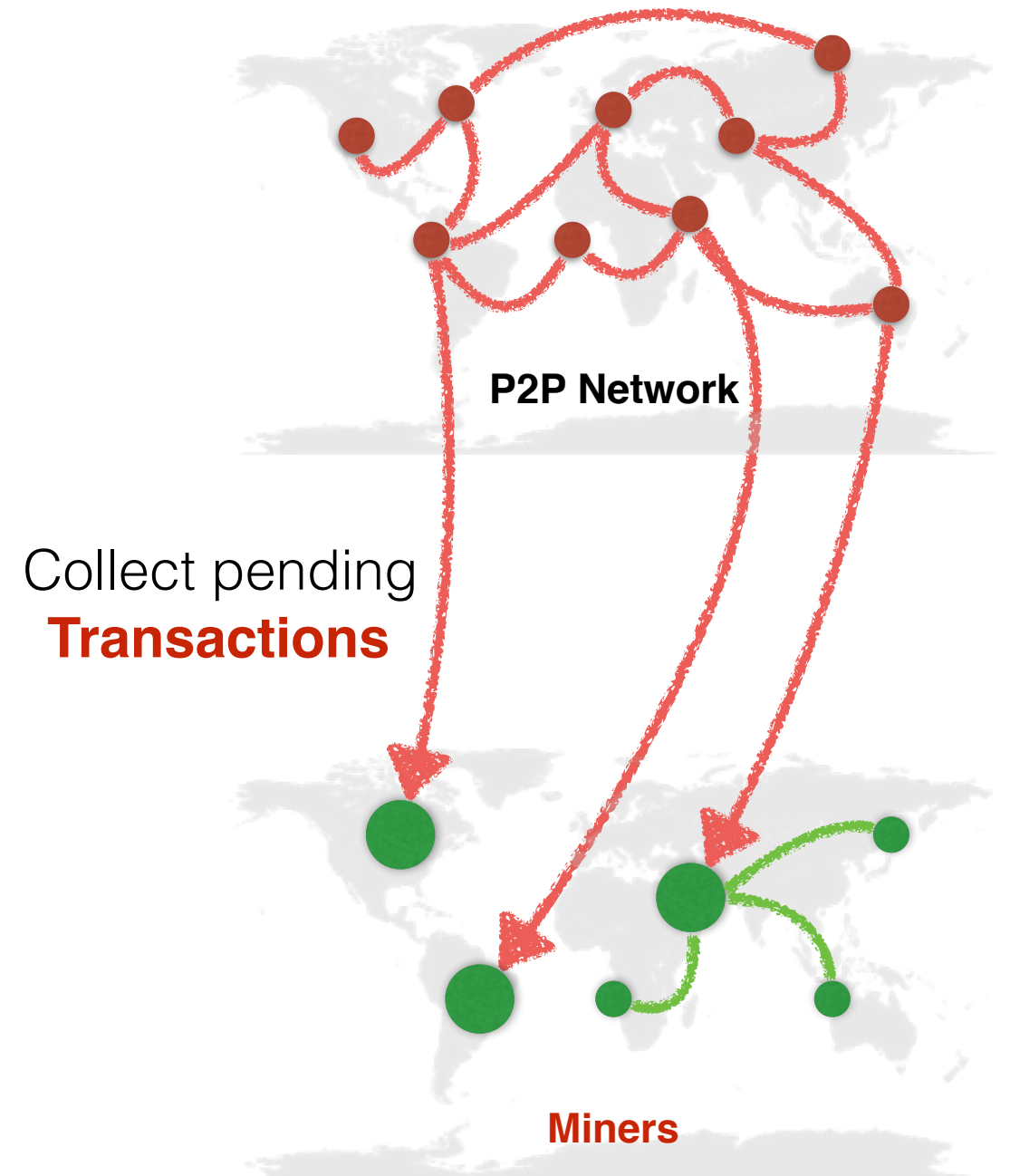
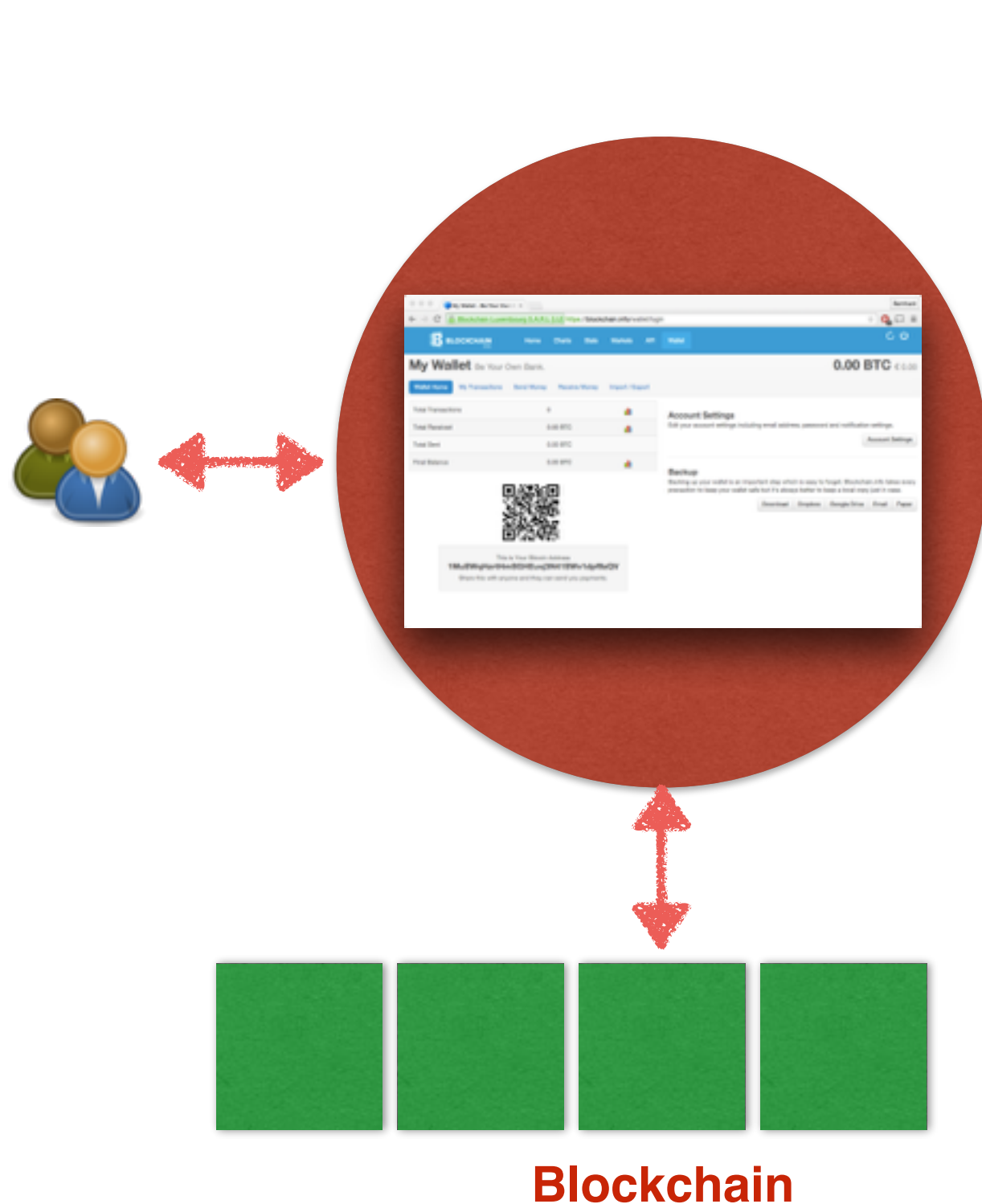


Transaction Processing

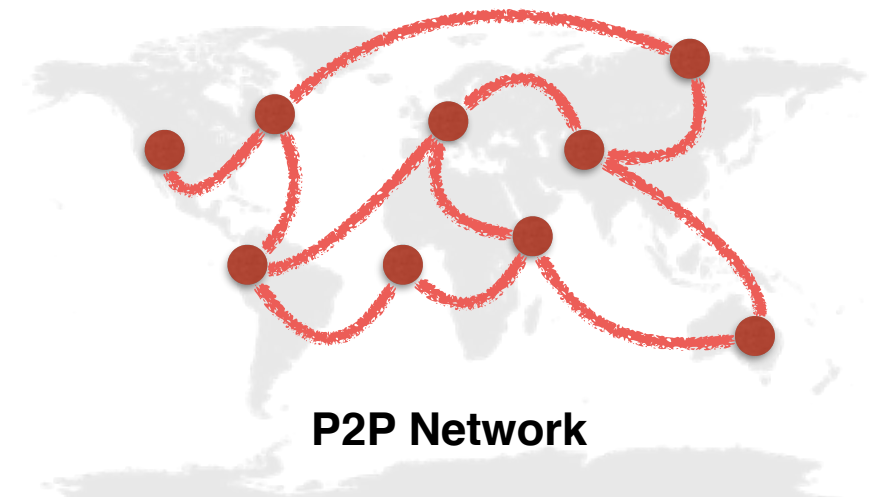
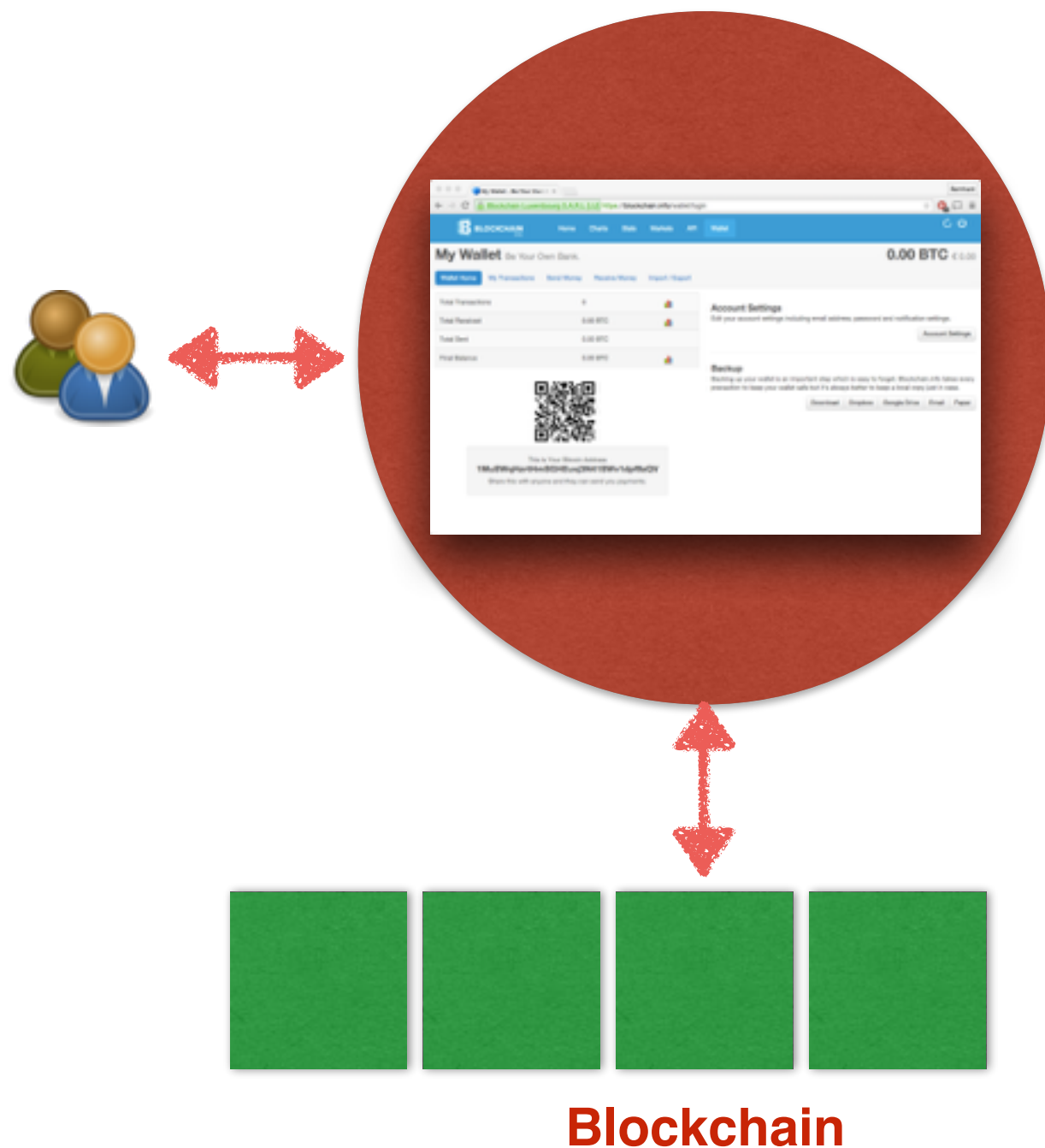
Transaction Processing



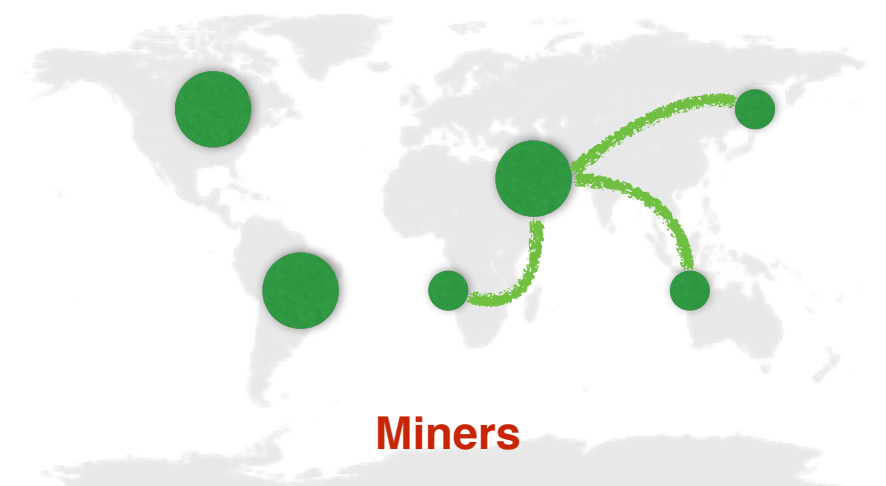
Transaction Processing



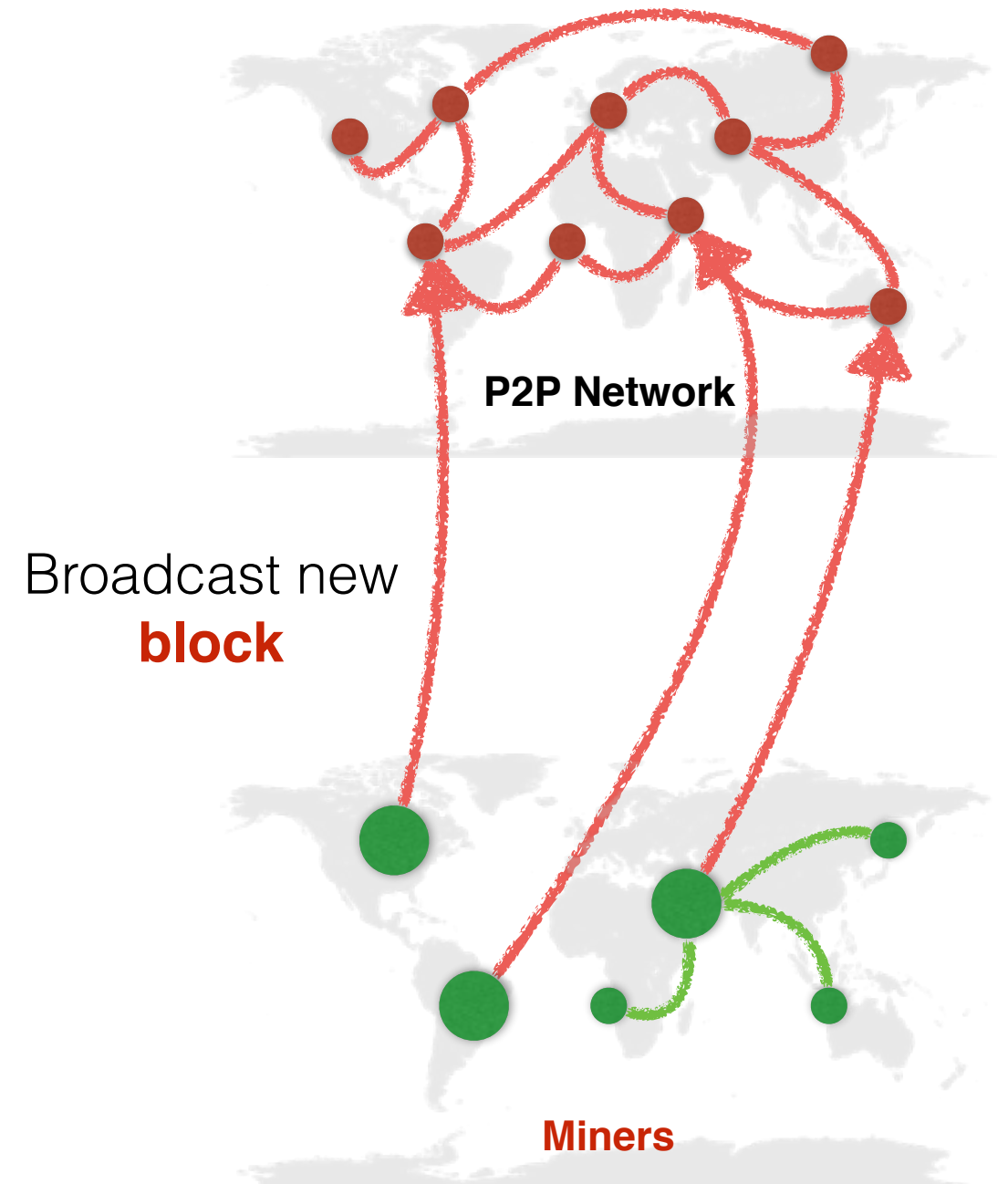
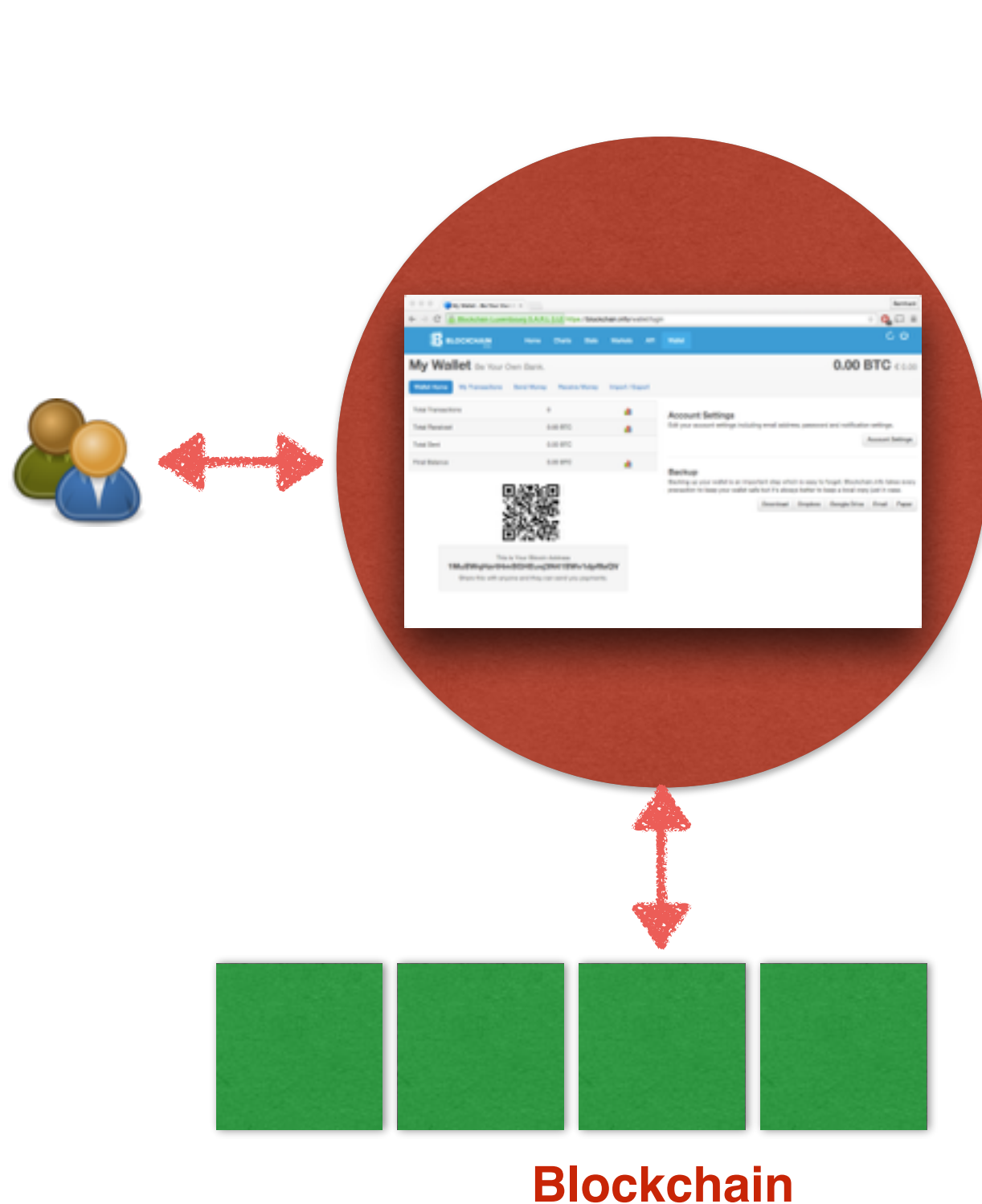
Transaction Processing



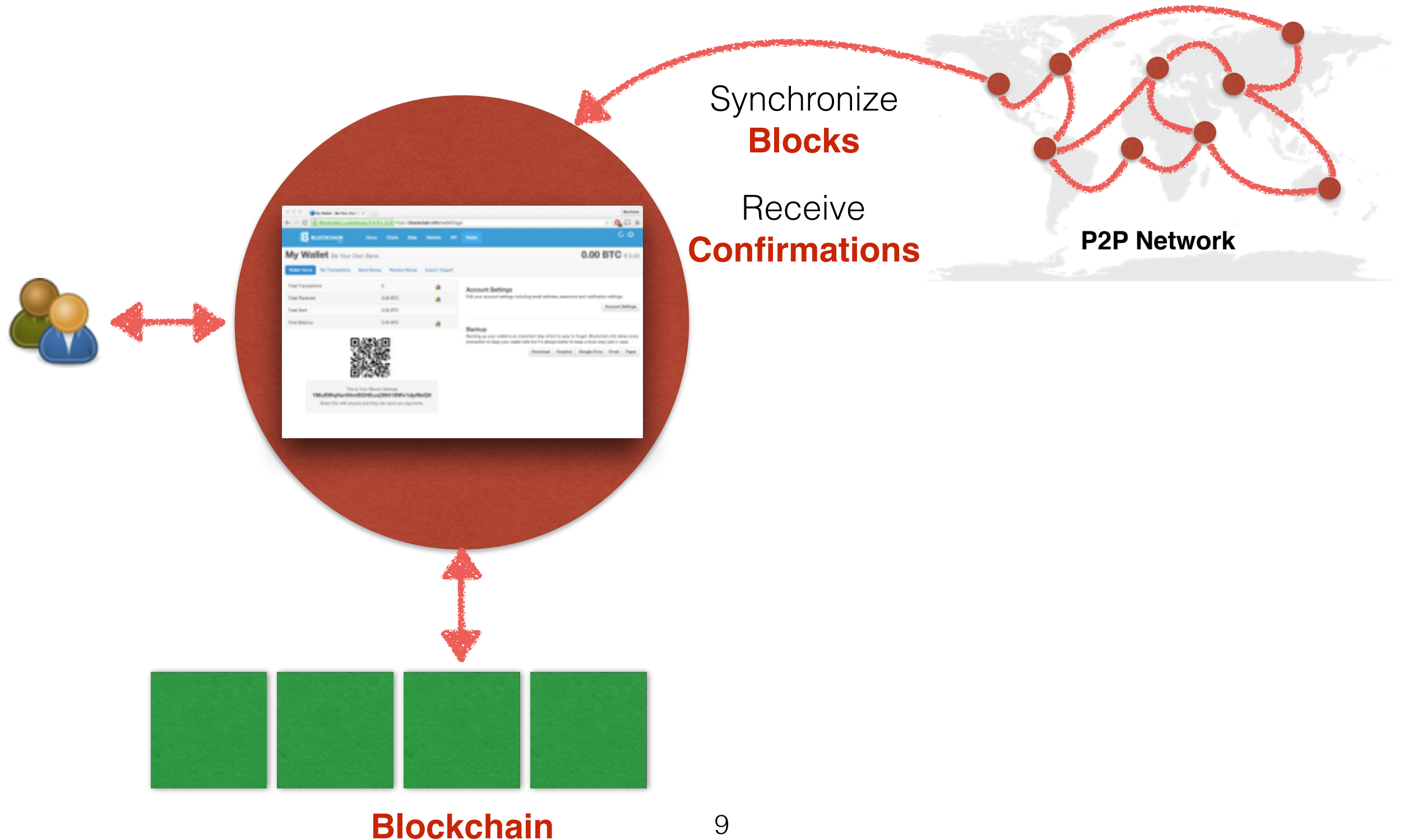
Find a **block**



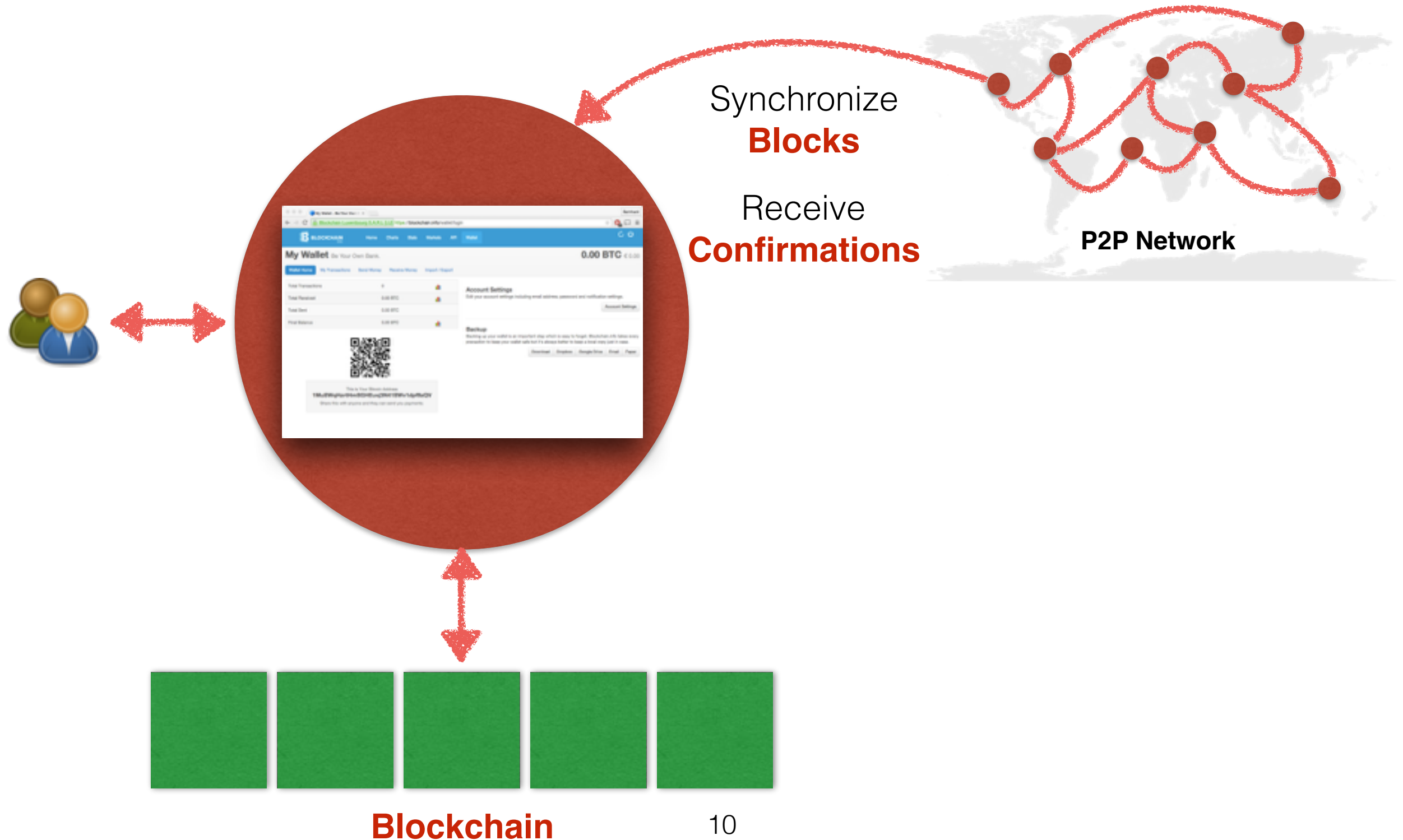
Transaction Processing

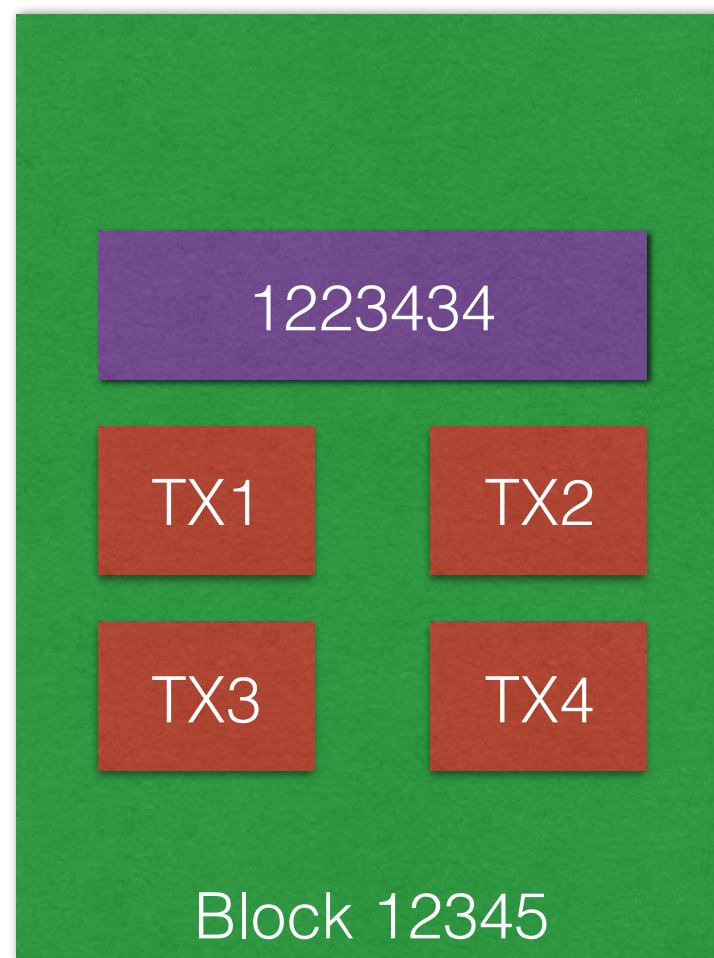


Transaction Processing



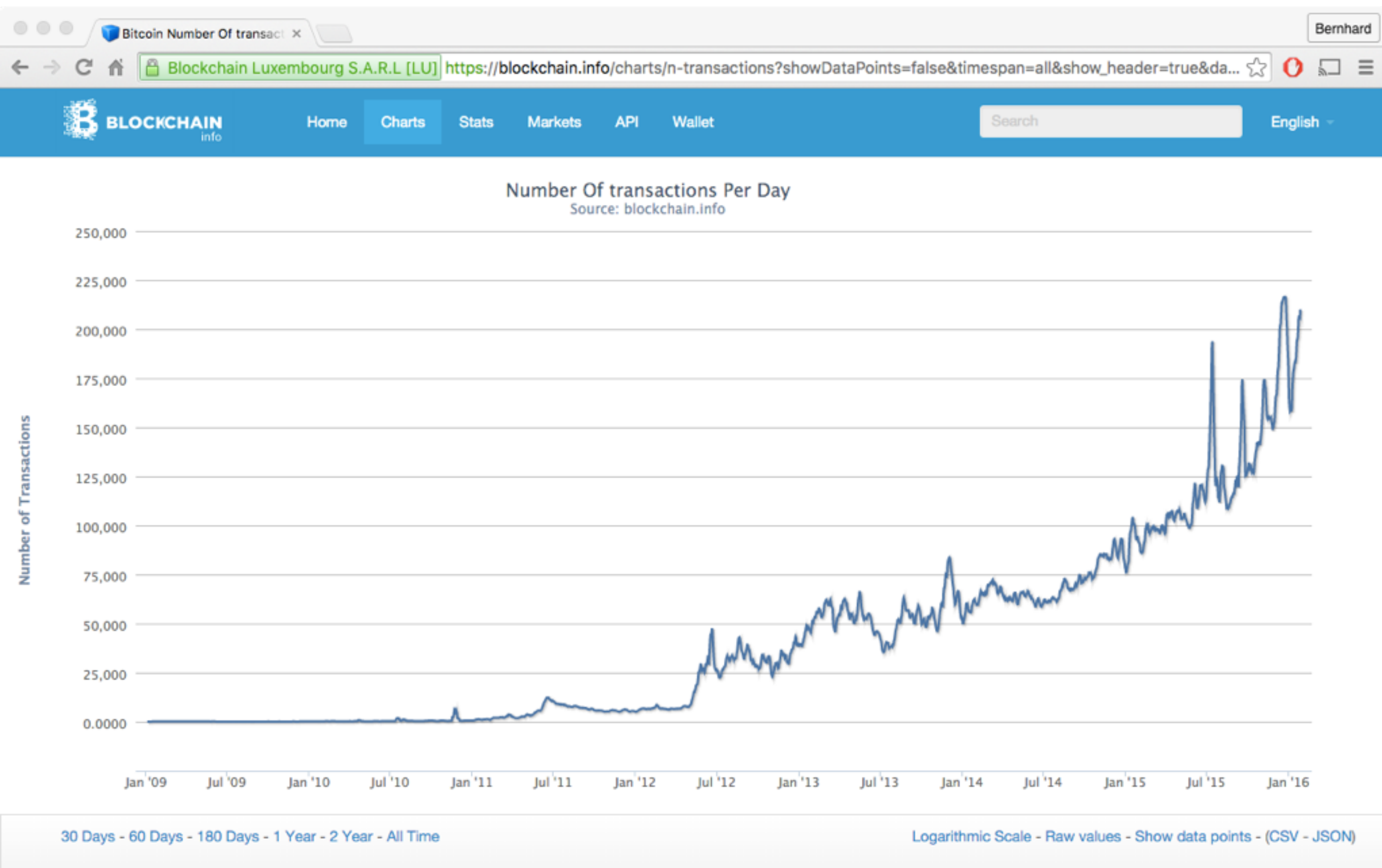
Transaction Processing





Spark R Use Case

Goal: plot such a figure...



Datasets

small blockchain dataset: **large** blockchain dataset:

180 K blocks
3.14 M transactions

3 CSVs (~ 630 MB)

380 K blocks
95 M transactions

3 CSVs (~ 17 GB)

Data Structure

blocks.csv

- block_hash (String)
- height (Integer)
- timestamp (Integer)

transactions.csv

- tx_hash (String)
- is_coinbase (Boolean)

rel_blocks_tx.csv

- block_hash (String)
- tx_hash (String)

Scenario

- Aggregate number of transactions per day and plot using **standard R**
- Aggregate number of transactions per day and plot using **SparkR locally (single node)**
- Aggregate number of transactions per day and plot using **SparkR on a cluster**