

TD-CHAIN: A System to Enhance Transparency in Data Flows

Emergent Research Forum Paper

Wei Chen

University of Arizona
weichen@email.arizona.edu

Laura Brandimarte

University of Arizona
lbrandimarte@email.arizona.edu

David Sidi

University of Arizona
dsidi@email.arizona.edu

Andrea Padula

University of Arizona
apadula@email.arizona.edu

Abstract

Data flows, or exchanges of information among data holders, collectors and brokers, constitute a fundamental part of several industries' modern business models (Awad & Krishnan 2006). The very basis of the data analytics industry, which has grown exponentially in the last decade, lies in the collecting and processing of vast amounts of data. Data about customers—whether it is static demographic data or, even more, dynamic behavioral data about preferences, attitudes, and habits—have become highly valuable assets for companies.

But customers' awareness of data collecting and processing is unclear. According to a recent report by the Pew Research Center on the state of privacy in the US (Pew Research Center 2016), 91% of surveyed adults believe they have lost control over personal information collection and use by companies, and 61% feel discouraged by the amount of effort required in order to understand how companies use their data. While transparency is at the core of industry self-regulatory efforts, FTC recommendations, and the White House Consumer Privacy Bill of Rights released by the Obama administration (FTC 2012), it seems that data collection and use processes are not transparent to users yet.

On the other hand, it is crucial for data brokers to collect accurate data about customers. Inaccurate information in databases results in less effective recommendations, less relevant targeted advertisements, and eventually in lower revenues for companies (Goh, Hui, & Png 2015).

In this project, we present TD-CHAIN (short for: Transparent Data Chain), a software system benefitting data holders by enhancing data flow transparency, and benefitting data brokers by tracking the integrity of data. At a very high level TD-CHAIN detects form submissions in network traffic (a typical example is the web form a user submits when signing up for a new service) and stores information about this data transfer, but not the data itself, with a transaction on a blockchain. If the data collector—e.g., the provider of the service the user signed up for—wants to resell or transfer such data to third party brokers, it can also record a transfer on the ledger, as can a broker transferring to another broker; in this way, a trace remains to indicate where data is flowing. By design, TD-CHAIN provides no technical means for preventing data brokers from sharing data without recording the transfer. The justification for this design decision is that such technical solutions are often circumvented, or result in reduced adoption rates for the systems that employ them. Instead, TD-CHAIN incentivizes transfer recording by (1) improving the reputations of brokers who record transactions over those who do not, and (2) allowing data buyers to verify the integrity of purchased data with no need for a trusted third party. More generally, by increasing transparency of data flows, TD-CHAIN helps data brokers comply with EU data protection regulation, and thus facilitates business relationships across countries. Facilitating the establishment of an international data market is another significant incentive for data brokers to participate in the system.

Keywords

Data flows, transparency, privacy, blockchain