Blockchain-based Data Management and Analytics for Micro-insurance Applications

Hoang Tam Vo, Lenin Mehedy, Mukesh Mohania, Ermyas Abebe
IBM Research – Australia

ABSTRACT

In this paper, we demonstrate a blockchain-based solution for transparently managing and analyzing data in a pay-as-you-go car insurance application. This application allows drivers who rarely use cars to only pay insurance premium for particular trips they would like to travel. One of the key challenges from database perspective is how to ensure all the data pertaining to the actual trip and premium payment made by the users are transparently recorded so that every party in the insurance contract including the driver, the insurance company, and the financial institution is confident that the data are tamper-proof and traceable.

Another challenge from information retrieval perspective is how to perform entity matching and pattern matching on customer data as well as their trip and claim history recorded on the blockchain for intelligent fraud detection. Last but not least, the drivers’ trip history, once have been collected sufficiently, can be much valuable for the insurance company to do offline analysis and build statistics on past driving behaviour and past vehicle runtime. These statistics enable the insurance company to offer the users with transparent and individualized insurance quotes. Towards this end, we develop a blockchain-based solution for micro-insurance applications that transparently keeps records and executes smart contracts depending on runtime conditions while also connecting with off-chain analytic databases.

1 INTRODUCTION

Motivation. In growth markets like developing countries, up to 35% of 4-wheelers and 70% of 2-wheelers are uninsured [2]. This is because of the hefty premium that one has to pay independent of mileage and driving behavior and pattern, as well as the lack of transparency in the insurance quotes. Additionally, in developed countries, drivers who have an extra car for using during the weekends or road trip holidays may find it costly to pay yearly premium for cars that are rarely in use. In order to mitigate this situation and reach this broad and large user base, several usage-based insurance or micro-insurance models have found their way to the market, e.g., pay as you drive, pay how you drive and mile-based auto insurance. Nevertheless, there are still rooms for improvement and novel solutions in this space.

Firstly, there is a lack of transparency in the insurance quotes as well as claim process with existing solutions. We propose to use blockchain technology [7, 8] in order to transparently persist all the information pertaining to the actual trip and premium payment made by the users. Blockchain has its roots in the technology underpinning the digital currency known as Bitcoin [1] where transactions are stored and transferred using a distributed ledger on a peer-to-peer network that is open, public and anonymous. Nevertheless, the technology also has the potential to transform every industry when applied to business network environments.

In our application, by transparent keeping records on blockchain every party in the insurance contract including the driver, the insurance company, and the financial institution is confident that the data are tamper-proof and traceable. This guarantees that any insurance claim request regarding to a trip can be processed quickly and indisputably, hence offering a better customer experience. In addition, well-designed smart contracts deployed on the blockchain, which work in similar manner to stored procedures in traditional databases, provide mechanisms for the system to intelligently reacting to dynamic variables such as driver’s diverting from the original selected path or exceeding road speed limits.

Secondly, insurance fraud is estimated to account for about 10 per cent of all general insurance claims and costs the industry more than $2 billion each year. In fact, storing claims and customer information on a shared blockchain network has the potential to simplify the underwriting process as well as enable intelligent fraud detection. Particularly, as transactions from seeking quotation to binding a policy contract can be persisted on the blockchain, the immutable life record of that policy, e.g., the trip history travelled by the driver, and the record of the policy holder can be traced for fraud detection purpose.

Data management issues. Several technical decisions need to be made when applying blockchain technology as part of database solution for micro-insurance applications. The first issue is to decide which blockchain platform is best suited for this type of business application. We chose Hyperledger Fabric [6] in our implementation rather than public blockchain networks such as Bitcoin [1] or Ethereum [4] because it is specially designed for permissioned and performant business networks. Further, Bitcoin scripting model for implementing smart-contracts is restrictive and not designed for general application use-cases. The second issue arises when we start to store data on the blockchain and develop smart contracts maninpulating these data. Most blockchain platforms adopt key-value data model instead of classical relational model, e.g., Hyperledger Fabric’s persistent state uses a key-value store interface backed by CouchDB [3]. Hence, it is necessary to design a data schema for efficiently maintaining and accessing all information related to the journey and payment made by the users on the blockchain. The third problem here is that there could be throughput implications

CIKM’17, November 6–10, 2017, Singapore.

ISBN 978-1-4503-4918-5/17/11…$15.00
DOI: https://doi.org/10.1145/3132847.3133172

in sending high volumes of data such as real-time sensor and GPS data from millions of vehicles to the blockchain. This raises open research questions: Could we aggregate and consolidate this data before sending to the blockchain? Who does this and can that party be trusted? What implications would it have on the traceability?

Information retrieval issues. As discussed, information extracted from the data stored on blockchain can be useful in reducing fraud related to the integrity of a policy or claim. For example, historical claims stored on a shared blockchain help reduce fraudulent non-disclosure, i.e., misrepresentation of facts material to the insurance policy such as drivers’ failure to disclose their history of accidents. A driver will find it difficult to hide their bad record with one insurer when moving to another insurer. Even though the fraudulent driver may try to tweak their identity when insuring with different insurers, information retrieval techniques such as entity matching can be used to match the tweaked customer information with the existing identity managed on the blockchain. Similarly, for opportunistic fraud, i.e., the exaggeration of otherwise legitimate claims, and premeditated fraud, i.e., deliberate fabrication of a claim, a technique for detecting these types of fraud is pattern matching on the blockchain data, i.e., the same patterns that are used in multiple claims across insurers by the same people or other people can be a red flag that requires further fraud investigation.

Data analytics issues. In addition, even though the blockchain database is useful for transparent persistence of streaming business data, there is no one-size-fits-all database solution for an application. As the micro-insurance application also requires accessing multiple risk analytic databases such as past driving behaviour statistics and past vehicle runtime statistics for computing premiums, a system architecture that allows for maintaining and analyzing both on-chain and off-chain databases is essential. More importantly, an offline analysis mechanism is also needed to periodically access the data maintaining on the blockchain and build relevant statistics for enhancing the risk analytic databases. These statistics enable the insurance company to offer the users with transparent and individualized insurance quotes. Furthermore, advanced machine learning techniques such as deep learning can also be applied to analyze data persisted on blockchain so that the system can learn sophisticated risk patterns.

2 SYSTEM OVERVIEW

2.1 Overall Architecture

Figure 1 provides an abstraction of the entire architecture of the micro-insurance front-end and back-end systems. The architecture consists of three major components, namely driver mobile app, analytic server, and on-chain and off-chain databases.

2.1.1 Micro-insurance Driver Mobile App. This main front-end component provides the driver with user-friendly interfaces for interacting with the pay-as-you-go insurance service provided by the car insurer. Through this mobile app, the driver can perform essential activities such as getting insurance quotes for their desired trips, receiving notification and updates from the pay-as-you-go insurance service while travelling, and calling for emergency road assistance. During the journey, the mobile app also periodically pushes information of road segments travelled by the insured car, which is digitally signed with the driver’s identity, to the micro-insurance analytic server for transparent persistence on the blockchain via a consensus algorithm with other parties in the network such as other insurance companies and transportation authority. In this way, both the driver and the insurer can be confident that the data are tamper-proof and traceable.

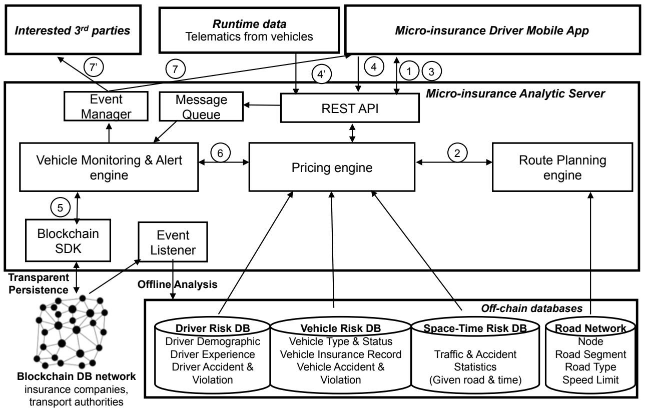


Figure 1: System architecture and basic data flow.

2.1.2 Micro-insurance Analytic Server. The key services provided by the back-end analytic server includes finding alternative routes together with estimated insurance costs, monitoring the insured vehicles and alerting confirmation or changes of insurance cost at runtime. These services are enabled by the following core components.

Route planning engine. This component is responsible to find alternative routes between a source and a destination at a given departure time. This engine is designed to be pluggable. That is, the system can invoke external routing services such as Google Direction Service [5], or alternatively choose to implement its own routing algorithm running on top of a road network in order to generate a set of possible routes based on the driver’s request.

Pricing engine. This component is where the calculation of insurance cost takes place. Given the set of alternative routes generated above, this pricing engine estimates the probability of accidents happening for each route based on the combination of factors such as driver risk, vehicle risk and space-time risk maintained in the underlying historical databases. The likelihood of accidents happening while following a route determines the insurance cost of that route. Note that, the actual insurance cost may be recalculated at runtime depending on real-time situations as will be discussed shortly in Section 2.2. Also, we do not focus on proposing new insurance premium models in this work but rather letting insurance companies use their existing models.

Vehicle monitoring and alert engine. This component receives updates from the front-end such as road segments travelled by the insured vehicle sent by the mobile app or vehicle status and driving behaviours sent by other Internet of Things (IoT) devices equipped with the vehicle. As discussed, these data are then transparently persisted onto the underlying blockchain database. In addition, this engine also sends acknowledgement to the driver for every road segment in the original selected path that they have consistently followed, and more importantly, notify the driver changes in the insurance costs due to dynamic variables such as route diversion, weather events and driver's exceeding speed limit.

2.1.3 On-chain and Off-chain Databases. This back-end layer consists of all databases necessary for the above micro-insurance analytic server. These databases are either stored on blockchain – Hyperledger Fabric [6] in our implementation – for meeting the data transparency requirement between parties, or off the chain – in MySQL database engine – for facilitating risk analytics purposes.

Blockchain database. All data streamed from the vehicle monitoring engine, e.g., the actual trip history and premium payment made by the driver, are transparently persisted onto this blockchain via a consensus algorithm with other parties in the network such as other insurers and transport authorities. This guarantees data protection and traceability of every step in the journey covered by the insurance contract. Insurance companies and transport authorities use this verifiable journey data to process claims related to vehicle damages and human injuries caused by traffic accidents respectively. Financial institutions may also join the blockchain network as they can stamp any payment made by one party for the other parties in the network.

In order to support high volume of transactions, our approach is to run multiple instances of blockchain networks, each of which is responsible for maintaining data related to a subset of the entire user base. More importantly, smart contracts are also implemented and deployed on the blockchain to ensure that the insurance process is handled properly. For example, one smart contract is invoked when the insured vehicle reports its new location in order to persist this new data on blockchain. Another smart contract is triggered by dynamic variables such as weather events, route diversion and driver's speeding. It recalculates the insurance cost and notifies the driver about the changes via the alert engine. A smart contract for handling driver's emergency assistance request is also in place to verify if the driver has consistently followed the agreed path and automatically generate blockchain events that trigger the back-end service to contact emergency assistance.

Risk analytic databases. The calculation of insurance cost is mainly relied on statistics maintaining in these databases including driver risk, vehicle risk and space-time risk. It is important to note that these risk databases are periodically updated via an offline analysis process that accesses the data maintaining on the blockchain and builds the statistics of past driving behaviour, past vehicle runtime and past traffic and incidents.

Road network database. A road network for the geographical area covered by the insurance company is also maintained as several statistics such as past traffic and incidents are mapped directly to each segment in the road network. Furthermore, the route planning engine in the analytic server needs this network data for their custom routing algorithm in order to generate alternative routes based on the user request.

2.2 Basic Data Flow

We now describe the basic data flow in a demonstration scenario in the following, while other sophisticated demonstration scenarios are provided in Section 3. The reference number for each step in the data flow is circled in Figure 1.

Step 1. Via the mobile app, the driver inputs desired trip information including source, destination and departure time, and asks for an insurance quote.

Step 2. Receiving the client request, the pricing engine contacts the route planning to get a set of alternative routes.

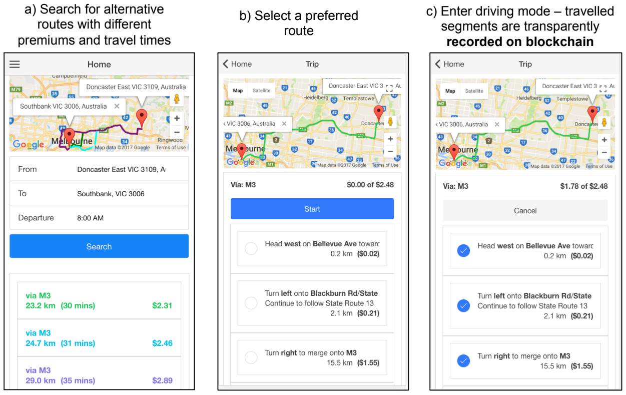
Step 3. Once getting alternative routes, the pricing engine caculates their respective insurance cost based on driver risk, vehicle risk and space-time risk extracted from historical databases. These routes together with their insurance quotes are returned to the mobile app for user selection.

Step 4 and 4'. The driver chooses one route that matches their personal preference the most, e.g., the fastest route or the cheapest route, and instructs the vehicle monitoring engine via the mobile app to start travel mode. During the journey, the mobile app also periodically pushes information of road segments travelled by the insured car, which is digitally signed with the driver's identity, to the vehicle monitoring engine for transparent persistence on the blockchain. In addition, other in-car IoT devices built in the insured vehicle also send information to the vehicle monitoring engine such as vehicle status, driving speed and events.

Step 5. The vehicle monitoring engine invokes the API of the underlying blockchain database in order to transparently persist all the data it has received so that the trip history made by the driver is completely recorded, tamper-proof and traceable.

Step 6. Depending on runtime situations, there could be incidents such as driver's exceeding speed limit of the road or diverting from the original selected route that the vehicle monitoring engine needs to contact the pricing engine for getting updated insurance cost. Note that, changes to insurance cost could be triggered due to external reasons as well, for instances, extreme weather events happening at some point in a long journey.

Step 7 and 7'. The alert engine acknowledges the driver for every road segment that they consistently follow the original selected path or communicates the changes in the insurance cost to the driver due to runtime situations. In addition, the alert engine can also notify other interested third parties such as transport authorities in the event of critical incidents, e.g., traffic accidents.


a) Search for alternative routes with different premiums and travel times

From Doncaster East VIC 3109, A

To Southbank, VIC 3006

Departure 8:00 AM

via M3 23.2 km (30 mins) $2.31

via M3 24.7 km (31 mins) $2.46

via M3 29.0 km (35 mins) $2.89

b) Select a preferred route

Via: M3 $0.00 of $2.48

○ Head **west** on **Bellevue Ave** towarc 0.2 km **($0.02)**

○ Turn **left** onto **Blackburn Rd/State** Continue to follow State Route 13 2.1 km **($0.21)**

○ Turn **right** to merge onto **M3** 15.5 km **($1.55)**

c) Enter driving mode – travelled segments are transparently **recorded on blockchain**

Via: M3 $1.78 of $2.48

☑ Head **west** on **Bellevue Ave** towarc 0.2 km **($0.02)**

☑ Turn **left** onto **Blackburn Rd/State** Continue to follow State Route 13 2.1 km **($0.21)**

☑ Turn **right** to merge onto **M3** 15.5 km **($1.55)**

Figure 2: Persisting data from client app onto blockchain.

3 DEMONSTRATION

The goal of the demonstration is to help the attendees understand the benefits of blockchain data management and analytics in the context of micro-insurance applications. Our demo setup consists of a micro-insurance back-end analytic server built on top of Hyperledger Fabric [6] that provides web services for external applications. To demonstrate real-world scenarios, in this implementation the route planning engine of the analytic server uses Google Direction Service to generate alternative routes and passes this route information into the pricing engine for calculating insurance premiums. In order to effectively demonstrate the system, we developed a front-end mobile application that allows attendees to interact with the pay-as-you-go car insurance service. Part of our demonstration places attendees in the shoes of a person who rarely uses a particular car to travel and prefers to pay car insurance per trip rather than yearly premium. In addition, we also developed an analytic dashboard that pulls and analyzes data from the blockchain for assisting insurance company's staffs to verify claims.

3.1 Finding Optimal Routes with Preferable Insurance Cost

The user wants to quickly find alternative routes from a source to a destination and their corresponding insurance costs at a given time. The demonstration consists of two stages. First, we invite the audience to provide via the mobile app their desired trip information including the start and end locations as well as expected departure time as depicted in Figure 2a. The audience then can fire this request for insurance quote which will be sent to the back-end system.

After receiving the user request, the system finds alternative routes between the source and destination. The risk travelling along these routes is calculated based on route characteristics – extracted from historical data on blockchain – such as traffic statistics, accident statistics, type of roads and speed limitation. This route risk is then combined with other risk factors such as time frame risk, driver risk and vehicle risk in order to estimate the likelihood of accidents and degree of damage, which determine the insurance cost for each route. These alternative routes with their insurance quotes are returned to the mobile app for the user to select.

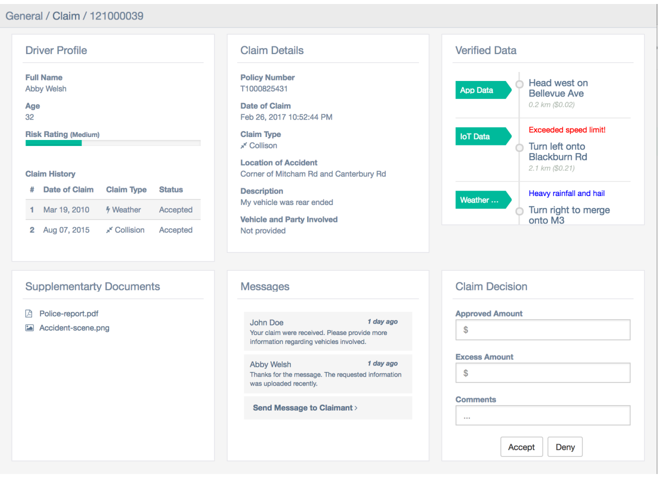
3.2 Transparent Bookkeeping of Trip History on Blockchain

After the above demonstration, the audience has been presented with several possible routes and insurance quotes for their trip request. In this demonstration, they decide which route is best suited for their personal preference based on multiple criteria such as total travel time and insurance cost. Some user may want to take the fastest route which comes with the most expensive insurance cost due to higher risk travelling on the main road. In contrast, another user may prefer the cheapest insurance cost and choose the route that takes longer time but travels on safer roads.

We let the audience select their preferred route on the mobile app (see Figure 2b) and this pay-as-you-go insurance app turns to travel mode, i.e., the road segments travelled by the insured car are periodically recorded and sent to the back-end system for persisting on the blockchain. This data is transparent to both the user and the insurance company, and will be used to verify if the user has consistently travelled along the agreed path in the case there is an insurance claim made. Every time the new location of the insured car has been successfully persisted on the blockchain, the mobile app marks a tick on the road segment that the insured car has just travelled so that the user is confident that the insurance company has acknowledged their actual trip (see Figure 2c).

3.3 Analysis of Blockchain Data for Verifying Insurance Claims

In the above demonstration, the audience has seen that all the major trip segments travelled by the driver have been persisted onto the back-end blockchain. Now, assume that the driver would like to report a car collision during that trip and make an insurance claim. The data related to this particular trip retrieved back from the blockchain (see top right corner of Figure 3) become an undisputable source of evidences to assist the insurer in claim processing.


General / Claim / 121000039

Driver Profile

Full Name
Abby Welsh

Age
32

Risk Rating (Medium)

Claim History

#	Date of Claim	Claim Type	Status
1	Mar 19, 2010	Weather	Accepted
2	Aug 07, 2015	Collision	Accepted

Claim Details

Policy Number
T1000825431

Date of Claim
Feb 26, 2017 10:52:44 PM

Claim Type
Collison

Location of Accident
Corner of Mitcham Rd and Canterbury Rd

Description
My vehicle was rear ended

Vehicle and Party Involved
Not provided

Verified Data

App Data — Head west on Bellevue Ave
0.2 km ($0.02)

IoT Data — Exceeded speed limit!
Turn left onto Blackburn Rd
2.1 km ($0.21)

Weather ... — Heavy rainfall and hail
Turn right to merge onto M3

Supplementarty Documents

Police-report.pdf
Accident-scene.png

Messages

John Doe *1 day ago*
Your claim were received. Please provide more information regarding vehicles involved.

Abby Welsh *1 day ago*
Thanks for the message. The requested information was uploaded recently.

Send Message to Claimant ›

Claim Decision

Approved Amount
$

Excess Amount
$

Comments
...

Figure 3: Analysis of blockchain data for verifying claims.

In particular, if the blockchain data show that the driver consistently followed the original selected route then the insurance company can be confident that this is a genuine claim. On the contrary, if the blockchain records reveal that the driver actually diverted from the original route or sped over the road's speed limit (shown as red font in Figure 3) prior to the collision event, then the insurance company could deny this claim. In addition, data from in-car IoT devices being persisted onto blockchain and, more especially, pattern of claims extracted from historical data on blockchain can also help detect fraudulent claims when there was actually no such collision event as reported by the driver.

REFERENCES

[1] 2008. Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf. (2008).
[2] 2016. Growth Market in South Africa . http://www.timeslive.co.za/thetimes/article1508818.ece. (2016).
[3] 2017. CouchDB. http://couchdb.apache.org. (2017).
[4] 2017. Ethereum. https://www.ethereum.org. (2017).
[5] 2017. Google Direction Service's API. https://developers.google.com/maps/documentation/directions. (2017).
[6] 2017. Hyperledger. https://www.hyperledger.org. (2017).
[7] Joseph Bonneau, Andrew Miller, Jeremy Clark, Arvind Narayanan, Joshua A. Kroll, and Edward W. Felten. 2015. SoK: Research Perspectives and Challenges for Bitcoin and Cryptocurrencies. In *IEEE Symposium on Security and Privacy*. 104–121.
[8] Florian Tschorsch and Bjorn Scheuermann. 2016. Bitcoin and Beyond: A Technical Survey on Decentralized Digital Currencies. *IEEE Communications Surveys and Tutorials* 18, 3 (2016), 2084–2123.