DATA

Decentralized AI-Powered Trust Alliance

Blockchain Data Foundation Ltd.
In Collaboration with Yomob International Co., Ltd.

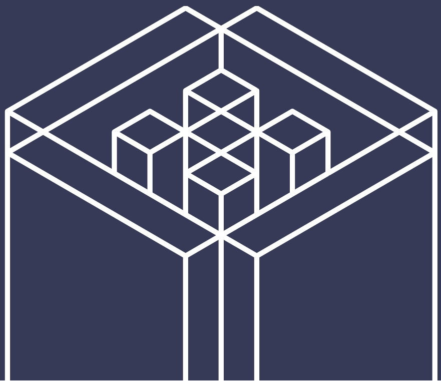
White Paper Draft Version 1.52

This Whitepaper is subject to the disclaimers set out on pages [52] to [60]. Please read them in full and carefully. Contents are subject to change along with project progress.

Table of Contents

Abstract

Forrester Report recently published a study titled "Mobile Fraud: Marketers' Massive Hidden Threat". It is a profound and alarming reflection from the digital marketing and advertising industry to look inside of the beast that has been fed by the industry itself and grown in pace with the latter. This beast is no longer easily containable as the speed of mobile device penetrating in every aspect of our daily life is heightened and our appetite is bigger and hungrier for better quality of content and information of the goods we want to consume.



One of the most overwhelming challenges in the digital advertising industry is the abundance of fraudulent data flowing into the entire network, commonly termed as “Ad fraud”.

Currently, more than 40% of digital ad traffic is generated by automated bots which have defrauded advertisers in billions of US dollars annually [29]. Such fraudulent performance marketing deteriorates efficiency of the whole industry, and results in exorbitant costs for advertisers and thereby cuts into the actual revenue for legitimate publishers. Ad fraud is more of a systematic problem. As the digital marketing and advertising ecosystem becomes more and more complex, the middle layers between advertisers and publishers also swell into multiple stacks. Although the pattern of fraud traffic can be detected, gradually more and more by the evolving machine learning technologies, the ever deeper origin or the source is hardly traceable because of the multi-stacked middle layers that typically excel in extremely complex redistribution, routing and redirecting. Such complexity inevitably has itself invigorated a new set of problems into the digital advertising ecosystem, such as “walled garden” and “data isolation”, etc. The entire AdTech industry is constantly facing the prisoner’s dilemma in that while data sharing globally would maximize the overall performance and effect, yet the optimal strategy for individual user is to keep the behavioral data exclusive locally on the device, be it a PC or mobile. The reason for such dilemma is due to lack of strong incentive to encourage data sharing, and inability to deliver actual value of such incentive if and when it's available. Autonomous, fair and transparent governance, while simply self-contradictory in a centralized environment, constitutes main thesis in a decentralized and consensus based infrastructure, especially with the blockchain technology.

> ***"We are entering a new era, where people expect the value of the services and communities they create is shared with them......Blockchain tokens offer a new way of sharing that perceived value of the community to its members."***
>
> **– Mikko Alasaarela, CEO/founder, inbot.io**

Since monetization relying on advertising has contributed to large share of revenue stream for most internet companies, the current model developed in the past decades has been built on a centralized economy by gauging end user's attention to the ads for better targeting. The demand for user attention has been rising as the competition gets tougher and push for volume number of users becomes a tug of war for the gain and for the survival of Internet companies. The stake is so high and the gain is lucrative, thus created loopholes for the birth of fraud data with fake users, installs and fake activation, etc. that continues to flood the entire communication network. Many advanced fraud detection technologies have been developed to tackle the problems, but most of them focused on the after-fact detection and then to analyze patterns of the data to determine the validity of end user if it even exists.

The technology needs to be able to build into preventing the fraud being propagated into the network traffic flow. The data of the user's attention patterns need be logged and encrypted to be distributed over a decentralized P2P protocol in order to deliver the heightened security with consensus from the community auto-programmatically. Thus it can be validated any time, while stored across the network and shared in real-time.

Applying the blockchain technology and the Artificial Intelligence plowing on deep learning is the foundation technology that can set the standard to ensure the decentralization of the least risky information data flow and storage requirement, where every player in the community or the ecosystem becomes the ultimate winner.

Introduction

DATA is a blockchain project initiated and governed by Blockchain Data Foundation Limited (“Blockchain Data Foundation”, or BDF from hereafter), a company limited by guarantee to be incorporated in Singapore and governed in a not-for-profit manner. This project is being developed in collaboration with Yomob International Co., Ltd. (“Yomob”), a mobile monetization-as-a-service company.

The DATA project is devised to address the root of the fraud at its core. To ensure that every participant in the ecosystem is motivated to "do good", the DATA platform applies a reward system to incentivize end users with their attention contribution and to publishers with pruning their sell-side inventory. More granular information is collected and privately encrypted into a built-in peer-to-peer data storage for sharing across the network. Because every user is made aware of varying rewards only appropriate for his/her attention consumed on ads served in games or apps, it is highly catered and more personal. Therefore a user is less inclined to reject, ignore or block advertising. The experience institutes a more intimate relation between the advertisement and its consumer.

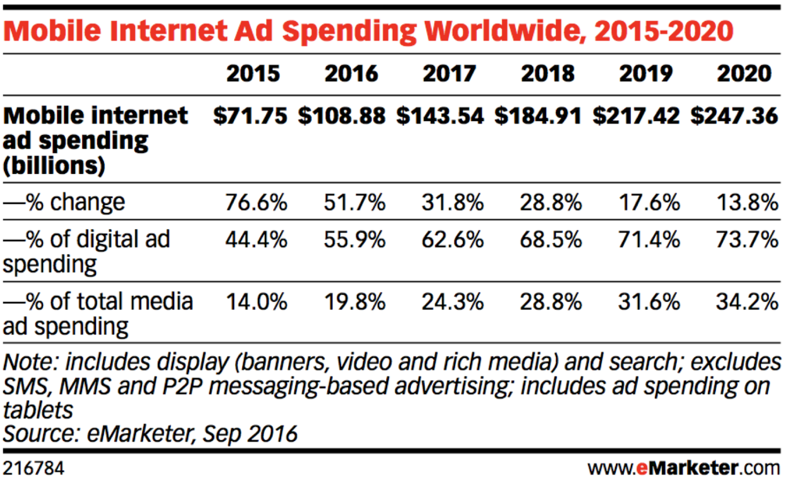
On top of that, advertisement consumption, no matter what format, be it banner or interstitial or a video or a playable advertisement, is tracked at the device level. Besides being the base for the hugely beneficial rewarding for the true user, the analysis of the behavioral data allows the subsequent conclusion of whether the user is a real person or a fake bot.

The DATA project's most innovative edge is to proactively identify the fraudulence based on the pattern or behavior analysis via our proprietary AI algorithm, prior to the fraudulent user data being injected into the entire network. Such preventive measurement is the other side of the coin where reward is granted to a true user for his/her attention consuming the advertisements and to the associated device contributing to in the P2P storage sharing.

The DATA platform and its token, DATA Token (DTA) will serve as the backbone of its blockchain based advertising infrastructure. The system supports micropayment where tokens can be used across applications, such as purchasing virtual items and premium services, or for publishers- redeemed as currency in ad networks, DSPs (Demand Side Platforms), ad exchanges, third-party SaaS (Software as a Service) service and etc.

1. Limitations of Centralized Internet Advertising Bureau

As Internet services derive their revenue from paid subscriptions or advertisements, the latter has been proven to be taking a much larger share behind the duopoly, namely Google and Facebook. Digital advertising is fundamentally a targeted delivery of marketing message, and measurement of whether and how such messages are consumed. When the end users spend their attention to consume services of their choice, the publishers/service providers/media/developers manage to gather user attention data for analysis and reporting, in order to index any errors and to refine future precision in targeting.



Mobile Internet Ad Spending Worldwide, 2015-2020

	2015	2016	2017	2018	2019	2020
Mobile internet ad spending (billions)	**$71.75**	**$108.88**	**$143.54**	**$184.91**	**$217.42**	**$247.36**
—% change	76.6%	51.7%	31.8%	28.8%	17.6%	13.8%
—% of digital ad spending	44.4%	55.9%	62.6%	68.5%	71.4%	73.7%
—% of total media ad spending	14.0%	19.8%	24.3%	28.8%	31.6%	34.2%

Note: includes display (banners, video and rich media) and search; excludes SMS, MMS and P2P messaging-based advertising; includes ad spending on tablets
Source: eMarketer, Sep 2016

216784 www.eMarketer.com

Figure 1. Mobile Internet Ad Spending Worldwide, 2015-2020 [1]

With the rising of mobile era, digital advertising has been drastically shifted to formulate mobile centric strategy. According to [1], mobile advertising now accounts for 68.5% of total global digital advertising expenditure and the trend is only going upward.

Bad actors in the ecosystem have deliberately created alarming amounts of malvertisement that have been extremely difficult to track and eliminate. Along that line, there is also high risk of user's personal data being exposed to unknown and unaccountable layers of intermediaries.

To police the network, IAB (Interactive Advertising Bureau) was created to govern the industry’s standard fraud reporting by way of blacklist and whitelist of malicious nodes, like Ads.txt[10]. Despite their good intentions and tremendous effort put in to serve the industry as it has been, these centralized reports are facing enormous challenge:

DATA INTEGRITY

These lists are centrally managed and organized. It is reasonable to challenge that the lists may be compromised due to single threaded programmatic mistakes or human errors.

TRANSPARENCY

Almost acted as a block box operation, there has not been much detail shared as to the ranking and measurement process over the data it collected and dissected, or whether the 3rd party authentication is certified. Lack of transparency opens windows for prolific fraud activities.

REAL-TIME PRECISION

With the real time dynamics occurring in every micro second in the ad trading, bidding and delivering, these lists quickly became out of date and therefore lack accuracy in a timely manner to reflect the fast changing of all the actors involved. For example, a benign publisher may introduce bot traffic after it is put into the whitelist, making it unpredictable and almost unpreventable in the real time manner.

MONOPOLY OF DATA

It lacks check and balance. Out-of-date technology is overwhelmed with the ever increasing volume of information it must process and index to provide the guidance through the black- and white-list. The reporting cannot be as fast as the fraud being generated and propagated in the internet speed, as a consequence it perpetuates the Mouse-and-Cat game.

2. Severe Ad Fraud

According to [6], intermediaries take more than half of advertising revenues. For example, if an advertiser's budget is $100, less than $50 goes to the end publisher while intermediaries take more than $50.

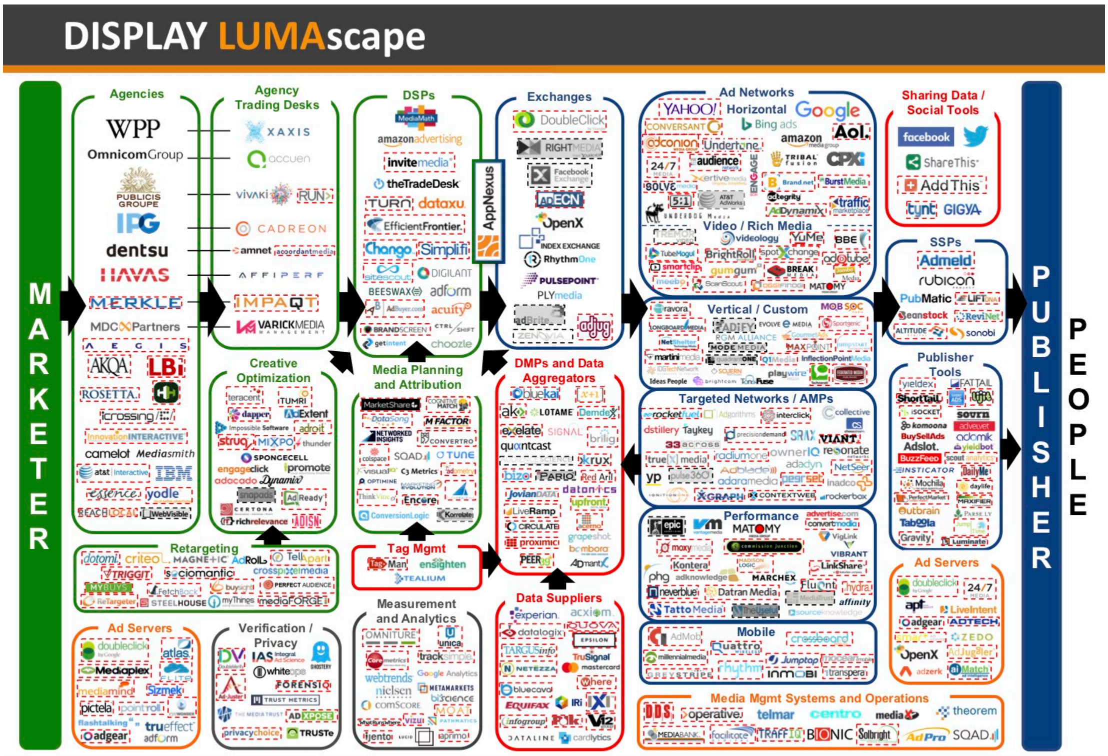


Figure 2. U.S. Digital Ads Landscape by Lumar Partners [6]

Primarily ad ecosystems are primarily composed of advertisers, publishers and users. Advertisers budgeted their marketing expenditure and let known of the target and expected KPI to publishers who own various networks directly or indirectly that can reach to the end users. Creative of an ad campaign would be eventually displayed to the end users in order to gain their attention. To bridge enormous upstream demand of ads serving with downstream supply of user reach, the mid-tier services have evolved to broker the matching, each to differentiate by offering more value-added services. The programmatic I/O has inevitably emerged in all levels to facilitate and fulfill the appetite of the global marketing needs, and driven by speedier and wider spread of the mobile devices. These mid-tier services include demand side platform (DSP), ad exchanges (ADX), ad networks (ADN), supply side platform (SSP), data vendor, ad attribution, analytical and reporting tools, optimization, trading desks, agencies etc.

Along with these services evolved to tackle a number of general or specific issues in the daily transaction of unthinkable volume of ad trading, the scale of complexity and cost also skyrocketed. The increasing complexity with its resulting negative impact on the health of the ecosystem, prompts the industry to search for better solutions. Looking deeper into the formation of the complexity, one could identify several issues, namely the insolation of information and limited sharing of data.

Service providers typically offer programmatic interfaces such as APIs (Application Programming Interface) or SDKs (Software Development Kit) to facilitate functions or features as well as certain data exchange.   Fundamentally every actor desires data, but inherently every one also guards certain data as part of its own purpose of mining the data, thereby between parties, there exists "walled garden". Understandably, owning deeper level and extra amount of user data adds to the competitive edge over the rivalry's, it is at the same time, exactly this mentality that fosters the warm bed of volume amount of fraud by bad actors for lucrative gains. True end user is becoming a pawn in the game without even being aware of it.

Currently some of the ad fraud prevention service providers such as TUNE or Appsflyer would track converted user data from the advertiser's promoting apps as well as data provided by ADNs and DSPs. Such solution is as passive as an after-fact by identifying the fraud actors. So far such solution has not been able to prevent the fraud from being deliberately leaked into the ad marketplace. The critical challenge of the current state of ad fraud prevention methodology is the ability to track down the original source of fraud downstream instead of just detecting the fraud occurrence based on the data collected from upstream.

3. Rewarding User's Attention on Ads Consumption

The ultimate source of all the advertisement income is user attention. A common Internet business model is users get free access to digital services, content and information, like news or media websites and App's, Facebook, Instagram, blogs, etc. While users contribute their attention by viewing the advertisements, the service providers then collect advertisement revenue. This business model had been worked well, but the landscape changed as the mobile Internet develops. First, publishers try to get advertisement revenue more and more aggressively, they exhaustively collect users' personal data, draw the user profiles, try their hardest to insert more ads into the original contents (and in a more ambiguous, hardly distinguishable way), and build extremely sophisticated models just trying to make users click more ads. In the end, user privacy is placed into jeopardy but they got less and less valuable original content. In addition, next generation mobile ads, such as video advisement, interactive advertisement, playable advertisement, Augmented Reality advertisements, are consuming increasing amount of users' attention. More and more users started to employ ads blocking tools and technologies, according to [8], more than 600 million mobile and desktop devices now use ad blocking. They are downloading ad-blocking tools on their mobile devices and this number is growing rapidly. The relationship between publishers and users are no longer win-win, but a combat. The problem to that is users are not well and systematically rewarded but they deserve.

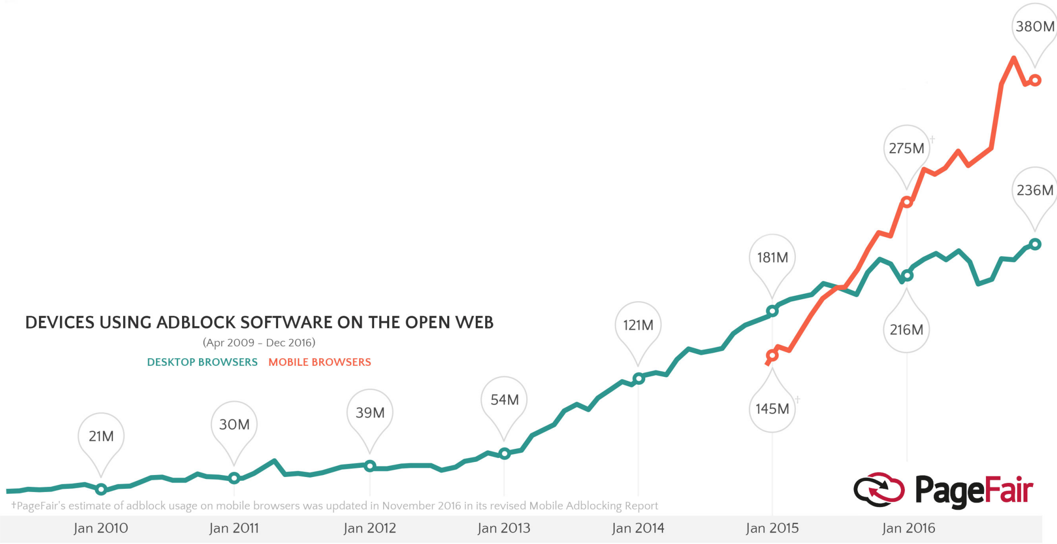


Figure 3. Devices Using Adblock Software on Open Web [8]

With the emergence of mobile ad technology, the value exchange based mobile ads model become very popular among certain mobile ad verticals such as rewarded video ads adopted by lots of mobile games. It is a great balance between user experience and ad monetization. For instance, game player gets some special in-app reward for viewing a video ad, such as Piano Tile 2 from Cheetah Mobile [24]. Also, users can view an ad to get offline rewards such as coupons. These are good initiatives to reward users for their attention contribution. However, these kinds of value exchange based ads have several limitations:

a) They are only feasible for specific type of mobile application publishers;

b) They may deliver benefits that's not needed by the users;

c) They frequently come with more distractive content that bothered users;

d) Users' attention have different opportunity costs given different applications and time they spend viewing the advertisements. However, rewards nowadays are not tailored to the value of the individual user's attention.

To summarize, the time has come for users to be well and fairly rewarded for the attention they contribute, but in a systematic, user-friendly, and widely adopted way.

4. The Solution: Blockchain Based Attention Reward to Prevent Ad Fraud

Technically, fraud traffic is not hard to detect as long as advertisement tracking data is authentic and available. But the previously mentioned isolation and complexity in the middle layers make it extremely hard to gather all available tracking data and trace back fraud traffic to its original sources. The "prisoners' dilemma" as mentioned calls for a out-of-the-box solution, some practical and effective incentives to make ads tracking data publicly shared and available. The incentive needs to be conducted in a decentralized manner as no centralized entity in the online ads industry is capable of handling this, or can be trusted by everyone else. Decentralized governance and incentive distribution are just what blockchain is best at, therefore we have come up with DATA, the blockchain based cyber advertising protocol for ad fraud prevention.

DATA is a blockchain based advertising protocol and technology stack. It consists of 4 layers: P2P Mobile Storage Layer, Blockchain Layer, SDK Management Layer and Application Process Layer.

P2P MOBILE STORAGE LAYER

is a P2P based decentralized mobile storage and distributed hash table (DHT). User activity logs are encoded using erasure coding and distributed in the DHT.

BLOCKCHAIN LAYER

is the DATA chain, which is a standalone public chain, built from a fork of Ethereum on Tendermint, the Ethermint project. The DATA chain has extra logic of working with the rest of the system: storage layer, SDK layer, and fulfills the objective of attention rewarding and ad fraud detection. In the meantime, DATA chain will incorporate state channel technologies to support micropayment.

SDK MANAGEMENT LAYER

contains SDK and protocol for collecting data and perform initial data processing. The open-source SDK transforms user's mobile device into a node in a P2P data storage. It serves as "oracle" to collect data and store data into the storage layer. In the meantime, it contains reputation modeling and management algorithm in calculating the proper amount of attention reward.

APPLICATION PROCESS LAYER

consists of protocol and smart contracts to perform applications of the system, the most important of which is fraud detection. The application layer supports other utilities like micropayment based user application and decentralized ad exchanges etc.

Websites and app developers can integrate DATA SDK to their stacks and process the following functions:

a) Transforms the computer or device into a node of the mobile P2P storage and DHT.

b) Collects users' activity logs.

c) Reputation modeling calculates the reputation score based on the activity data.

d) Activity logs and reputation data are stored into the DHT.

e) The DATA blockchain reads data from DHT, and periodically determines and distributes the attention token reward and handles fraud traffic detection.

The following chapters entitled "Competitive Landscape" is a rather extensive study on a few of the digital advertising projects based on the state of the art blockchain technology. They are analyzed and compared with DATA. In the chapter entitled "DATA Project Technology Design", the high level design of DATA illustrates its technical core competency, followed by why Yomob is the perfect launch pad for DATA, and this Whitepaper will end with introducing our key partners, team, advisors and the road map.

Competitive Landscape

As the blockchain technology has become more broadly adopted, few of its applications are aiming to resolve the larger issues facing the digital advertising industry, namely the fraud, the numerous intermediaries, and the centralized agency dictating the obsolete rules and the more indifference of end users toward the ad serving.

Some of these blockchain application projects do try to focus on certain specific problems although in a narrower sense. However, the level of feasibility or efficiency of them is the subject of this chapter that we will analyze and make reasonable comparison.

1. Basic Attention Token (BAT) [3]

BAT is developed by the Brave browser team and is the pioneer in blockchain based attention rewarding. Users' are rewarded for their advertisement viewing time within the Brave browser with digital tokens; and these tokens can be used in user applications as well as a currency in the advertiser - publisher - user business flow. BAT is limited by the following factors:

CENTRALIZATION

It's a centrally managed application token issued by the Brave company. The tokens can only be used within Brave ecosystem, and not anywhere else on the Internet. The governance of the Brave ecosystem is under the company's control, not the holders of the tokens and end users.

EFFICIENCY

BAT is an ERC20 token on Ethereum blockchain, therefore the efficiency is subject to the capacity of Ethereum. The transaction latency of Ethereum, which is typically several minutes and the scalability of Ethereum, which is around 10 transactions per second on the whole network, makes BAT tokens not feasible in real-life applications like micropayment or ad bidding.

DEPENDENCY

Dependence on the Ethereum network not only introduces efficiency problems, but also feasibility problems. Transactions involving BAT tokens require burning ETH as gas. For an ordinary end user, who does not possess ETH, or doesn't have knowledge of crypto tokens or the Ethereum blockchain, this presents a practical problem - how can Brave company persuade an end user to purchase ETH from places like exchanges and use it as currency to pay for any desired BAT transactions?

PRACTICABILITY

In BAT's model, the advertisers need to buy tokens from exchanges and pay them to publishers and users. This is not much different from current money flow other than the fact that users receive a portion of tokens immediately. It's not made clear in BAT's white paper whether this token flow runs parallel with current money flow or will substitute current money flow. In either case, it's not practically or easily achievable to coach all advertisers to buy BAT tokens as the currency in the advertisement flow. Currently, BAT lacks the infrastructure to do so: they do not have a digital ad exchange system on blockchain, and the performance of Ethereum prevents them from building an effective system.

AD FRAUD

BAT does not present a solution to ad fraud. To make matters worse, it is vulnerable to fraud traffic. As the token distribution is based on a simple timestamp-based formula, there is incentive for fraudulent end users to create fake traffic to get more tokens, like those performed in clicking farms.

In contrast, DATA works perfectly in these scenarios. It is a set of blockchain based protocol and technology stack. Being a standalone chain means the governance is managed by the community of all players in the ecosystem, and the area of application is not restricted to any specific browser or application, but anyone who adopts the protocol. Our technology stack, the P2P based mobile storage layer, Tendermint based consensus engine, as well as the incorporation of state channels for micropayment makes the DATA blockchain highly performant and scalable. Being a standalone chain means DATA does not have the "burn ETH for BAT transaction" problem.

In DATA, reward tokens are not purchased by the advertisers, but distributed by the system in a mining like process. This enables the token flow to coexist with current money flow with the money flow unchanged. Since advertisers need to do nothing different, the protocol is easy to get accepted and bootstrapped. In DATA, token distribution is not based on a simple time-based model, but a more sophisticated reputation-based model. Having users' activity logs and reputation data available in the P2P mobile storage enables the system to effectively detect fraud while in simultaneously as determining the reward distribution.

	Basic Attention Token	DATA Token
Type of Token	Application Token	Protocol Token
Area of Usage	Within Brave bowser	Entire internet
Blockchain Technology	Ethereum ERC20 Token	A Standalone Blockchain
Where's Users' Reward From	Advertisers buy Token (Impractical)	Distributed by the system in a "Mining" Process
Governance	Centrally Managed by Brave company	Managed by community
Efficiency	Low efficiency due to Ethereum's own limitation	Highly efficient via Tendermint & State Channels
Reward Determination	Simple time-stamped	User Reputation Modeling
Ad Fraud	Detection post-fact and still prone to fraud	Preventive at device level before damage is propagated

Table 1. Comparing DATA Token (DTA) with Basic Attention Token (BAT)

2. AdChain [4]

AdChain uses an ERC20 token, named adToken (ADT), as well as a set of smart contracts on Ethereum network. The idea is to build and maintain a decentralized registry of none-fraud publishers, in other words a shared whitelist, via decentralized voting. The project and its token does not aim to solve other problems in the industry, nor are they capable of doing so. The idea of shared whitelist seems good, but faces many practical challenges.

a) Whitelist is not the holy grail for fraud prevention. First it would not be up-to-date. There are thousands of new sites and applications being created and released every single day, but it takes time for those sites to be part of whitelist. Second it would not be comprehensive, with millions of different sites and applications on Internet, many of them being long-tail sites, and thus it is almost impossible to generate a comprehensive whitelist of legitimate publishers.

b) Benign voters are not familiar with fraudulent sites. As stated in the introduction section, detecting fraud is more of a political problem than a technical one. Without sufficient and clear tracing data, one can hardly distinguish the fraudulent sites from the legitimate ones.

c) The voting mechanism itself is subject to attack. As voting is done by individual accounts and the mapping from real-life entities to Ethereum accounts is one-to-many, malicious voters can easily hack the system by voting from many accounts under their control.

d) The project does not clarify many edge cases to make the system a really functional one in all scenarios. These edge cases include how to handle unpopular sites and sparse voting, how to prevent attacks, what the incentives for token holders are, and how to prevent legitimate publishers from becoming malicious after they are classified into the whitelist.

To summarize, AdChain is a nice initiative to use smart contracts on Ethereum and crowd source of wisdom to build publisher whitelist. However, the whitelist itself is not a holy grail for fraud prevention, as AdChain's voting mechanism is subject to attack, as its approach in relying on voters' wisdom falsely underestimates the difficulty of fraud detection for individual players, especially in absence of traceable data.

3. AdEx [5] and Papyrus [23]

The AdEx project aims at building a decentralize d ad exchange with Ethereum smart contracts. To support that, the AdEx project contains a publisher portal, an advertiser protocol, an AdEx profile, and a publisher side AdEx SDK. The publisher protocol is a client-side DApp used for publisher registration, registering websites/channels and advertising properties, and most importantly for accepting particular bids. Likewise, advertiser protocol is a client-side DApp used for advertiser registration, creating different campaigns and placing bids over advertising space. The AdEx profile is a client-side DApp that allows user to change their preferences regarding advertising. AdEx SDK is a HTML5 based publisher SDK for ads display and reporting to the blockchain.

Papyrus is a similar project to AdEx. It also builds a decentralized ad exchange solution, with different decomposition from AdEx, but mostly the "d" counterparts of traditional digital ads players, like "dSSP", "dDSP", "dDMP", etc. Papyrus is also built on smart contracts but with integration of state channels, Papyrus enables real-time bidding following the "dRTB" protocol they propose.

AdEx and Papyrus present feasible smart contract based decentralized ad exchange solutions. AdEx's focus is direct bids for large conversion goals, instead of real-time bidding what Papyrus handles. However decentralized ad exchange does not solve current pain points in the digital ads industry. It's good to have if a lot of advertisers and publishers already moved to the blockchain space, but it doesn't help solve current problems like fraud traffic, ad blocking, lack of user incentive etc. Overall, we see AdEx and Papyrus as good trials into the blockchain space with DApps, but they do not constitute the infrastructure of such a system. They can be nice complements to DATA ecosystem, and we will look into the possible integration.

DATA Technology Design

DATA technology stack is composed of four layers: P2P Mobile Storage layer, Blockchain layer, SDK Management layer and Application Process layer.

It is our attempt to steer toward a clear path where a set of wholesome solutions can be presented.

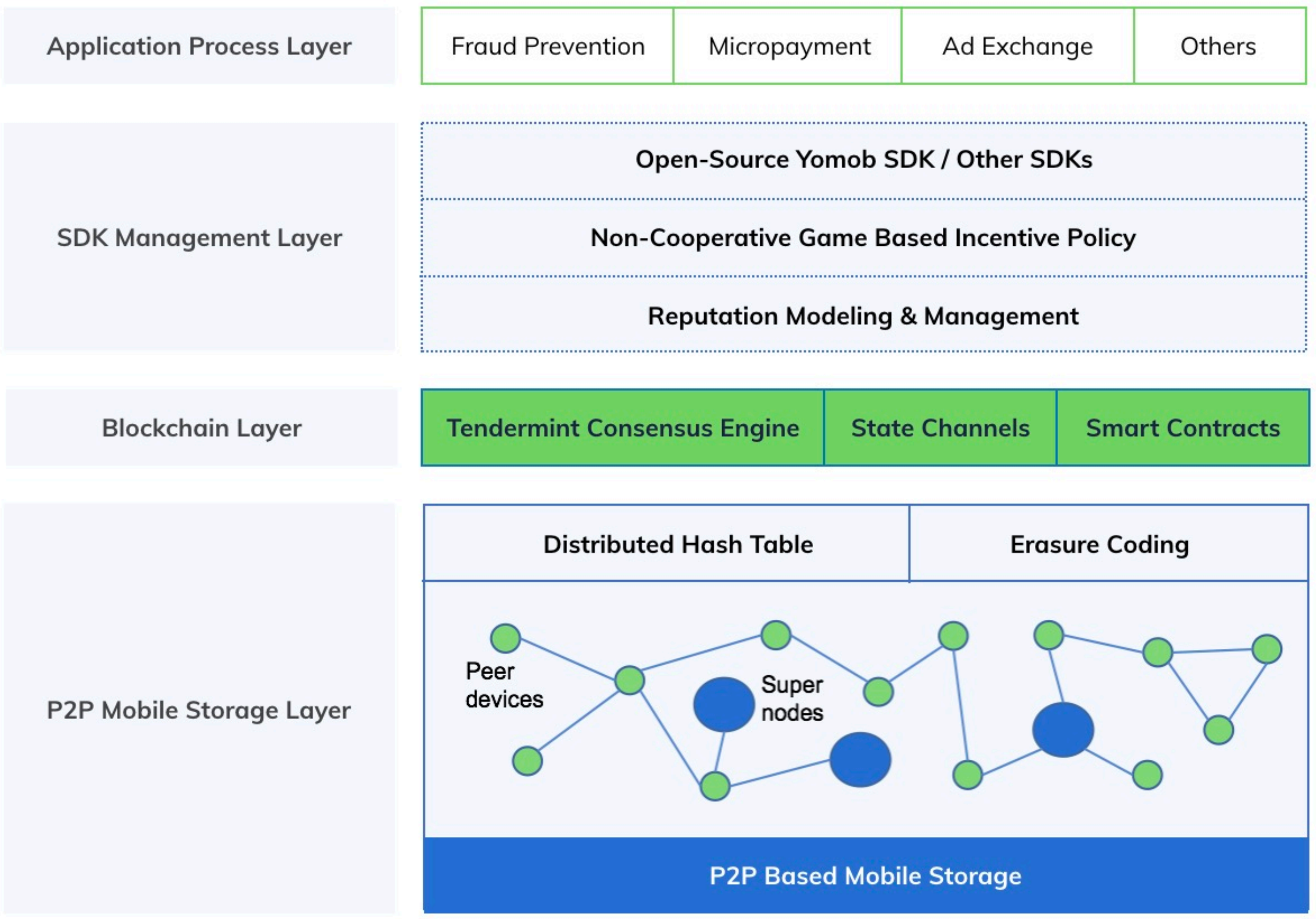


Figure 4. DATA Technology Stack

1. P2P Mobile Storage Layer

In order to prevent the message from broadcasting in the system and ensure the fully decentralization of the system, DATA system leverages on the distributed *Mobile Metadata Management* system (***M³*** or ***M Cube***) to manage the node reputation value and evidences for blockchain. ***M³*** uses distributed hash table (e.g. Chord[17] , Kademlia[18]) protocol to manage nodes in a distributed manager. Inspired by the IPFS system, ***M³*** is targeting on lightweight metadata management in mobile devices.

In DHT, every node only needs to maintain connections with O(log(n)) of nodes in the system. For a node query, instead of broadcasting the query messages to all the nodes, it only needs query O(log(n)) nodes. Figure 4.1 shows that as the number of nodes increase in the system, the number of nodes that need to be contacted is small.[28]

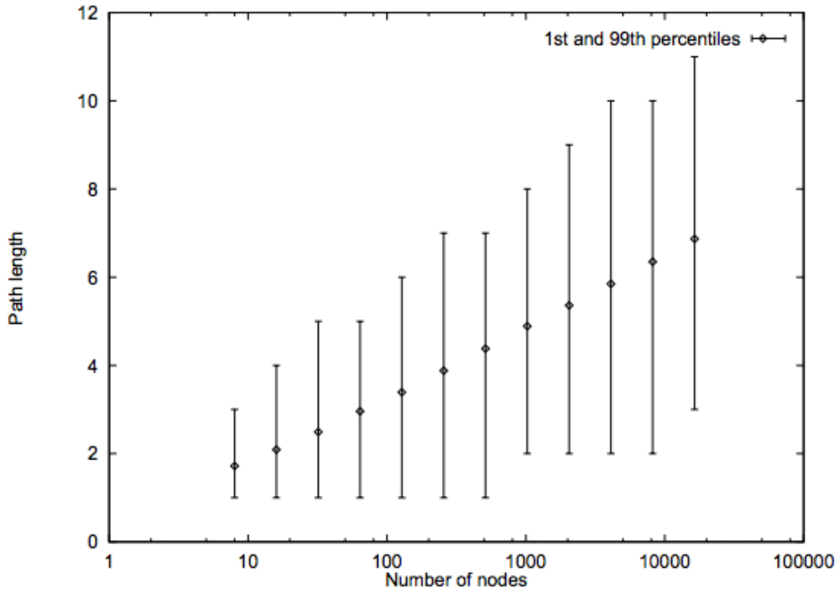


Figure 4.1 Scalability of the nodes

The DHT is widely used in IPFS[16], BitTorrent [19] to increase the scalability of the system. A number of open source implementations of DHT are available on GIthub [26] as well. ***M³*** is a mobile version of the DHT tailored for blockchain metadata management of mobile devices. According to a survey from Pew Research Center [27], cell owners under 50 rarely turn their phones off. Their mobile devices maintain a constant existence online that turns into stable points in the internet to form a DHT as the churn is rather insignificant. To scale up the availability of DHT, ***M³*** also uses super nodes as part of DHT, which are the static VMs in the cloud maintained to provided high availability and reliability. These static VMs are bootstrap nodes that if there is no mobile peers available in the system, the storage layer is still functioning at these nodes. They also serve as NAT servers for mobile nodes for IP address tracking if the cached IP in mobile nodes has expired. The super nodes are contributed by the community in order to ensure the transparency of the system, which is similar to the miner machine.

Figure 4.1 shows an overview of the **M³** overlay. **M³** consists of two types of nodes: Super Node and Mobile Node. A Mobile node does not necessarily have to be a mobile phone- it can be a laptop, desktop or any device that has storage availability accessible via internet. A mobile node acts as a storage node until the device is shut down. Both types of nodes will be awarded with tokens for the duration of time and the amount of storage each contributes to the net.

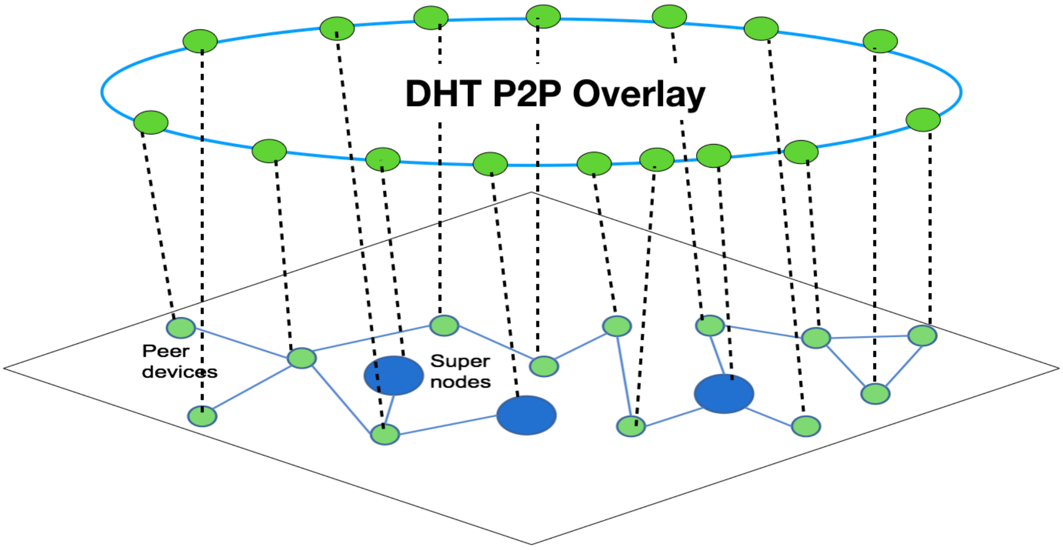


Figure 4.2 DHT P2P Overlay

In the following sections, we will take Chord protocol as an example to elaborate how **M³** leverages DHT for efficient file storage and lookup. Basically, the **M³** is designed to do the following：

a) uploading evidence and reputation values generated and stored in the mobile storage in a distributed manner;

b) managing churns (i.e. length of time when a node joins and leaves the topology)

c) providing APIs for file look up.

In **M³** protocol, each node has two operations for the file management:

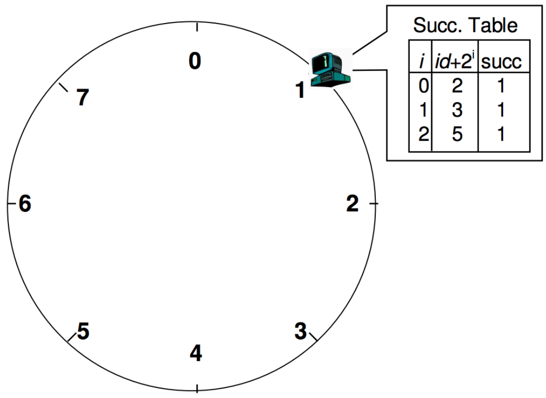
(1) $Store(key, val)$, which puts the file with its hash key into the DHT

(2) $val = Retrieve(key)$, which retrieves the file based on the key

Unlike the broadcasting protocol that broadcasts the query to all the nodes in the topology, the store and retrieve protocols can achieve $O(log(n))$ complexity, where n is the number of nodes in the system. To achieve this, each node needs to manage a finger table. As shown in Figure 5, each node keeps a finger table containing up to *m* entries, where *m* is the number of bits in the hash key. The ith entry of node *n* will contain successor $((n + 2^{i-1}) \mod 2^m)$. The first entry of finger table is actually the node's immediate successor. Every time a node wants to look up a key *k*, it will pass the query to the closest successor or predecessor (depending on the finger table) of *k* in its finger table (the "largest" one on the circle whose ID is smaller than *k*, until a node finds out the key is stored in its immediate successor. DHT is like a distributed skip list. Each time you go *1/2* way towards the destination.

In order to further enhance the file availability, we use erasure code to encode the file chunks and distribute them in the Chord. An erasure code is a forward error correction (FEC) code under the assumption of bit erasures (rather than bit errors), which transforms a message of *k* symbols into a longer message (code word) with *n* symbols such that the original message can be recovered from a subset of the *n* symbols.

We will discuss more details on how **M³** works in the following sections.



i	id+2^i	succ
0	2	1
1	3	1
2	5	1

Figure 5. Example of DHT Finger Table

1.1 Mobile Node Join/Leave

For each node in the system, it must have a node id. We use SHA-1[21], which is a consistent hash function to generate the node id based on the device id. The consistent hash function can ensure the hash collision probability be extremely small. At beginning, all the super nodes will form a initial bootstrap DHT. When a node *i* joins the DATA system for the first time, it generates a node id ID_i using the SHA-1 hash function. Then,

a) It contacts any bootstrap node n' (i.e., super nodes) for its successor by using *find_successor(n')* as shown in Figure 5.1. The super nodes will help the new join node to establish its finger table as shown in Figure 5.2. The finger table also store the IP address associated with the Node IDs. The file querier can use the IP address to directly ask for files

b) Node *i* then Notify the nodes to update their finger tables and predecessors make sure they must point *i*. The nodes are in the ranges of $[i - 2^m, pred(i) - 2^i + 1]$.

c) Then the network need to move responsibility for all the keys to the nodes that is now the successor. It would involve moving the data associated with each key to the new node. Node can become the successor only for keys that were previously the responsibility of the node immediately following. So only needs to contact that one node to transfer responsibility for all relevant keys.

```
// ask n to find the successor of id
    n.find_successor(n'){
      if  id  in (n, successor]
         return successor
      else  n'= closest_ preceding_node (id)
           return n'.find_successor(id)
    }
// search for the highest predecessor of id
      n. closest_preceding_node(id) {
          for i = log N downto 1
              if (finger[i] in (n,id)
                 return finger[i]
          return n
      }
```

Figure 5.1 The pseudocode to find the successor node of an id

```
n.init_finger_table(n') {
     n.successor = n'.find_successor(finger[i].start)
     predecessor = successor.predecessor
     successor.predecessor =n
     for i = to m-1{
          if (finger[i+1].start in [n,finger[i].node]){
              finger[i+1].node = finger[i].node
          } else {
          // ask n' to find other nodes
          finger[i+1].node=n'.find_successor(finger[i+1].start)
          }
     }
}
```

Figure 5.2 The pseudocode to init_finger tables of a new join node

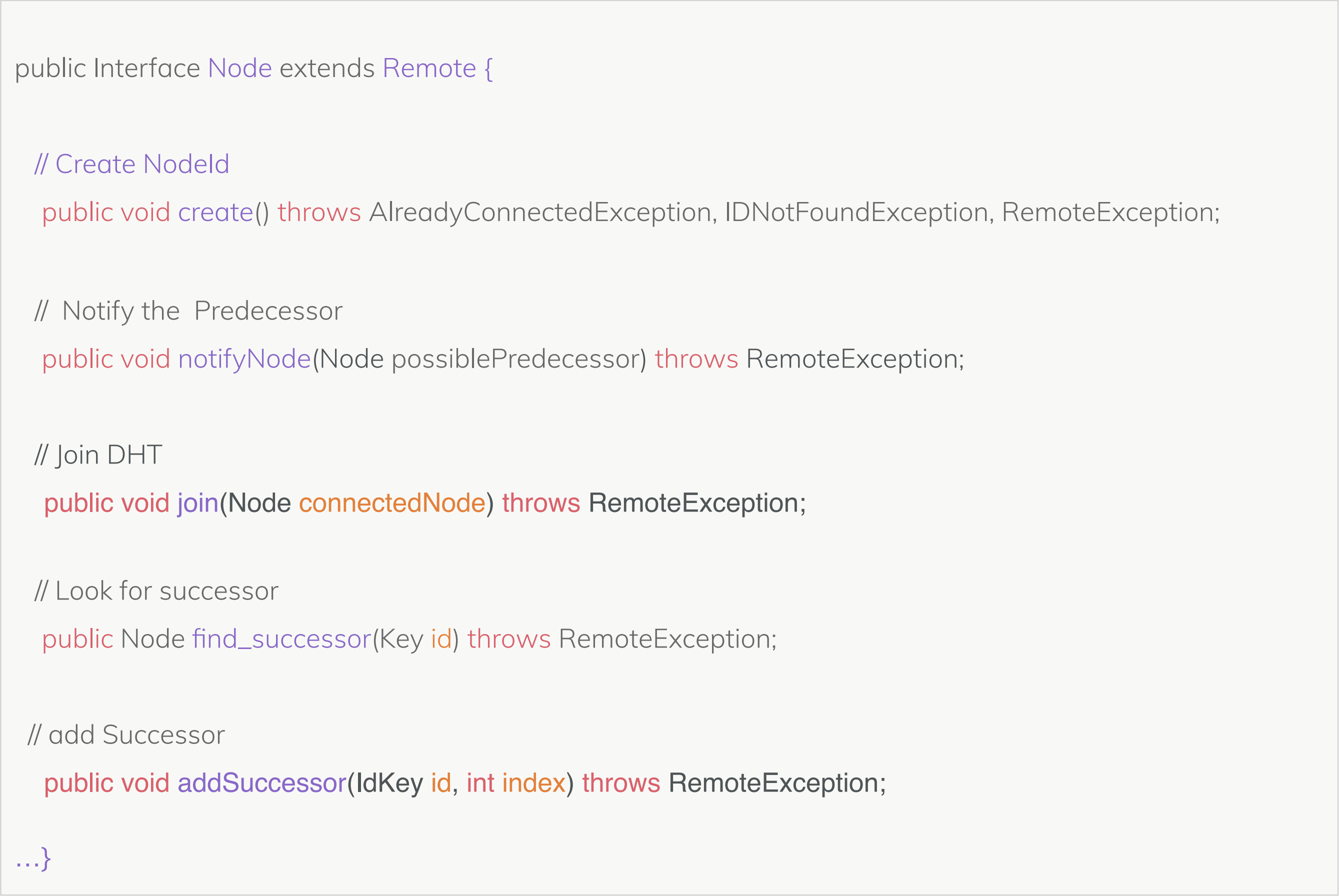
When a node leaves the **M**^{3}, the node sends a leaving message to all the nodes in the finger table and the corresponding nodes will update their finger table. The leaving nodes will also transfer their responsible files to its successor.

To ensure correct lookups, all successor pointers must be up to date. That is, a stabilization protocol as shown in Figure 5.3 should be running periodically in the background. Each node *i* asks its successor for its predecessor. If the predecessor is not itself and let says it is node *p*, it will contact the node *p* and set *p* as the successor. node *p* will also set node *i* as predecessor.

When a node joins or leaves the network, the responsibility of at most O(K/N) keys changes hand (only to or from the node that is joining or leaving. When the N is large, the impact of the key transfer is small.

```
n.stabilize() {
      x = successor.predecessor;
     if (x in (n, successor))
          successor = x;
    successor.notify(n);
}
```

Figure 5.3 The pseudocode to stabilize the DHT



```
public Interface Node extends Remote {

  // Create NodeId
   public void create() throws AlreadyConnectedException, IDNotFoundException, RemoteException;

  //  Notify the  Predecessor
   public void notifyNode(Node possiblePredecessor) throws RemoteException;

  // Join DHT
   public void join(Node connectedNode) throws RemoteException;

  // Look for successor
   public Node find_successor(Key id) throws RemoteException;

 // add Successor
   public void addSuccessor(IdKey id, int index) throws RemoteException;

...}
```

Figure 5.4 Sample Interface of the DHT Nodes

With high probability, any node joining or leaving, an N node Chord network will use $O(\log^{N})$ messages to re-establish the Chord routing invariants and finger tables.

1.2 File Erasure Encoding

In order to increase the availability of the files in the DHT, we use erasure code to store redundant files in the DHT.

For a reputation evidence of size M, we split the log into *k* chunks, each of the same size M/k. Then we apply the (n, k) code on these k chunks to get *n* chunks, each of the same size M/k. Now the effective size is nM/k. Thus the file is expanded n/k times. We need to set *n* to be greater than or equal to *k*, so that n/k is at least 1. If *n* equals *k*, you will have just split the file and there is no coding performed. Any *k* chunks out of the *n* chunks can be used to retrieve the specific file. So this also means that the code can tolerate up to $(n-k)$ erasures. Figure 6 shows an example of a $(4,\ 2)$ code.

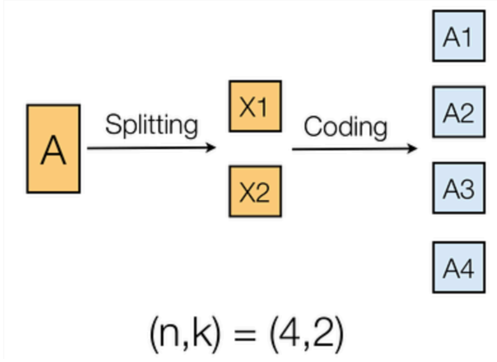


Figure 6. Example of Erasure Coding

Those erasure encoded chunks are distributed into different nodes in the DHT. For example, if the file is $(4,\ 2)$ encoded, we will distribute the 4 chunks to 4 different nodes in the DHT. If any two of the nodes are available in DHT, the original file can be recovered.

1.3 File Store/Lookup

In DATA system, each node is preloaded with a set of hash functions with a certain order. The set of hash functions will be used in File store and look up process.

After we generate the file chunks based on erasure coding, we need to publish the file chunks into the DATA storage system. The file storing process consists of the following 4 steps:

a) Generate the file chunks $(n,\ k)$ based on erasure coding as discussed in Section 1.2.

b) Since we need to publish *k* chunks into the DHT system, we use *k* consistent hash functions to generate *k* chunk ids based on the node id as $ChunkID_k = Hash_k(NodeId, Chunk_k)$. The node id refers to the node that originally own the file.

c) Publish *k* chunks into the DHT using Store Function: $Store(ChunkID_k)$. Since $ChunkID$ and $NodeID$ are generated using the same suits of hash functions, the $ChunkID$ and $NodeID$ are in the same storage space.

d) Once node *i* tries to validate the reputation values of node *j*, it uses the same set of hash functions as used by node *j* to generate a set of $ChunkID_k = Hash_k(NodeID)$. Then, the node calls $LookUp(ChunkID_k)$ to retrieval $ChunkID_k$. For each node relaying nodes in the DHT, upon receiving a query for $ChunkID$, a node checks whether it stores the item locally. If not, the node forwards the query to the largest node in its finger table that does not exceed id. Otherwise, it will forward back the Chunk to the verifier.

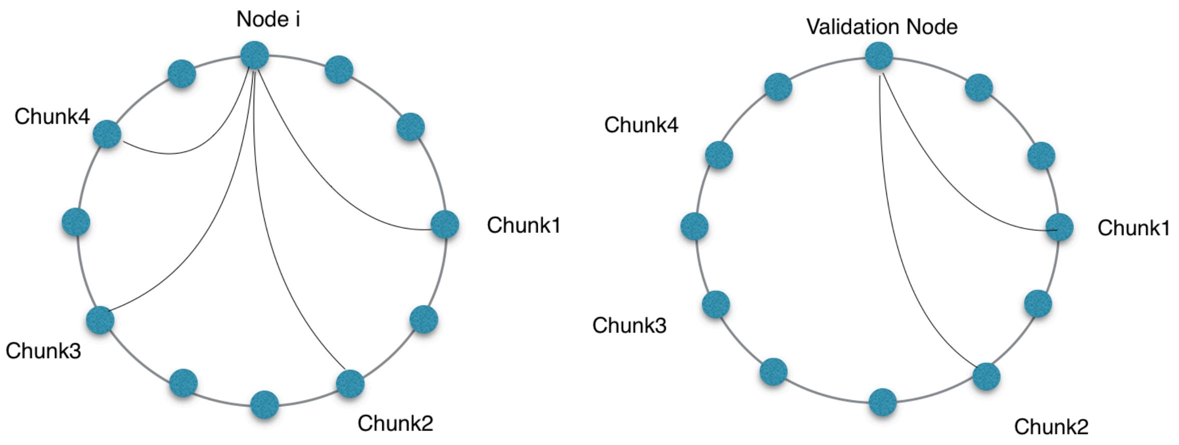


Figure 7. Example of Erasure Coding

The validator will send queries to the nodes in the DHT one by one until it receives k chunks so that it can reduce the whole traffic in the DHT. The verifier recovers the original encrypted reputation logs and use node *i*'s public key to decrypt the logs. Then the verifier needs to run the reputation model based on the behavior log to verify if the reported reputation value is corrected or not.

1.4 File Logging

Since we need to prevent the local user to change the activity logs, those logs are encrypted with owner's private key. Any changes on the file will prevent the file being decrypted. Similar to the Kafka logging system, the file in M^3 is append only. Every file chuck has an expiration date in order to keep the disk usage of the mobile device as small as possible.

The distributed P2P based mobile storage **M³** can be separated out from the DATA system as an independent decentralized file system solution. It can serve as a mobile centric data storage solution to support mobile based blockchain applications for content delivery purpose. Also the mobile devices in the storage system can be mobile phone, mobile PC, tablet and so on.

1.5 Comparison with the IPFS[16].

The InterPlanetary File System (IPFS) is a peer-to-peer distributed file system that seeks to connect all computing devices with the same system of files. IPFS could be seen as a single BitTorrent swarm, exchanging objects within one Git repository consists of a distributed hash table, an incentivized block exchange, and a self-certifying namespace. **M³** shares some similarities with the IPFS that tries to offer a decentralized services based on a distributed hash table. The unique features of DATA system includes:

a) Unlike IPFS that treats all the nodes the same, **M^3** is designed with super nodes and peer nodes to deal with unique challenges in the mobile storage environments.

b) Unlike IPFS that heavily use GIT to manage file versions which is unnecessarily heavy in the mobile environments, **M^3** uses append only log similar to the Apache Kafka[25], which is super light weight.

c) Unlike IPFS uses BitTorrent protocol to coordinate networks of untrusting peers (swarms) to cooperate in distributing pieces of files to each other, which is too heavy to the mobile environment, **M^3** is leveraging awarding tokens and AI based reputation management system to encourage the node cooperation.

d) **M^3** further uses ensure code to better enhance the file distribution efficiency, which is missing in IPFS.

e) In **M^3**, the user identity is presented as a one-way hashed value, which can protect the user privacy. Meanwhile, the file chunks stored in each node are append only and are encrypted with private keys with a salt value. The security of the file can be protected in the mobile storage level as well.

In general, **M^3** is tailored for lightweight metadata storage with high mobility, security, trustworthy, privacy and efficiency. **M^3** not only can support DATA use case, but also can serve as a fundamental layer and service for other suitable scenarios for metadata storage in the whole blockchain community.

2. Blockchain Layer

2.1 Consensus Engine

DATA Chain is built on top of Tendermint [11] Consensus Engine, which is very well suited for scaling delegated proof-of-stake (DPoS) blockchains. More specifically, we're going to build DATA from a forked code base of EtherMint [12], which is a high speed implementation of Ethereum on top of Tendermint. Tendermint is chosen because of its system sophistication and availability:

a) It offers Byzantine Fault Tolerant DPoS consensus. Byzantine fault tolerant means the network is able to tolerate up to 1/3 of nodes being malicious. DPoS is the advancing target of blockchain technology, as it's more energy efficient. Individual validators (like miners in PoW systems) in DPoS systems are strongly incentivized against acting malicious.

b) Tendermint sits at the consensus layer and provides a set of ABCI (Application Blockchain Interface) for developers to build customized blockchain solutions on top of it. This kind of layered architecture is very developer friendly to us.

c) Tendermint is much faster and more scalable than Bitcoin and Ethereum because of its DPoS engine. In a PoW system, users typically wait for a few confirmations (around 6) to conclude that the transaction is “final”. Therefore, it takes about 60 minutes for Bitcoin and 2 minutes for Ethereum to mark the finality of a transaction. Throughput wise, Bitcoin only allows for about 3 transactions per second and Ethereum allows for about 10 transactions, whereas Tendermint reliably supports 10,000 transactions per second at 1 second latency. The numbers make Tendermint a very favorable system choice.

2.2 Proof of Attention (PoA) based Token Reward

In DPoS systems, validators (like miners in PoW systems) need to be rewarded for their contribution to the network. In DATA, users and publishers get most of the reward for their attention and data contribution.

Tokens are to be distributed as following manner:

Initial Launch:

The token supply of DATA Tokens (DTA) at initial launch will not exceed 12 billion tokens, out of which no more than 40% will be sold to Token investors, and the rest belongs to team, advisors, and Blockchain Data Foundation, subject to vesting terms.

PoA Mining-like Process:

The same amount of tokens as the supply at initial launch will be distributed as Attention Reward in the mining-like process with the following allocation:

(1) 10% to validator nodes in the system

(2) 20% to publishers

(3) 70% to end users

Each relevant contributor is properly incentivized according to the following principles:

- Rewards for users and publishers are issued on a daily basis and follow this “Proof of Attention” (PoA) rules.
- Users daily rewards depends on its attention contribution represented by its reputation determined by validators.
- Rewards for validators are issued on each block creation following DPoS rules.

The attention reward distribution amount has a halving schedule similar to that of Bitcoins (e.g. halving about every 4 years). And when the newly minted amount diminishes, we expect DATA and DATA Tokens to become industry protocol and widely accepted. That is the time when advertisers start to purchase DTA as the currency in the ad flow, and the token flow gradually substitutes current money flow in the digital ads ecosystem. The DTA currency represents the real value of the opportunity cost of user attention, which is highly relevant to the user acquisition cost. Given that the aggregate market value of Internet is continuously growing while the total cyber attention of humankind is limited to the entire population on the Internet, theoretically the value of DTA will keep increasing with the time being.

2.3 State Channels

Canonical blockchains are limited to their scalability due to the decentralized nature of transactions: each transaction needs to be processed every single network node. Current blockchain latency and throughput cannot even satisfy the need for currency transactions, not to mention the need for massive digital advertising transaction: clicking and displays, and real-time bidding; they are both orders magnitude more than currency transactions.

To scale blockchains, the blockchain community has studied three directions: larger blocks, sharding and state channels:

LARGER BLOCKS

is a linear scaling solution, and it introduces centralization: only advanced entities with specialized hardware can run blockchain nodes when the block size grows too large. In the meantime, larger blocks improves throughput but not latency; it brings down transaction fees but does not support zero fee micropayment.

SHARDING

technology subdivides the network into subnetworks; transactions between parties within a subnetwork are processed much faster and subnetworks are synchronized with the main network. However sharding is not fully mature yet; there are proposals like Ethereum 2.0 and Zilliqa but they are not in production and not well tested yet.

STATE CHANNELS

in contrast, is a more mature and favorable choice of technology. The idea of state channel is to use off-chain channel between pairs of parties and networks of such paired channels to process fees less peer-to-peer micropayments. Those micropayments do not need confirmation from the whole network, therefore they are virtually instantly fast. There are already projects on state channels for high-speed payments such as the Raiden Network [13] for Ethereum and Lightning Network [14] for Bitcoins and Litecoins. In fact, Lightning network is already in production for Litecoins.

Having studied three directions for scaling, we had decided that state channel is a more favorable choice. Therefore we will implement similar technologies like Raiden Networks in DATA.

2.4 Smart Contracts

Building based on Ethermint means DATA has natural support for Ethereum virtual machines and smart contracts. We will develop smart contract APIs to interact with the P2P mobile storage and the state channels network. That makes smart contracts on DATA more powerful, and makes DATA one of most advanced and full fledged blockchain stacks for various kinds of online advertising applications, or even applications in other areas.

3. SDK Management Layer

3.1 Open Source SDK

Since we treat SDK as an Oracle to manage the P2P storage system and the reputation of a user, we will release the frontend SDK as an open-source project maintained by DATA community so that the public can audit the credibility of the algorithms and the code. The mobile SDK has the following functionalities:

a) Provide application level services such as ad fraud prevention service for ad networks and DSPs and etc. We will discuss details in Section 4;

b) AI based reputation modeling to determine the reputation value of a user based on the activities logs;

c) P2P mobile metadata storage managements

3.2 Game theory based Incentive Modeling

Game theory is a theory of applied mathematics that models and analyzes interactive decision situations. Based on whether the players make binding agreements in the game, game theory models can be classified into cooperative game models and non-cooperative game models. In the former, the players act based on their binding agreements. In the latter, the players are self-enforcing entities (i.e., nodes can change their strategies at any time to maximize their benefits).

We use $\mathcal{N} = \{1,2,...,n\}$ to denote the set of all mobile nodes (i.e., game players) in the system. We use A_i to denote the action set for node i, and $A_i = \{I, C\}$; the C (i.e., cooperative) action means the node will not hack the system, while the I (i.e., incooperative, non-cooperative) action means it hacks the system and be malicious. The action chosen by node *i* is denoted by a_i, and the actions chosen by other nodes are denoted by an action set:

$$\mathbf{a_{-i}} = \{a_1, a_2, a_3, \ldots a_{i-1}, null, a_{i+1}, \ldots, a_n\} .$$

$\mathbf{a} = (a_i, \mathbf{a_{-i}}) = \{a_1, a_2, a_3, \ldots a_{i-1}, a_i, a_{i+1}, \ldots, a_n\}$ denotes the action set of all the nodes in the system when report reputations. We use D to denote the Cartesian product of the action set of a node, use $U_i(a_i, \mathbf{a_{-i}})$ to denote the utility (i.e., payoff, benefit) function of a node *i* given the strategies used by other nodes and use ($U(\mathbf{a})$) to denote the sum of the utilities of all nodes. The game theory model for the system is denoted as: given a normal form of game G, $G = <\mathcal{N}, D, U_i(a_i, \mathbf{a_{-i}})>$. Every rational node in the system intends to choose an action that maximizes its utility for a given action tuple of the other nodes. That is, the best action $a_i^* \in A_i$ is the best response of node i to $\mathbf{a_{-i}}$ iff for all other $a_i \in A_i$, $U_i(a_i^*, \mathbf{a_{-i}}) \geq U_i(a_i, \mathbf{a_{-i}})$.

Definition 1. A Nash Equilibrium (NE) is an action tuple that corresponds to the mutual best response. Formally, the action tuple $\mathbf{a}^* = (a_1^*, a_2^*, a_3^*, \ldots, a_n^*)$ is a NE if $U_i(a_i^*, \mathbf{a}_{-\mathbf{i}}^*) \geq U_i(a_i, \mathbf{a}_{-\mathbf{i}}^*)$ for $\forall a_i \in A_i$ and $\forall i \in \mathcal{N}$, where A_i denotes the action set (cooperative, non-cooperative) for node *i*.

Therefore, a NE is an action set where no individual rational node can benefit from unilateral deviation.

Definition 2. An outcome of a game is non-Pareto-optimal if there is another outcome which would give all players higher payoffs, or would give partial players the same payoff but the other players a higher payoff. An outcome is Pareto-optimal if there are no other such outcomes.

Therefore, the ultimate goal of the non-cooperative model is to make the cooperation behavior as both Nash equilibrium and Pareto-Optimal.

In DATA system, three kinds of nodes will be awarded: DATA nodes, infrastructure contributing nodes and user attention contributing nodes. Those cooperative nodes will be awarded with tokens to increase their incentives. The malicious nodes will be punished by putting them into the blacklist. The awards/punishments that we are designed will guarantee the cooperative behaviors are both Nash Equilibriums and Pareto-Optimal.

3.3 Reputation Management

DATA SDK leverages Artificial Intelligence (AI) to calculate the reputation of the users, which reflects the trustworthiness of the users. DATA also evaluate advertisers' reliability based on the collaborated rating from peers. In order to ensure the peer consensus for the reputation value, the rating and activities are encrypted with private key of owners and put them into blockchains. The AI model is published with the SDK so that the peers can use the model to verify the reputation value based on the activity logs. The reputation models can be easily extended to other scenarios such as online retail, online banking and cyberbully and etc. for reputation management to prevent fraud. The sections below will give the details of the reputation management.

3.3.1 Publisher Reputation Modeling

In order to distinguish the malicious users and cooperative nodes and give a evaluation of the quality of the user, we are using Long Short-Term Memory (LSTM) [20] to learn the historical malicious nodes and cooperative nodes based on the historical logs. The model is trained offline based on TensorFlow[22] and is supplied with the SDK so that the user that uses SDK can use the same model to calculate/verify the reputation values.

An LSTM block is composed of four main components: a cell, an input gate, an output gate and a forget gate. The cell is responsible for "remembering" values over arbitrary time intervals; hence the word "memory" in LSTM. Each of the three gates can be thought of as a "conventional" artificial neuron, as in a multi-layer (or feedforward) neural network: that is, they compute an activation (using an activation function) of a weighted sum. Intuitively, they can be thought of as regulators of the flow of values that goes through the connections of the LSTM; hence the denotation "gate". There are connections between these gates and the cell. Some of the connections are recurrent, some of them are not.

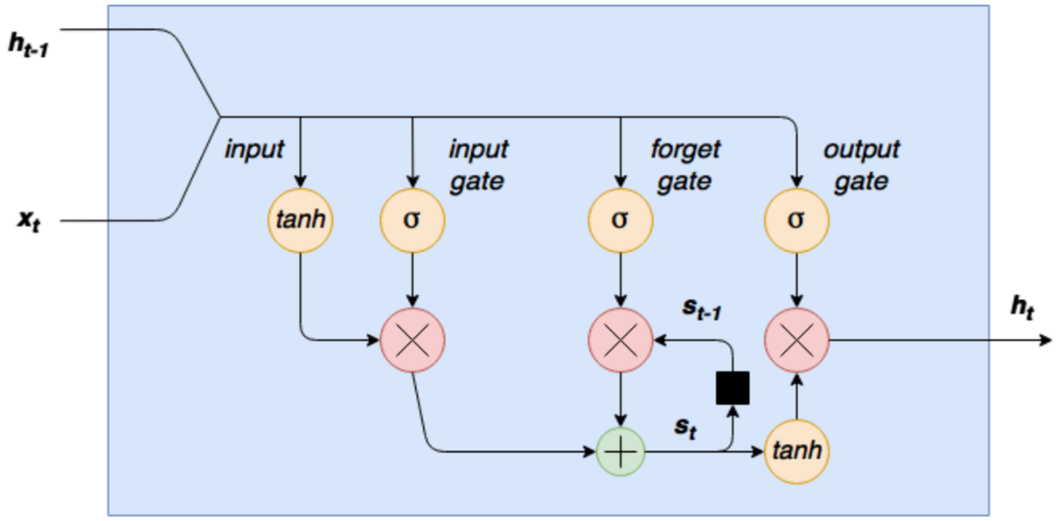


Figure 8. LSTM Cell

The SDK layer will track the click activities of users before/during/after viewing advertisement. Since this activity data is time-series data, we use a sliding window of size N across the data to generate training sets. The basic sliding window size is 1 day. Each sliding window is labelled with a reputation value scale from 0 to 1 with 0.1 for each step. 0 indicates the malicious nodes and $1$indicates the dedicated attention contributor. The value in between indicates the different level of attention each user has put into the model. Those time series activity data will encode the information such as the number of user views a day and the length of time a video advertisement is viewed for. It can also encode whether the user has clicked on the advertisement. Together with these time series activity data, the inputs of the model also include device type (e.g. iOS, Android, AR device), App type (e.g. Game App, Social App and etc.), location (e.g. Beijing, New York) and so on. All these advertisement related features are fused into the model to measure the quality of a user in a multiple-dimensional way.

The LSTM's training model looks like the figure below. The time-series training sets are the normalized users click logs, which are the inputs of the model. The outputs of the model are the reputation values. The reputation values are calculated based on SoftMax functions.

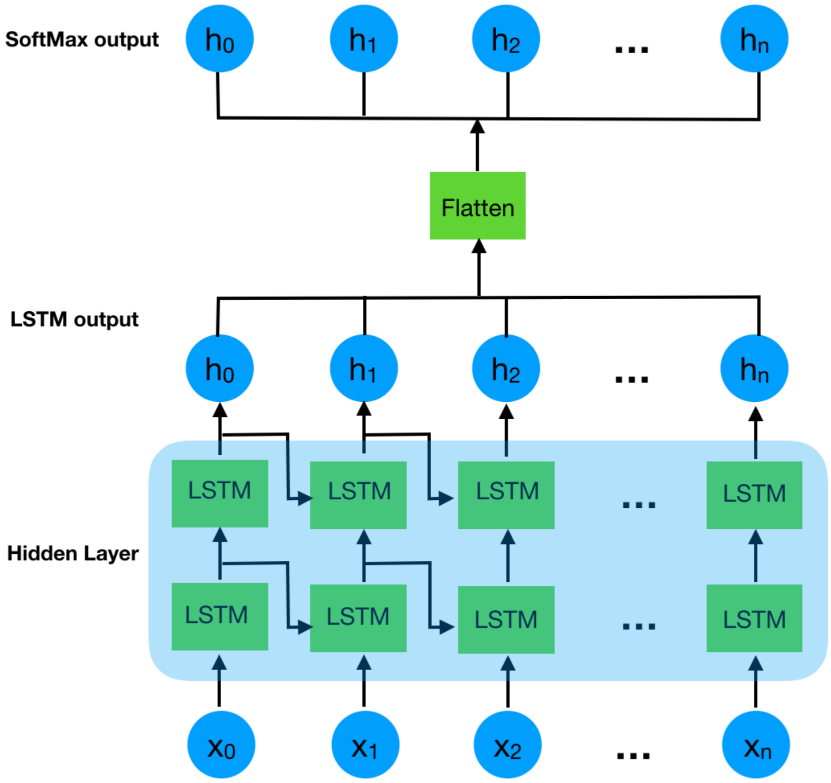


Figure 9. LSTM Architecture

The trained model is loaded into the SDK Management Layer. In the SDK layer, we will calculate the reputation value within a monthly moving window. That is, the user must accumulate 1 month's worth of click logs before get a reputation value based on the LSTM model. Section 3.3.3 will discuss why we need to have a one month's worth activity log. When calculating a reputation, a user creates a feature vector including a week's worth of activity logs, location, device type, app category. The input vector is supplied into the model which then calculates the reputation value of the user based on the inputs.

The click log and the associated reputation value will be transferred into the distributed mobile storage and blockchains. When a user tries to verify the reputation value, it will find the activity log of the user and use the same model to generate the reputation value of a user.

3.3.2 Advertiser Reputation Modeling

In order to prevent advertiser/ad networks from supplying malicious advertisements to the publishers, each viewer in the system can assign a like/unlike rating of the advertisements based on advertisement user experiments. If the advertisement is malicious, the SDK also supplies a API to report the malicious behaviors. All these ratings and malicious reporting will be stored in the distributed mobile storage and reported to the blockchain. The reputation of an advertiser j is a weighted sum of the rating from different users *i* as:

$$R_j = \sum_{1}^{n} R_i \cdot W_i / n,$$

where R_i is the rating of user *i* and W_i is the reputation of user *i*. That is, we will give more rating weights to the nodes with high reputation values.

3.3.3 Attack Analysis on Malicious Nodes

Since DATA system use an oracle to calculate the reputation value of a user based on the activity logs, we need to prevent or reduce the impacts of malicious nodes if they fake their activity logs and transmit them into the blockchains. First, we grant user tokens only when they generate 1 month's worth of valid activity log. The AI model will validate the activity log. This policy can ensure that the malicious nodes are unable to generate a bunch of device ids and changing them repeatedly. Secondly, before granting the tokens to the users, Blockchain Data Foundation can check the activity data with the advertiser's data management platform (DMP) data to ensure these activities are actually valid. We will also encourage advertisers to put their observed user click data into the blockchain to automate the verification process.

4. Application Process Layer

The built-in application of DATA is fraud detection. Since it supports EVM and smart contracts, all applications available on Ethereum are also available on DATA. As it supports state channel based micropayments, DATA Tokens can be used as digital currency in scenarios broader than Bitcoins and Ethereum, like premium content purchasing, online gifting, real-time bidding etc. Here we describe a few.

4.1 Fraud Prevention

The Fraud Prevention in DATA System is achieved through 3 ways.

a) Reputation Management. As discussed in Section 3.3, based on the AI models, both advertiser and publisher are rated with a reputation value, which reflects their trustworthiness. The AI models can detect the malicious nodes with suspicious long term activities and punish them. The peers can rate the advertisers to collaboratively identify malicious advertisers and blacklist them.

b) Blockchain. Since the activity logs can be accessed in the blockchain consensus layer, users and advertisers can verify the reputation of each other anytime, everywhere, with high transparency.

c) Reward and Punish Incentives. The DATA Tokens together with the punishments can provide high incentives to users in the system to be cooperative.

4.2 Micropayment for User Application

We will build desktop and mobile wallets for DATA. The wallet software will automatically determine the assigned address based on device ID. We will also release SDK too to integrate the wallet into third party applications. As DATA supports lightening fast micropayment, DTA can be used in all places where fiat currency can be used online, for instance, purchasing virtual goods, premium content, online subscription, online gifting, etc. Figure [11] illustrates mock ups of DATA wallets and example applications.

4.3 Decentralized Ad Exchange

As discussed previously, we'll look into porting of existing smart contracts based decentralized ad exchange solutions, like AdEx [5] and Papyrus [23] into DATA. The DATA infrastructure naturally supports that.

5. DATA System Flows

Figure [10] illustrates the flows in DATA ecosystem. There are 5 categories of players in the ecosystem: advertisers, publishers, users, DATA network and Blockchain Data Foundation, and there are three functionality flows: ad flow, data flow and token flow.

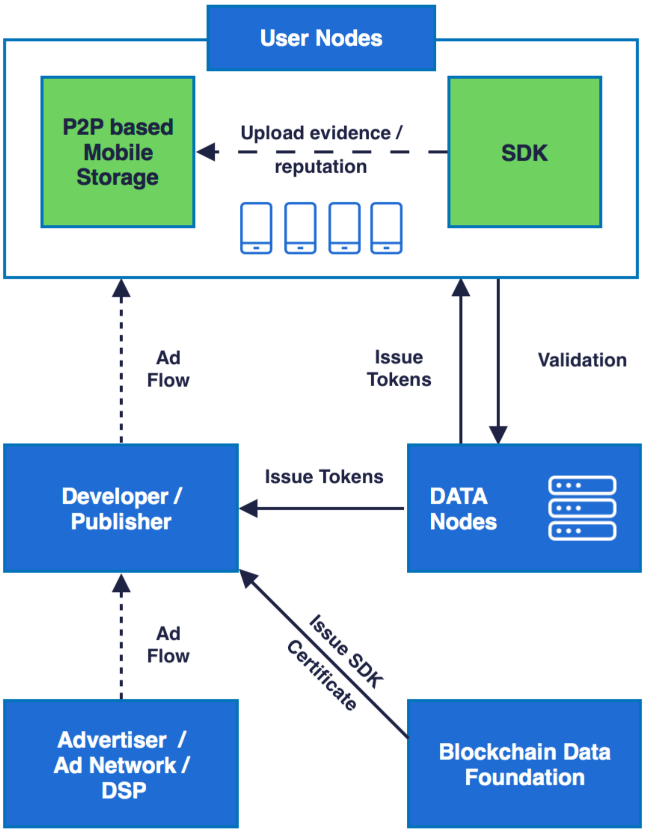


Figure 10. DATA System Flows

AD FLOW

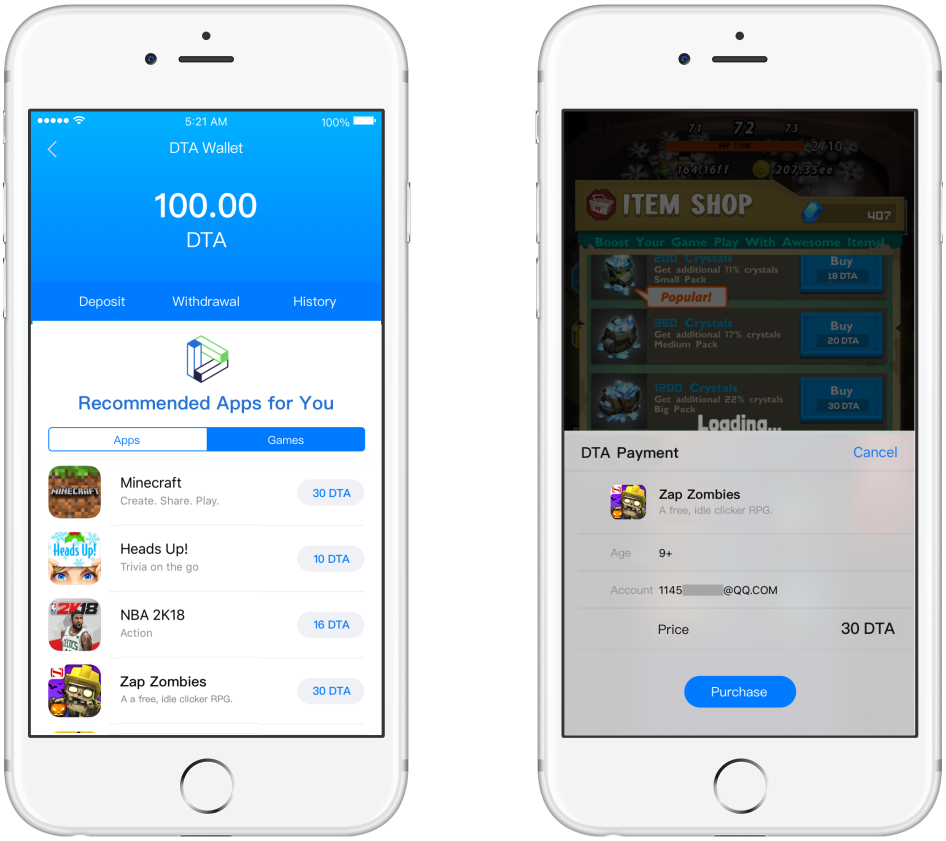
is from advertisers/ad networks/DSP to publishers/developers and then to end users. This is no different from traditional ad flow. In future development of DATA, when tokens substitute money in the ad flow, decentralized exchanges, bidding, DSP and SSP will be involved as discussed in the "Application layer" section.

DATA FLOW

is from user device to P2P based mobile storage. The SDK transforms individual devices into nodes of the mobile storage network. Then data in the storage is queried and validated by the DATA nodes.

TOKEN FLOW

is from DATA network to users, publishers/ developers as well as DATA Nodes. After DATA network validates data, detect fraud, it issues tokens as "coinbase transactions" to relevant contributors.



100.00 DTA

Recommended Apps for You

Minecraft
Create. Share. Play.
30 DTA

Heads Up!
Trivia on the go
10 DTA

NBA 2K18
Action
16 DTA

Zap Zombies
A a free, idle clicker RPG.
30 DTA

200 Crystals
Get additional 11% crystals
Small Pack
18 DTA

350 Crystals
Get additional 17% crystals
Medium Pack
20 DTA

1200 Crystals
Get additional 22% crystals
Big Pack
30 DTA

DTA Payment

Zap Zombies
A free, idle clicker RPG.

Age 9+

Account 1145[redacted]@QQ.COM

Price 30 DTA

Figure 11. DATA Coin Use Case

A Developer Registry, a DApp itself, allows developers and publishers to register, create wallets and manage their applications.

DATA can serve as a fundamental service layer in the decentralized and shared mobile ecosystem, and collaborate with other platforms running with DATA layer, such as decentralized App Stores and mobile SaaS platforms. Use of DATA Tokens can be applied to any App Store or platform that supports DTA as a third party payment, for users to purchase on premium services or IAP (in-app purchase) in apps or games. Figure 11 demonstrates a typical use case of DATA Tokens.

Innovation Partner

The vision and mission of DATA is only achievable through the support of platform partners and developers who trust the value of DATA Project. DATA is currently working with a strategic partner, Yomob International Co., Ltd., a unified mobile ad global monetization platform to develop the blockchain platform and early applications.

Yomob is a global mobile SaaS platform focusing on monetization service for developers. Its mission is to empower mobile developers to optimize and monetize efficiently based on its programmatic IO experience, big data and AI technology. The company is based in San Fransisco, Los Angels USA and has offices in Beijing, Guangzhou and Berlin Germany. Yomob partners with over forty leading global mobile ad networks and DSPs such as mobile video advertisements, playable advertisements and AR advertisements demand side platforms, including Facebook, Google, Bluefocus, Unity, Vungle and etc. Yomob has served more than 1 billion end users worldwide from over 2000 mobile developers on its platform while the numbers of both users and developers are still growing dramatically.

Yomob's Technology

SAAS

Yomob provides professional BI and SaaS platform services focusing on advertising and in-app purchase monetization optimization.

BIG DATA

The Yomob team has industry experience in big data. The team operates a global distributed cloud based network for supporting big data service on Yomob. Currently Yomob global big data platform can process tens of billions of data per day.

ARTIFICIAL INTELLIGENCE

The Yomob team has extensive experience in utilizing AI technology in monetization of mobile applications. With a team of industry leading technologists and experts, Yomob is proficient in applying cutting-edge artificial intelligence technologies in monetization.

Yomob's Business

The next generation mobile ad format has increasingly become the top performer of mobile advertisements for games and applications – but along with it, the cost for developers to manage it is also on the rise. With the application of machine learning and AI, Yomob offers a solution: global mobile advertisement monetization by matching and optimizing the flow of buyer demand with seller inventory.

Yomob coined the phrase “MaaS,” or Monetization as A Service, to describe the objective: to take a developer audience and monetize with higher return. Yomob has risen quickly as the leader in optimizing the placement of advertisements, substantially increasing the advertisements revenue stream for mobile developers.

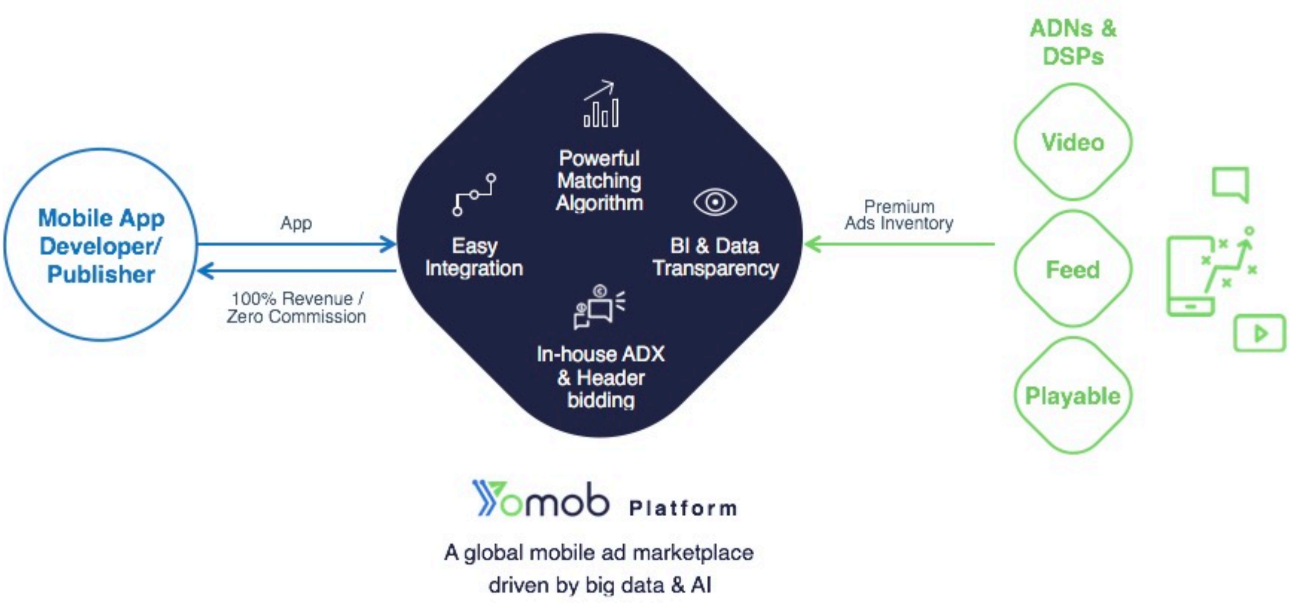

Mobile App Developer/ Publisher
App
100% Revenue / Zero Commission
Powerful Matching Algorithm
Easy Integration
BI & Data Transparency
In-house ADX & Header bidding
Premium Ads Inventory
ADNs & DSPs
Video
Feed
Playable
Yomob Platform
A global mobile ad marketplace driven by big data & AI

Figure 12. Illustration of Yomob’s Business

Ineffective advertising has always been a key problem in retaining the best user experiences and achieving the best timing and placement for every advertisement across every demand-side advertiser and ad network SDK -- from the cost of testing multiple iterations to the opportunity cost of siphoning off the time and attention that mobile game and application developers should be spending on their core competencies.

As an ad exchange platform, Yomob focuses on video advertisement at the current stage and is integrated with more than 50 major video ad networks and DSPs. As an aggregated ADX, Yomob can claim impartiality through a wide range of advertiser bids. AI-powered programmatic head bidding and a BI dashboard also offer greater transparency, so that both sellers and publishers can see the payout, or remove unsuitable ads.

Yomob’s approach integrates an understanding of the applications’s user, the users’ psychology, and the nature of the ad to match them at the right time, with the right users, at the time when they will be most interested in clicking. The focus is achieving the best performance and least disruption for users, within applications.

In order to reduce costs, improve user experience, and boost revenue, Yomob offers what it calls a “zero rev share” initiative, in which the developer gets the full payout from the advertising campaign that is displayed and watched by the users.

Team

Josh Burns

Josh is an experienced consumer internet business leader focused on helping to drive revenue, reach and engagement for both early to mid-stage consumer internet companies developing compelling products for both mobile and desktop as well as businesses from traditional industry entering the mobile landscape. He has a specific focus on the video game industry and melds a data driven approach coupled with a creative eye to drive increased product engagement and revenue. He has deep experience with partnerships as well as international markets in the mobile gaming sector. Previously, he worked at 6waves, co-founding the US office and leading the US product management team for one of the largest publishers of games for Facebook, iOS & Android, where he managed & launched over 100 apps including those from top developers like Kabam, Nexon & Atari, as well as games based on IP from Eminem, Disney, Dungeons & Dragons, Starz & BBC. Prior to 6waves, he worked at Electronic Arts in EA's Pogo.com division, supporting one of the largest casual game websites, focusing on product management, customer insights, new platforms, market strategy & analytics. Prior to his work at Electronic Arts, Josh worked at various companies providing market research and strategy consulting to Fortune 100 companies in more traditional industries.

Shirley Lin

Ms. Lin is a serial entrepreneur, a seasoned executive in business development in the mobile AdTech sector serving mobile game/app industry, with extensive relations in Silicon Valley, Europe and China, and specializing in globalization for cross border operations. Ms. Lin was VP of Business Development at Nexway (French), Yeahmobi (one of the top Chinese mobile performance marketing networks) responsible for the global expansion in San Francisco and Berlin. She also served at iConsole.tv and Beintoo (Italian) in the executive rank. Ms. Lin has technical experiences in operating system level of programming and was an aerospace software engineer in Space Shuttle Program at NASA, Houston. Shirley holds a M.S. in Computer Science/Math/Statistic from Texas A&M University and a B.A. in History from National Taiwan University.

Dr. Eric Li

Dr. Li worked as principal data scientist and architect in Microsoft Azure, Capital One and MicroStrategy in U.S. focusing on artificial intelligent, machine learning, cloud computing, business intelligent and big data system design and implementations, FinTech, P2P system and etc. Dr. Li has published more than 50 peer reviewed papers in the fields of peer to peer networks, distributed systems, online reputation management, social networks and big data with more than 800 citations, such as ACM/IEEE Transactions on Networking, IEEE Transactions on Distributed System, IEEE Transactions on Mobile Computing, IEEE Transaction on Computers, ACM Multimedia and etc. He also has multiple US patents. He served as committee member of multiple international conferences. He was the recipient of Chinese Government Award for Outstanding Students Abroad 2011 and Harris Outstanding Researcher Award in Clemson University 2011. Dr. Li received Ph.D of Computer Engineering from Clemson University, US and BS from Huazhong University of Science and Technology.

Victor Ye

Mr. Ye has worked for LinkedIn, Twitter and Snapchat in the U.S. as senior software engineer. Victor has technical expertise in multiple technical dimensions and years of cutting-edge R & D experiences. As academic experience, Mr. Ye contributed to the fields of data privacy, temporal-spatial data management and distributed database systems with more than 10 peer reviewed publications and multiple patents. During his time working in the industry, he worked on multiple cutting-edge open source distributed systems including Apache Kafka and project Manhattan in Twitter. Mr. Ye has received Master of Computer Science from Columbia University and Bachelor of Science from Tsinghua University.

Franklin Song

Mr. Song is a serial entrepreneur and cofounded several start-ups including a global mobile game developing and publishing company SOULGAME which earned the most Apple App Store global game features among all Chinese companies. Mr. Song has also worked for Microsoft Research Asia and Oracle in the U.S. Mr. Song joined Decentralized and Distributed Systems Research at Yale University advised by Prof. Bryan Ford and worked on Dissent Project (a Dining-cryptographers Shuffled Send Network, which is a protocol for accountable anonymous messaging, voting, and other interactions among members of a decentralized group). Mr. Song has received MS from Yale University and BS from Tsinghua University.

Henry Zhao

Mr. Zhao worked at Perfect World and Tencent. During that period, Mr. Zhao led the operating and marketing teams for several extreme successful mobile games with total revenue over $100 mil. Mr. Zhao has received both BS/MS of Engineering in Tsinghua University.

Han Liao

Mr. Han worked as supply lead of Vungle and drove tens of times revenue growth for Vungle in China. Han also served as Director of Global BD for WQ Mobile. Mr. Han received M.B.A. from University of La Verne, advanced in Marketing Planning and Financial Analyzing.

Ashley Zhou

Ms. Zhou worked as product marketer in LinkedIn and led the marketing & operation of LinkedIn career product and LinkedIn digital advertising sources. Ms. Zhou also served as digital data analyst in Goodby Silverstein & Partners and provided social media strategy & solution for Cisco. Ms. Zhou received Master of Integrated Marketing Communications from Northwestern University and Bachelor from Renmin University.

Board Of Advisors

Prof. Shoucheng Zhang (Chief Advisor)

Shoucheng Zhang is the JG Jackson and CJ Wood professor of physics at Stanford University. He is a condensed matter theorist known for his work on topological insulators, quantum spin Hall effect, spintronics, quantum Hall effect and high temperature superconductivity. He is a fellow of the American Physical Society and a fellow of the American Academy of Arts and Sciences. He received the Guggenheim fellowship in 2007, the Alexander von Humboldt research prize in 2009, the Europhysics prize in 2010, the Oliver Buckley prize in 2012, the Dirac Medal and Prize in 2012, the Physics Frontiers Prize in 2013, the "Nobel-class" Citation Laureates by Thomson Reuters in 2014 and the Benjamin Franklin Medal in 2015. He is identified as one of the top candidates for the Nobel Prize by Thomson Reuters in 2014. He has been elected as the member of the National Academy of Science in 2015. Mr. Zhang is also the founding chairman of Danhua Capital, a VC fund that invests primarily in early stage and growth stage company with disruptive technology/business model, big market and excellent team.

Michael Arrington

J. Michael Arrington is the American founder and former co-editor of TechCrunch, a blog covering the Silicon Valley technology start-up communities and the wider technology field in USA and elsewhere. Magazines such as Wired and Forbes have named Arrington one of the most powerful people on the Internet. In 2008, he was selected by TIME Magazine as one of the most influential people in the world. In 2017, he founded Arrington XRP Capital, a digital asset management firm in blockchain-based capital markets.

Shuoji Zhou

Zhou Shuoji is the founding partner of FBG Capital, with extensive experience in digital assets trading and investment. Vincent is also an early investor of a broad spectrum of blockchain companies and projects. He is considered as one of the most well-connected and visionary crypto hedge fund managers in Asia.

Bo Shen

Founding Partner of Fenbushi Capital and Blockasset Fund.

Paul Veradittakit

Paul is partner of Pantera. Prior to joining Pantera in 2014, Paul worked at Strive Capital as an Associate focusing on investments in the mobile space. Previously, he was at Hatch Consulting and LECG and performed business development and marketing for Urban Spoils, an early stage startup in the daily deal aggregation space. Paul graduated from the University of California, Berkeley with a B.A. in Psychology and a B.A. in Political Science.

Junfei Ren

Junfei Ren is Secretary of the Board of Directors of Huobi Global, world's largest crypto exchange. She is also the founder of Huobi Labs, a blockchain incubator for early stage startups. Prior to joining Huobi, Mrs. Ren worked as secretary of Mr. Xiaoping Xu, Founder of ZhenFund. Mrs. Ren has also cofounded a blockchain based fin-tech company with successful exit.

Yusen Dai

Yusen is a partner of ZhenFund. Before joining ZhenFund, Yusen dropped out of Stanford University in 2009 and cofounded Jumei International LTD (NYSE: JMEI). Jumei is the leading Chinese cosmetics e-commerce retailer and one of the largest Chinese e-commerce companies. Yusen oversaw Jumei's product design and development, operation management, online marketing, and management of several categories. Jumei was listed on NYSE in May 2014 at a market cap of $3 billion, only four years after its inception. Yusen attended Stanford University Graduate School, and received his bachelor degree from Tsinghua University, Beijing.

Bman Lee

Consultant of Alibaba (NYSE: BABA), Lenovo (HKG: 0992), Qihoo (NYSE:QIHU) Co-founder of Lijiaoshou (Acquired by Baidu) Lead of Baidu AOD platform (served 500,000+ businesses) Active Investor in Blockchain and active blockchain community contributor since 2013

Satoshi

Satoshi is Partner of DFund. He was VP at Top investment bank in Japan, PM at Top telecom company in Japan and MD at Global investment funds in Hong Kong and Japan

Kevin Wen

C.S., UT-Ausin Serial Entrepreneur; Co-founder of LightInTheBox.com (NYSE:LITB); Co-founder of Blogchina

Scientists

Jia Tian

Mr. Tian served as senior developer at Baidu, Inc. and Alibaba, Inc. During that period, he developed large scale computer systems including the technology behind so.com, supporting more than 100 million page views per day, and large scale recommendation systems. Mr. Tian is also a serial entrepreneur. He joined the founding team of several companies focusing on AI and related areas. The first company Wolongyun was acquired by Alibaba, Inc. Thereafter he joined the Beijing Machine Learning Information Technology company as CTO, designed and built AI systems including recommendation systems, chatbots, medical image recognition systems. After that he joined Pony.ai, invested by Sequoia Capital and IDG Capital in the angel round, to build autonomous driving systems. Mr. Tian serves as the Chief Scientist in bitfundpe.com, a bitcoin fund dedicated to supporting the bitcoin community since 2013, founded by Xiaolai Li. Mr. Tian is also advisor to multiple blockchain tech startups such as ZCash. Mr. Tian received BS/MS of Engineering in Tsinghua University, majoring in distributed systems.

Dr. Harry Liu

Harry is Staff Engineer / Senior Manager of Research at Alibaba Group. He was a researcher in Microsoft Research Redmond Lab from 2014 to 2017. Harry's research focus is on designing and implementing fundamental network protocols, algorithms and systems for the future world with prevalent BigData, Internet of Things and Virtual Reality applications with areas including cloud computing, edge computing, network functions virtualization, software-defined networking, network management driven by BigData, online-service and content delivery, ultra-low-latency networking, peer-to-peer networking. Harry has published research papers in top conference like SOSP, NSDI and SIGCOMM as first author. Harry also won the "2014 ACM SIGCOMM Doctoral Dissertation Award - Honorable Mention" which is the only honorable mention of this award in this year. He obtained his Ph.D. in Dept. of Computer Science at Yale University. Previously he received the B.S and M.S. degrees from Tsinghua University.

Investors & Partners

Investors & Partners

FBG F|B|G CAPITAL

PANTERA

Huobi Labs

ZhenFund

DHVC

arrington XRP CAPITAL

ALPHABIT DIGITAL CURRENCY FUND

Kenetic Capital

BLOCKTOWER

DFund

INBlockchain

Ce Yuan Ventures

METROPOLIS VC

LINKVC

ChinaGrowthCapital

ZMT Capital

Nirvana Capital

SCIENCE

NODE CAPITAL

CAPITAL YOUBI Your Coin, Best Coin

GENESIS

cointime

BiTs Angel

Bixin

BiShiJie.com

GMIC

50+ Global Ad Network & DSPs Partners of Yomob

Audience Network

AdMob by Google

Tencent Social Ads

unity ADS

Vungle

APPLOVIN

Mobvista.

BlueFocus

ADCOLONY

LEADBOLT

CHANCE

LoopMe

INMOBI

smaato

Avazu

appnext

ironSource

WWW.KSYUN.COM

Oneway.mobi

Tapjoy

YOU·APPI

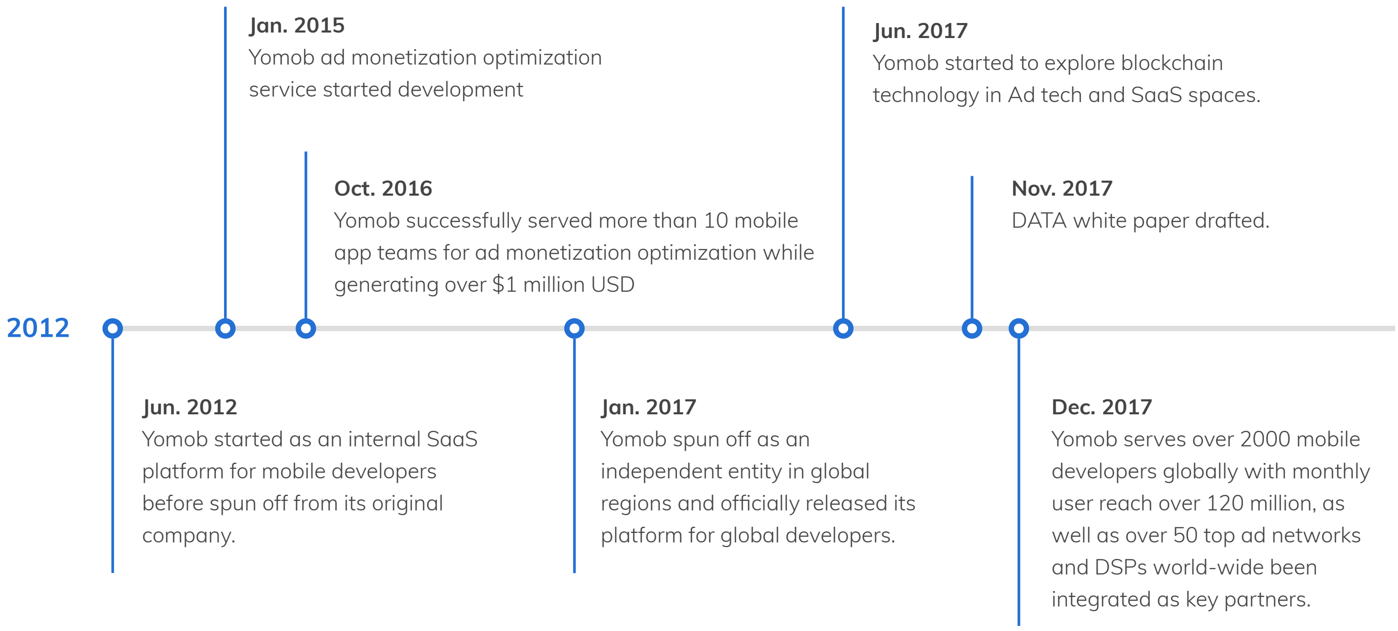
Chartboost

Centrixlink

sunteng

Roadmap

BOOTSTRAP (YOMOB)



2012

Jun. 2012
Yomob started as an internal SaaS platform for mobile developers before spun off from its original company.

Jan. 2015
Yomob ad monetization optimization service started development

Oct. 2016
Yomob successfully served more than 10 mobile app teams for ad monetization optimization while generating over $1 million USD

Jan. 2017
Yomob spun off as an independent entity in global regions and officially released its platform for global developers.

Jun. 2017
Yomob started to explore blockchain technology in Ad tech and SaaS spaces.

Nov. 2017
DATA white paper drafted.

Dec. 2017
Yomob serves over 2000 mobile developers globally with monthly user reach over 120 million, as well as over 50 top ad networks and DSPs world-wide been integrated as key partners.

PHASE I: DEVELOPMENT AND MARKETING

2018 Q1 Blockchain Data Foundation founded.

2018 Q1 DATA Project institution-only private token sale as ERC20 token on Ethererum.
Marketing strategy formed.

2018 Q2 DATA first prototype initial testing including DTA distribution system and etc.

PHASE II: INTERNAL RELEASE TO TEST ON YOMOB

2018 Q3 Alpha launch of DATA Chain and system, integrated inside Yomob's SDK (DATA version).
First Go-to-market strategy executed.

2018 Q3 Finish Closed Alpha test with selected participants from developers on Yomob platform.
Distribute DTA with developers, users and miners.

2018 Q4 Conversion of ERC20 based DATA Token to native token.
Closed Beta launch of DATA on Yomob platform with most of its developers.

2019 Q1 Open Beta launch of DATA on Yomob platform. Release Beta version of DATA SDK Protocol.

PHASE III: SOFT LAUNCH TO EARLY ADOPTERS

2019 Q1 Start industry collaboration with initial strategic partners including both ad networks and developers.

Start building DATA ecosystem and generate DTA.

2019 Q1 Release Alpha Version of ***M***^{3} for initial testing.

PHASE IV: LAUNCH FOR BROADER ADOPTION THROUGH ECOSYSTEM

2019 Q2 Fully functional DATA system to include wallet, micropayment.

Provide external APIs / SDK solutions for developers to enable in-app purchase for virtual goods and services to their end users.

2019 Q2 Release Beta Version of ***M***^{3}.

2019 Q3 Partner with selected industry strategic partners such as mobile SaaS service providers, App Stores, developers, ad networks , DSPs, ad exchanges, SSPs and etc. for utilizing DTA as a currency for more use cases.

2019 Q4 Official launch of DATA Project and parallel marketing promotion for broad industry adoption.
Open invite to engage partners up- and down-stream of the ecosystem.

References

1. Mobile Internet Ad Spending Worldwide, 2015-2020
   www.eMarketer.com

2. AdMaster Anti-Fraud White Paper
   http://www.admaster.com.cn/eng/index.php?c=downloads&a=view&id=102

3. Basic Attention Token
   https://basicattentiontoken.org/

4. AdChain Project
   https://www.adchain.com/

5. AdEx Project
   https://www.adex.network/

6. Lumar Partners
   https://www.lumapartners.com/lumascapes/display-ad-tech-lumascape/

7. RTB China
   https://rtbchina.com

8. PageFaire 2017 Adblock Report
   https://pagefair.com/blog/2017/adblockreport/

9. “Ad Fraud Estimates Double”. WPP. Business Insider. March 16, 2017.
   http://www.businessinsider.com/ad-fraud-estimates-doubled-2017-3

10. ads.txt by IAD lab
    https://iabtechlab.com/ads-txt-about/

11. Tendermint Consensus Engine
    https://github.com/tendermint/tendermint/wiki

12. EtherMint: Ethereum on Tendermint
https://blog.cosmos.network/a-beginners-guide-to-ethermint-38ee15f8a6f4

13. Raiden Network
http://raiden.network/

14. Lightning Network
http://lightning.network/

15. Zilliqa
https://www.zilliqa.com/

16. IPFS
https://ipfs.io/

17 Chord
https://github.com/sit/dht/wiki

18. P. Maymounkov and D. Mazieres. Kademlia: A peer-to-peer information system based on the xor metric. In Peer-to-Peer Systems, pages 53–65. Springer, 2002.

19. BitTorrent
http://www.bittorrent.com/

20. Sepp Hochreiter; Jürgen Schmidhuber (1997). "Long short-term memory". Neural Computation. 9 (8): 1735–1780. doi:10.1162/neco.1997.9.8.1735. PMID 9377276.

21. SHA-1 https://tools.ietf.org/html/rfc3174

22. TensorFlow https://www.tensorflow.org/

23. Project Papyrus
https://papyrus.global/

24. Piano Tiles 2™
https://itunes.apple.com/us/app/piano-tiles-2/id1027688889?mt=8
https://play.google.com/store/apps/details?id=com.cmplay.tiles2

25. Apache Kafka
https://kafka.apache.org/

26. WebTorrent
https://github.com/webtorrent/bittorrent-dht

27. Cell Owners Under 50 Rarely Turn Their Phones Off
http://www.pewinternet.org/2015/08/26/americans-views-on-mobile-etiquette/2015-08-26_alone-together_1_01/

28. Ion Stoica , Robert Morris , David Karger , M. Frans Kaashoek , Hari Balakrishnan, Chord: A scalable peer-to-peer lookup service for internet applications, Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications, p.149-160, August 2001, San Diego, California, USA

29. The ad fraud issue could be more than twice as big as first thought — advertisers stand to lose $16.4 billion to it this year
http://www.businessinsider.com/thepartnership-msix-and-adloox-ad-fraud-2017-2017-3

Whitepaper Disclaimers

IMPORTANT: YOU MUST READ THE FOLLOWING DISCLAIMER IN FULL BEFORE CONTINUING

The sale ("**Token Sale**") of the DATA Token ("**DATA Tokens**"), the exchange medium for participants of the DATA platform as detailed in this whitepaper (the "**Whitepaper**") is only intended for, made to or directed at, only certain persons. Moreover, this Whitepaper is not a prospectus or offer document of any sort and is not intended to constitute an offer of securities of any form, units in a business trust, units in a collective investment scheme or any other form of investment, or a solicitation for any form of investment in any jurisdiction. No regulatory authority has examined or approved of any of the information set out in this Whitepaper. This Whitepaper has not been registered with any regulatory authority in any jurisdiction.

By accessing and/or accepting possession of any information in this Whitepaper or such part thereof (as the case may be), you represent and warrant to Blockchain Data Foundation (to be incorporated) (the "**Foundation**") and Wealth Wise Ventures Limited (BVI Company No.: 1965363) (the "**Token Issuer**") that:

(a) you are not located in the People's Republic of China and you are not a citizen or resident (tax or otherwise) of, or domiciled in, the People's Republic of China;

(b) you are not located in the Republic of Korea and you are not a citizen or resident (tax or otherwise) of, or domiciled in, the Republic of Korea;

(c) you are not located in the United States of America and you are not a citizen, resident (tax or otherwise) or green card holder of, or domiciled in, the United States of America, unless you are a U.S. Qualified Person (as defined in the Token Sale Terms (as defined herein));

(d) you are not located in a jurisdiction where the Token Sale is prohibited, restricted or unauthorized in any form or manner whether in full or in part under its laws, regulatory requirements or rules;

(e) you agree to be bound by the limitations and restrictions described herein; and

(f) you acknowledge that this Whitepaper has been prepared for delivery to you so as to assist you in making a decision as to whether to purchase the DATA Tokens.

IMPORTANT NOTICE

This Whitepaper in current form is being circulated for general information and to invite investor feedback only on the DATA platform as presently conceived, and is subject to review and revision by the directors of the Token Issuer and/or the Foundation, the advisers, and/or legal advisers of the Token Issuer and/or the Foundation. Please do not replicate or distribute any part of this Whitepaper without this note in accompaniment. No part of this Whitepaper is intended to create legal relations between a recipient of this Whitepaper or to be legally binding or enforceable by such recipient against the Token Issuer and/or the Foundation. An updated version of this Whitepaper may be published on a date to be determined and announced by the Token Issuer and/or the Foundation in due course.

PLEASE READ THIS SECTION AND THE FOLLOWING SECTIONS ENTITLED “DISCLAIMER OF LIABILITY”, “NO REPRESENTATIONS AND WARRANTIES”, “REPRESENTATIONS AND WARRANTIES BY YOU”, “CAUTIONARY NOTE ON FORWARD-LOOKING STATEMENTS”, “THIRD PARTY INFORMATION AND NO CONSENT OF OTHER PERSONS”, “TERMS USED”, “NO ADVICE”, “NO FURTHER INFORMATION OR UPDATE”, “RESTRICTIONS ON DISTRIBUTION AND DISSEMINATION”, “NO OFFER OF INVESTMENT OR REGISTRATION” AND “RISKS AND UNCERTAINTIES” CAREFULLY.

IF YOU ARE IN ANY DOUBT AS TO THE ACTION YOU SHOULD TAKE, YOU SHOULD CONSULT YOUR LEGAL, FINANCIAL, TAX OR OTHER PROFESSIONAL ADVISOR(S).

The DATA Tokens are not intended to constitute securities of any form, units in a business trust, units in a collective investment scheme or any other form of investment in any jurisdiction. This Whitepaper does not constitute a prospectus or offer document of any sort and is not intended to constitute an offer of securities of any form, units in a business trust, units in a collective investment scheme or any other form of investment, or a solicitation for any form of investment in any jurisdiction.

This Whitepaper does not constitute or form part of any opinion or any advice to acquire, sell, or any solicitation of any offer by the Foundation or the Token Issuer to acquire any DATA Tokens nor shall it or any part of it nor the fact of its presentation form the basis of, or be relied upon in connection with, any contract or investment decision.

The proceeds from the sale of the DATA Tokens will be deployed to support ongoing development and growth of the DATA platform, marketing, human resources, sales, and other operational activities.

No person is bound to enter into any contract or binding legal commitment in relation to the acquisition of DATA Tokens and no cryptocurrency or other form of payment is to be accepted on the basis of this Whitepaper.

Any agreement as between the Token Issuer and you as a participant in the sale of the DATA Tokens by the Token Issuer (the “**Token Sale**”), and in relation to any purchase of DATA Tokens is to be governed by only a separate document setting out the terms and conditions (the “**Token Sale Terms**”) of such agreement. In the event of any inconsistencies between the Token Sale Terms and this Whitepaper, the former shall prevail.

PLEASE NOTE THAT YOU ARE NOT ELIGIBLE AND YOU ARE NOT TO PURCHASE ANY TOKENS IN THE TOKEN SALE IF: (A) YOU ARE LOCATED IN THE PEOPLE'S REPUBLIC OF CHINA OR IF YOU ARE A CITIZEN OR RESIDENT (TAX OR OTHERWISE) OF, OR DOMICILED IN, THE PEOPLE'S REPUBLIC OF CHINA; (B) YOU ARE LOCATED IN THE REPUBLIC OF KOREA OR IF YOU ARE A CITIZEN OR RESIDENT (TAX OR OTHERWISE) OF, OR DOMICILED IN, THE REPUBLIC OF KOREA; (C) YOU ARE LOCATED IN THE UNITED STATES OF AMERICA OR IF YOU ARE A CITIZEN, RESIDENT (TAX OR OTHERWISE) OR GREEN CARD HOLDER OF, OR DOMICILED IN, THE UNITED STATES OF AMERICA, UNLESS YOU ARE A U.S. QUAILED PERSON; OR (D) SUCH TOKEN SALE IS PROHIBITED, RESTRICTED OR UNAUTHORIZED IN ANY FORM OR MANNER WHETHER IN FULL OR IN PART UNDER THE LAWS, REGULATORY REQUIREMENTS OR RULES IN THE JURISDICTION IN WHICH YOU ARE LOCATED, AT THE TIME OF YOUR INTENDED PURCHASE OR PURCHASE OF THE TOKENS IN THE TOKEN SALE.

No regulatory authority has examined or approved of any of the information set out in this Whitepaper. No such action has been or will be taken under the laws, regulatory requirements or rules of any jurisdiction. The publication, distribution or dissemination of this Whitepaper does not imply that the applicable laws, regulatory requirements or rules have been complied with.

There are risks and uncertainties associated with the Foundation and the Token Issuer and their business and operations, the DATA Tokens, the Token Sale, and the DATA platform. Please refer to the section entitled “Risks and Disclosures” set out at the end of this Whitepaper.

This Whitepaper, any part thereof and any copy thereof must not be taken or transmitted to any country where distribution or dissemination of this Whitepaper is prohibited or restricted.

No part of this Whitepaper is to be reproduced, distributed or disseminated without including this section and the following sections entitled “Disclaimer of Liability”, “No Representations and Warranties”, “Representations and Warranties By You”, “Cautionary Note On Forward-Looking Statements”, “Third Party Information and No Consent of Other Persons”, “Terms Used”, “No Advice”, “No Further Information or Update”, “Restrictions On Distribution and Dissemination”, “No Offer of Investment Or Registration” and “Risks and Uncertainties”.

DISCLAIMER OF LIABILITY

To the maximum extent permitted by the applicable laws, regulations and rules, the Foundation and/or the Token Issuer shall not be liable for any indirect, special, incidental, consequential or other losses of any kind, in tort, contract or otherwise (including but not limited to loss of revenue, income or profits, and loss of use or data), arising out of or in connection with any acceptance of or reliance on this Whitepaper or any part thereof by you.

NO REPRESENTATIONS AND WARRANTIES

The Foundation and the Token Issuer do not make or purport to make, and hereby disclaim, any representation, warranty or undertaking in any form whatsoever to any entity or person, including any representation, warranty or undertaking in relation to the truth, accuracy and completeness of any of the information set out in this Whitepaper.

REPRESENTATIONS AND WARRANTIES BY YOU

By accessing and/or accepting possession of any information in this Whitepaper or such part thereof (as the case may be), you represent and warrant to the Foundation and the Token Issuer as follows:

(a) you agree and acknowledge that the DATA Tokens do not constitute securities of any form, units in a business trust, units in a collective investment scheme or any other form of investment in any jurisdiction;

(b) you are not:

(i) located in the People's Republic of China or a citizen or resident (tax or otherwise) of, or domiciled in, the People's Republic of China;

(ii) located in the Republic of Korea or a citizen or resident (tax or otherwise) of, or domiciled in, the Republic of Korea;

(iii) located in the United States of America or a citizen, resident (tax or otherwise) or green card holder of, or domiciled in, the United States of America not being a U.S. Qualified Person; or

(iv) located in a jurisdiction where the Token Sale is prohibited, restricted or unauthorized in any form or manner whether in full or in part under the laws, regulatory requirements or rules in such jurisdiction;

(c) you agree and acknowledge that this Whitepaper does not constitute a prospectus or offer document of any sort and is not intended to constitute an offer of securities of any form, units in a business trust, units in a collective investment scheme or any other form of investment in any jurisdiction, or a solicitation for any form of investment, and you are not bound to enter into any contract or binding legal commitment and no cryptocurrency or other form of payment is to be accepted on the basis of this Whitepaper;

(d) you acknowledge and understand that no DATA Token should be construed, interpreted, classified or treated as enabling, or according any opportunity to, token holders to participate in or receive profits, income, or other payments or returns arising from or in connection with the DATA Tokens or the proceeds of the Token Sale, or to receive sums paid out of such profits, income, or other payments or returns;

(e) you agree and acknowledge that no regulatory authority has examined or approved of the information set out in this Whitepaper, no action has been or will be taken under the laws, regulatory requirements or rules of any jurisdiction and the publication, distribution or dissemination of this Whitepaper to you does not imply that the applicable laws, regulatory requirements or rules have been complied with;

(f) you agree and acknowledge that this Whitepaper, the undertaking and/or the completion of the Token Sale, or future trading of DATA Tokens on any cryptocurrency exchange, shall not be construed, interpreted or deemed by you as an indication of the merits of the Foundation, the Token Issuer, the DATA Tokens, the Token Sale, and the DATA platform;

(g) the distribution or dissemination of this Whitepaper, any part thereof or any copy thereof, or acceptance of the same by you, is not prohibited or restricted by the applicable laws, regulations or rules in your jurisdiction, and where any restrictions in relation to possession are applicable, you have observed and complied with all such restrictions at your own expense and without liability to the Foundation and/or the Token Issuer;

(h) you agree and acknowledge that in the case where you wish to acquire any DATA Tokens, DATA Tokens are not to be construed, interpreted, classified or treated as:

(i) any kind of currency other than cryptocurrency;

(ii) debentures, stocks or shares issued by any person or entity;

(iii) rights, options or derivatives in respect of such debentures, stocks or shares;

(iv) rights under a contract for differences or under any other contract the purpose or pretended purpose of which is to secure a profit or avoid a loss;

(v) units in a collective investment scheme;

(vi) units in a business trust;

(vii) derivatives of units in a business trust; or

(viii) any form of investment;

(i) you are legally permitted to participate in the Token Sale and all actions contemplated or associated with such participation, including the holding and use of DATA Tokens;

(j) the amounts that you use to acquire the DATA Tokens were not and are not directly or indirectly derived from any activities that contravene the laws and regulations of any jurisdiction, including anti-money laundering laws and regulations;

(k) if you are a natural person, you are of sufficient age and capacity under the applicable laws of the jurisdiction in which you reside and the jurisdiction of which you are a citizen to participate in the Token Sale;

(l) you are not obtaining or using DATA Tokens for any illegal purpose;

(m) you have a basic degree of understanding of the operation, functionality, usage, storage, transmission mechanisms and other material characteristics of cryptocurrencies, blockchain-based software systems, cryptocurrency wallets or other related token storage mechanisms, blockchain technology, and smart contract technology;

(n) you are fully aware and understand that in the case where you wish to purchase any DATA Tokens, there are risks associated with the Foundation and the Token Issuer and their business and operations, DATA Tokens, the Token Sale, and the DATA platform;

(o) you bear the sole responsibility to determine what tax implications a purchase of DATA Tokens may have for you and agree not to hold the Foundation, the Token Issuer and/or any other person involved in the Token Sale liable for any tax liability associated with or arising therefrom;

(p) you agree and acknowledge that the Foundation and the Token Issuer are not liable for any direct, indirect, special, incidental, consequential or other losses of any kind, in tort, contract or otherwise (including but not limited to loss of revenue, income or profits, and loss of use or data), arising out of or in connection with any acceptance of or reliance on this Whitepaper or any part thereof by you;

(q) you waive the right to participate in a class action lawsuit or a class wide arbitration against the Foundation, the Token Issuer and/or any person involved in the Token Sale and/or with the creation and distribution of DATA Tokens; and

(r) all of the above representations and warranties are true, complete, accurate and non-misleading from the time of your access to and/or acceptance of possession this Whitepaper or such part thereof (as the case may be).

CAUTIONARY NOTE ON FORWARD-LOOKING STATEMENTS

All statements contained in this Whitepaper, statements made in press releases or in any place accessible by the public and oral statements that may be made by the Foundation or the Token Issuer or their respective directors, executive officers or employees acting on behalf of the Foundation and/or the Token Issuer (as the case may be), that are not statements of historical fact, constitute “forward-looking statements”. Some of these statements can be identified by forward-looking terms such as “aim”, “target”, “anticipate”, “believe”, “could”, “estimate”, “expect”, “if”, “intend”, “may”, “plan”, “possible”, “probable”, “project”, “should”, “would”, “will” or other similar terms. However, these terms are not the exclusive means of identifying forward-looking statements. All statements regarding the Foundation and/or the Token Issuer’s business strategies, plans and prospects and the future prospects of the industry which the Foundation and/or the Token Issuer is in are forward-looking statements. These forward-looking statements, including but not limited to statements as to the Foundation and/or the Token Issuer’s prospects, future plans, other expected industry trends and other matters discussed in this Whitepaper regarding the Foundation and/or the Token Issuer are matters that are not historic facts, but only predictions.

These forward-looking statements involve known and unknown risks, uncertainties and other factors that may cause the actual future results, performance or achievements of the Foundation and/or the Token Issuer to be materially different from any future results, performance or achievements expected, expressed or implied by such forward-looking statements. These factors include, amongst others:

(a) changes in political, social, economic and stock or cryptocurrency market conditions, and the regulatory environment in the countries in which the Foundation and/or the Token Issuer conduct their business and operations;

(b) the risk that the Foundation may be unable to execute or implement their business strategies and future plans;

(c) changes in interest rates and exchange rates of fiat currencies and cryptocurrencies;

(d) changes in the anticipated growth strategies and expected internal growth of the Foundation and the DATA platform;

(e) changes in the availability and fees payable to the Foundation in connection with its businesses and operations or in the DATA platform;

(f) changes in the availability and salaries of employees who are required by the Foundation and/or the Token Issuer to operate their business and operations;

(g) changes in preferences of users of the DATA platform;

(h) changes in competitive conditions under which the Foundation operates, and the ability of the Foundation to compete under such conditions;

(i) changes in the future capital needs of the Foundation and the availability of financing and capital to fund such needs;

(j) war or acts of international or domestic terrorism;

(k) occurrences of catastrophic events, natural disasters and acts of God that affect the businesses and/or operations of the Foundation and/or the Token Issuer;

(l) other factors beyond the control of the Foundation and/or the Token Issuer; and

(m) any risk and uncertainties associated with the Foundation and the Token Issuer and their business and operations, the DATA Tokens, the Token Sale, and the DATA platform.

All forward-looking statements made by or attributable to the Foundation, the Token Issuer and/or persons acting on behalf of the Foundation and/or the Token Issuer are expressly qualified in their entirety by such factors. Given that risks and uncertainties that may cause the actual future results, performance or achievements of the Foundation and/or the Token Issuer to be materially different from that expected, expressed or implied by the forward-looking statements in this Whitepaper, undue reliance must not be placed on these statements. These forward-looking statements are applicable only as of the date of this Whitepaper.

None of the Foundation, the Token Issuer or any other person represents, warrants, and/or undertakes that the actual future results, performance or achievements of the Foundation and/or the Token Issuer will be as discussed in those forward-looking statements. The actual results, performance or achievements of the Foundation and/or the Token Issuer may differ materially from those anticipated in these forward-looking statements.

Nothing contained in this Whitepaper is or may be relied upon as a promise, representation or undertaking as to the future performance or policies of the Foundation and/or the Token Issuer.

Further, the Foundation and the Token Issuer disclaim any responsibility to update any of those forward-looking statements or publicly announce any revisions to those forward-looking statements to reflect future developments, events or circumstances, even if new information becomes available or other events occur in the future.

THIRD PARTY INFORMATION AND NO CONSENT OF OTHER PERSONS

This Whitepaper includes information obtained from various third party sources ("**Third Party Information**"). None of the publishers of Third Party Information has consented to the inclusion of Third Party Information in this Whitepaper and is therefore not liable for Third Party Information. While reasonable action has been taken to ensure that Third Party Information has been included in their proper form and context, neither the Foundation, the Token Issuer nor their respective directors, executive officers, and employees acting on their behalf, has independently verified the accuracy, reliability, completeness of the contents, or ascertained any applicable underlying assumption, of the relevant Third Party Information. Consequently, neither the Foundation, the Token Issuer nor their respective directors, executive officers and employees acting on their behalf makes any representation or warranty as to the accuracy, reliability or completeness of such information and shall not be obliged to provide any updates on the same.

TERMS USED

To facilitate a better understanding of the DATA Tokens being the subject of the sale conducted by the Token Issuer, and the business and operations of the Foundation and the Token Issuer, certain technical terms and abbreviations, as well as, in certain instances, their descriptions, have been used in this Whitepaper. These descriptions and assigned meanings should not be treated as being definitive of their meanings and may not correspond to standard industry meanings or usage.

Words importing the singular shall, where applicable, include the plural and vice versa and words importing the masculine gender shall, where applicable, include the feminine and neuter genders and vice versa. References to persons shall include corporations.

NO ADVICE

No information in this Whitepaper should be considered to be business, legal, financial or tax advice regarding the Foundation, the Token Issuer, the DATA Tokens, the Token Sale, or the DATA platform. You should consult your own legal, financial, tax or other professional adviser regarding the Foundation and the Token Issuer and their business and operations, the DATA Tokens, the Token Sale, and the DATA platform. You should be aware that you may be required to bear the financial risk of any purchase of DATA Tokens for an indefinite period of time.

NO FURTHER INFORMATION OR UPDATE

No person has been or is authorized to give any information or representation not contained in this Whitepaper in connection with the Foundation and the Token Issuer and their business and operations, the DATA Tokens, the Token Sale, or the DATA platform, if given, such information or representation must not be relied upon as having been authorized by or on behalf of the Foundation or the Token Issuer. The Token Sale shall not, under any circumstances, constitute a continuing representation or create any suggestion or implication that there has been no change, or development reasonably likely to involve a material change in the affairs, conditions and prospects of the Foundation and/or the Token Issuer or in any statement of fact or information contained in this Whitepaper since the date hereof.

RESTRICTIONS ON DISTRIBUTION AND DISSEMINATION

The distribution or dissemination of this Whitepaper or any part thereof may be prohibited or restricted by the laws, regulatory requirements, and rules of any jurisdiction. In the case where any restriction applies, you are to inform yourself about, and to observe, any restrictions which are applicable to your possession of this Whitepaper or such part thereof (as the case may be) at your own expense and without liability to the Foundation and/or the Token Issuer.

Persons to whom a copy of this Whitepaper has been distributed or disseminated, provided access to or who otherwise have the Whitepaper in their possession shall not circulate it to any other persons, reproduce or otherwise distribute this Whitepaper or any information contained herein for any purpose whatsoever nor permit or cause the same to occur.

NO OFFER OF INVESTMENT OR REGISTRATION

This Whitepaper does not constitute a prospectus or offer document of any sort and is not intended to constitute an offer of securities of any form, units in a business trust, units in a collective investment scheme or any other form of investment, or a solicitation for any form of investment in any jurisdiction. No person is bound to enter into any contract or binding legal commitment and no cryptocurrency or other form of payment is to be accepted on the basis of this Whitepaper.

No regulatory authority has examined or approved of any of the information set out in this Whitepaper. No such action has been or will be taken under the laws, regulatory requirements or rules of any jurisdiction. The publication, distribution or dissemination of this Whitepaper does not imply that the applicable laws, regulatory requirements or rules have been complied with.

RISKS AND UNCERTAINTIES

Prospective purchasers of DATA Tokens should carefully consider and evaluate all risks and uncertainties associated with the Foundation and the Token Issuer and their business and operations, the DATA Tokens, the Token Sale, and the DATA platform, all information set out in this Whitepaper and the Token Sale Terms prior to any purchase of the DATA Tokens. If any of such risks and uncertainties develops into actual events, the business, financial condition, results of operations and prospects of the Foundation and/or the Token Issuer could be materially and adversely affected. In such cases, you may lose all or part of the value of the DATA Tokens. Please refer to the risk factors set out in Exhibit 2 of the Token Sale Terms.