A New Approach to Crowd Journalism Using a Blockchain-Based Infrastructure

Leonardo Teixeira
FEUP and MOG Technologies
Porto, Portugal
up201502848@fe.up.pt

Ivone Amorim
CMUP and MOG Technologies
Porto, Portugal
ivone.amorim@mog-technologies.com

Alexandre Ulisses Silva
MOG Technologies
Maia, Portugal
alexandre.ulisses@mog-technologies.com

João Correia Lopes
FEUP - INESC TEC
Porto, Portugal
jlopes@fe.up.pt

Vasco Filipe
MOG Technologies
Maia, Portugal
vasco.filipe@mog-technologies.com

ABSTRACT

The significant evolution of smartphones has given ordinary people the power to create good-quality content which can then be spread, by the press, over multiple platforms. Citizens are almost always the first ones to arrive at a breaking news location and can provide the initial images of the scene. However, existing crowdsourced tools and platforms are predominantly centralized and are usually fed with unreliable and untrustworthy information.

This work introduces a Crowd Journalism ecosystem whose core is a video marketplace web tool based on an organization-level decentralized system that can store, visualize, rate, and execute transactions of live-made videos. Smart contracts ensure that all the transactions are transparent and secure.

This approach to Crowd Journalism exploits the inherent features of a blockchain such as offering trustful, anonymized, and immutable transactions, which has the potential to revolutionize the way news content is shared and commercially exploited.

CCS CONCEPTS

• **Computer systems organization** → **Peer-to-peer architectures**; • **Information systems** → *Crowdsourcing*.

KEYWORDS

crowdsourcing, crowd journalism, video marketplace, decentralized networks, blockchain

1 INTRODUCTION

In the era of journalism before the Internet (see Figure 1), the market worked differently from nowadays: there was a small number of companies regulating the journalistic content, most of them controlled by the government, and the readers could not choose the content they wanted to read. Instead, the press had the power to decide what was published, treating the readers as a mass — unable to choose the news according to personal interests. Also, the advertisers had in the press its primary channel to reach consumers, and consequently, a monopoly was well established.

As predicted by Nicholas Negroponte in 1995 [20], with the emergence of the Internet and social media, this market model eventually began to change (see Figure 2). Now the reader can choose what content to read, and the subscription method was increasingly abandoned since social media allows a market system where clicks and views generate revenues. The flexibility brought by the easy access to the Internet and social media platforms promoted the direct participation of readers in news creation, which is what Bowman and Willis call citizen journalism [4].

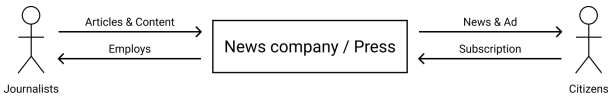

Figure 1: Journalism before Internet

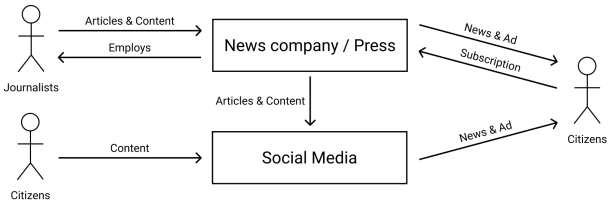

Figure 2: Journalism with Internet and Social Media

Citizen-based journalism, also called participatory and democratic journalism [21], is the act of a citizen, or group of citizens, playing an active role in the process of collecting, reporting, analyzing, and disseminating news and information [4]. It then evolved into more sophisticated mechanisms such as crowdsourcing or collaborative journalism [1, 26]. This collaboration in the news creation, often promoted by the journalistic organizations themselves, is facilitated by digital tools that allow specific tasks, such as knowledge-search or breaking news video coverage, traditionally performed by company employees, to be outsourced to the general public [18].

Some game-changing solutions, such as Civil [13], Steem [27], and Citizen [5], tried to revolutionize the way journalism interacts with its audience, by letting the latest contribute with meaningful content. Therefore, they define a paradigm in which the audience directly and personally relates to the content that is being created. These are solutions that address Crowd Journalism as a way to enrich the news creation process but they do not offer reliable and

transparent means to reward citizens and amateurs journalists for their contributions. Consequently, their adoption is not massive, and therefore news stories that reach the general public are not as rich as they could be if Crowd Journalism was widely used in a structured way.

Our main goal is to create an ecosystem for Crowd Journalism where citizens, amateurs journalists, and news organizations can co-collaborate in the news video creation process through a secure and transparent transaction system.

To this end, a decentralized network of organizations has to be created, composed by an independent and trustworthy leading entity (e.g. a news agency), that deploys the system, aggregating the content, and establishing an organization consortium. A permissioned blockchain seems to be the perfect fit to solve this issue since it provides an additional level of security over traditional blockchain systems and requires an access control layer. Besides, by using a permissioned blockchain, we automatically inherit all the features that characterize a blockchain, such as anonymity, transparency, immutability, traceability, among others, and which are crucial to ensure that citizens and news organizations will trust and use such a system.

The development of a video marketplace, which runs on top of this permissioned blockchain, enables transactions between crowd journalists video creators and potential buyers (e.g. media companies) belonging to the consortium. This marketplace includes features such as the storage of live videos created by crowd journalists, their pre-visualization, and a democratic crowdsourced content rating.

With this approach, we tackle the problem of trust and governance over user content, by designing a democratic system, with user anonymization and freedom of content creation. By ensuring user privacy and control over the content in a transparent fashion, while collaborating in the creation of delicate news, we can foster trust among its users.

To sum up, this work proposes an environment, based on a decentralized marketplace network (see Figure 3), that allows the general public and amateur journalists to share journalistic live-video content and be rewarded for it, while maintaining their privacy and security through a permissioned blockchain infrastructure. In this model, a news agency is responsible for aggregating the content created by citizens, which is then made available to media companies through decentralized transactions in the marketplace.

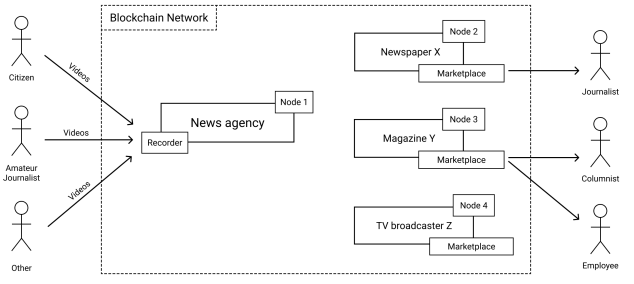

Figure 3: Decentralized system for crowdsourcing journalism

The rest of this paper is structured as follows. In Section 2 we discuss the state of the art of the different areas relevant to the problem. Section 3 explores the blockchain technology, emphasizing some of its concepts (blockchain, consensus algorithms, and smart contracts) and the technology adopted for our ecosystem: Hyperledger Fabric. Section 4 describes our approach to a Crowd Journalism decentralized ecosystem, referring to the components and functionalities that form the system. Finally, Section 5 introduces a validation prototype developed as an instantiation of a marketplace view for a given organization or media group.

2 RELATED WORK

The growth of the Internet and social media platforms has revolutionized the way people interact with each other. The ease of access to information and its subsequent sharing has turned the audience into an active part of the community.

In 2006, Jeff Howe [12] originally introduced the term crowdsourcing as "the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined network of people in the form of an open call." By this definition, crowdsourcing can be applied in a variety of contexts. As an example, reCAPTCHA (2008) [29] was created to improve the optic system in charge of digitizing old books, getting users to help recognize the words that are not correctly detected — a clear example of a crowdsourcing approach for a non-journalistic purpose.

In the journalism context, large news corporations started adopting crowdsourcing as a way to enhance their knowledge-search strategies. For example, in 2017, ProPublica explicitly asked their readers to support them in the creation of their news piece Lost Mothers [10], which seeks to explore the reasons why the US has the highest rate of deaths related to pregnancy and childbirth in the developed world. In the meanwhile, CNN, The New York Times, NBC, The Onion, and Time started using Twitter as a vehicle to collect information from its users, and for publishing and promoting their news. Twitter, Reddit and other platforms have been promptly adopted in newsrooms as a mechanism for creating news based on user-generated content [22].

Meanwhile, some platforms were created with the specific purpose of enhancing collaborative journalism. Google introduced the Google News Lab in 2015 [9], aiming to "collaborate with journalists and entrepreneurs to help build the future of media". The platform uses other Google tools in an environment where citizens, amateur and professional journalists can scout for real-time stories, updates, and breaking news, produce better reporting, and get data for insightful storytelling. In 2016, Crowdpondent [9] launched a web platform for citizen journalism where the readers rate articles and decide how valuable a piece can be to the website. The crowd does not only consume content; it creates, reads, and rates it to determine how much the author will earn for their piece.

Some crowd streaming solutions were also brought into the spotlight for revolutionizing some aspects of people's daily lives. In 2019, the crime app Citizen[5] had over a million active users, remaining constantly in the top news apps in the App Store. Citizen is a platform that encourages citizens to report crimes and police-related issues in their own city. The Citizen ecosystem is formed by a mobile app and a website. The mobile app can be used to record

or live-stream videos from incident situations, as well as to report new incidents or comment on existing ones. The website displays incidents in specific cities using a map-based approach.

According to Bowman [4], participatory journalism flourishes in social media. Social media has become a free-for-all information library that is updated 24 hours a day, by people from all around the world. The ordinary citizen can be at a news scene long before the journalists. Therefore, the content that citizens are able to gather in the form of text, images or videos has a very important role in the creation of the final news story. Social media platforms like Twitter, Facebook, and others have taken the concept to a global scale, making these networks a primary source for collaborative media production. Twitter was launched in 2006 and ended up becoming one of the primary levers that led to the current scenario of collaborative journalism.

Every day, a massive amount of content is produced on the Internet. For reference, over 500 million tweets are posted daily on Twitter alone [3], adding up to 12 TiB of data every day [7]. User-generated content creates billion-dollar revenues for social media companies like Facebook, Twitter, and Reddit. Despite the monetary contribution given to users who publish high-traffic content, most of the revenues associated with the content are directed to the social media platform owner. This circumstance, coupled with recurring data privacy and content ownership problems, has raised a call for decentralized social media approaches and peer-to-peer architectures based on blockchain to improve content governance and transparency while increasing its security.

As an example, Steem[27], created in 2016 by Ned Scott and Dan Larimer [24], is a social blockchain with application in various strands of social connectivity. Steem took the premise that its users are more active if they are directly rewarded for meaningful contributions. To verify and classify the content quality, every new post is assessed and voted by the community and rewarded accordingly. Moreover, the vote influence is also decentralized, as the more reputation a contributor has, the more influence its vote has on the reward given to other users.

A different application of decentralized social media is Po.et [23]. Officially established in 2017, Po.et is a shared ledger designed to record immutable and timestamped information about creative digital assets on a blockchain. Po.et blockchain uses Proof of Existence to verify the integrity of a document regarding its history, timestamps, copyright, and authorship. Proof of Existence uses cryptographic hashes to allow its users to upload digital documents and get them verified by the blockchain without disclosing the content itself [8].

Decentralized approaches using blockchain and distributed ledger technologies have also been considered for the development of journalistic-related platforms. In this context, news organizations' infrastructures could use a blockchain to securely store the timestamps of the publication date and provenance of news stories. Alternatively, there can be a solution around a network of stakeholder-driven organizations in a private blockchain, where news organizations — participants of the network — can access and manage the data stored in the blockchain, and also participate in the governance of the entire network, in an organization consortium [14].

One example of such a platform was the Civil blockchain, which was launched in 2016 and shut down in 2020. Civil [13] was a decentralized self-governing organization with a blockchain-based media platform implemented under its own cryptocurrency, the CVL. The platform, launched with the collaboration of Forbes, Denver Post alumni, and the Associated Press, worked as an open-market allowing the community to publish, assess, and trade the news by recording every single change on a news post, the information about its origin, modification, and time of publication.

Kim and Yoon [16] defined a blockchain for creating and storing news articles based on a participatory Proof of Sharing process with the curation of preapproved journalists, that defines which articles are breaking news or not. In order to fight the spreading of fake news, Shae and Tai [25] proposed a blockchain that uses Artificial Intelligence to rank news and provide incentives for citizens to work as news validators. Nevertheless, these approaches do not include a blockchain infrastructure that is organized in entities with different roles such as independent content aggregators (news agencies), content buyers (media groups), and a video marketplace, with views that can be customized for each of the buyers. Moreover, our approach ensures secure and transparent transactions between content owners and buyers without the intervention of third parties.

3 BACKGROUND ON BLOCKCHAIN TECHNOLOGIES

The blockchain concept was first introduced in 1991 by Stuart Haber and W. Scott Stornetta [11] but was popularized by Nakamoto, in 2008, with the cryptocurrency Bitcoin [19], used to solve the double-spending problem, which is the well-known problem of spending the same electronic money twice. However, the concept of blockchain makes sense on its own, outside the scope of a cryptocurrency. For example, blockchain technology is becoming one of the most promising technologies for the next generation of Internet interaction systems, such as smart contracts, public services, Internet of Things, reputation systems and security services.

Blockchains allow applications, that could previously run only through a trusted intermediary, to operate in a decentralized fashion, without the need for a central authority, and achieve the same functionality with the same amount of certainty.

3.1 Blockchain concept

A blockchain can be regarded as a public ledger, in which all committed transactions are stored in a chain of blocks that continuously grows when new blocks are appended to it. A single block is a set of information that is added to the system, with new transactions or new content. Each block is unique by having a hash value created from its content. The key point is that each block holds the hash of the previous block, linking them in a chain of blocks, and making the chain unbreakable [30]. The ledger is then kept by all participants guaranteeing the decentralized approach. However, it does not ensure that the system is trustful, in a way that it is not possible to avoid the injection of fraudulent blocks without a consensus method to decide which new blocks are valid and which ledger is the truthful one. For that, there are several consensus algorithms, with different pros and cons, to ensure that the system is inviolable and reliable.

3.1.1 Consensus Algorithms. Consensus algorithms can be divided into two categories: lottery-based and voting-based. Lottery-based algorithms are known for finding a way to choose a node that is responsible for adding the new block. The way the node that adds the block is chosen varies according to the algorithm.

For example, Proof of Work (PoW) is the lottery-based algorithm used by the Bitcoin blockchain network and it is based on the premise that a node must use computational power to prove that it has no malicious intent to tamper with the network. On the other hand, the Proof of Stake (PoS) stands out for its absence of complicated and expensive calculations. This method states that participants who have more digital assets of the cryptocurrency would be less likely to act maliciously on the network and so they are more likely to be chosen to forge the new block. Along with these two, there are many other lottery-based algorithms to define a network consensus (Proof of Authority (PoA), Proof of Elapsed Time (PoET), Proof of Burn (PoB), etc.).

However, the technology used in this work, Hyperledger Fabric, uses a different algorithm called Practical Byzantine Fault Tolerance (PBFT). Unlike the previous ones, PBFT is a voting-based algorithm, often used for permissioned blockchains, where the blockchain is aware of all the nodes that belong to the network. The consensus is achieved when there are at most m malicious nodes for a total of $3m + 1$ nodes. It is a less energy-consuming solution as it removes the unnecessary computations required to reach consensus on the network, improving the overall performance. Nevertheless, it is a less scalable solution due to the exponential complexity of the number of messages sent, generally used in systems with a limited and known number of nodes.

3.1.2 Smart Contracts. Another key feature of blockchain technology is that it enables smart contracts, which are computer programs that autonomously execute the terms of a contract when predefined conditions are met. The term "smart contract" was popularized in 1994 by Szabo [28], who gave an example of a smart contract that enforces car loan payments. Smart contracts consisting of transactions, such as the ones required for this work, are essentially stored, replicated and updated in distributed blockchains. This contrasts with conventional contracts that need a third party to be validated in a centralized manner. Zheng et al.[31] claim that smart contracts have three main advantages when compared with conventional contracts: reducing risks, cutting down administrative and service costs, and improving the efficiency of business processes, which is a crucial characteristic for the platform here proposed.

3.2 Hyperledger Fabric

The Hyperledger project, created by the Linux Foundation in 2015 and with a significant contribution from IBM [15], has been launching several tools, libraries, and frameworks to support blockchain-based software development. The most relevant framework for this work is Hyperledger Fabric [17]. Fabric is a modular and extensible open-source tool [2] to deploy and operate permissioned blockchains with the possibility to delineate functions between network nodes, run smart contracts, and configure authentication and consensus services.

The Hyperledger Fabric technology implements a *ledger*, which is a fundamental concept for storing the information within the system. This ledger stores relevant factual information about a set of *business objects*. A ledger, in this particular context, consists of two distinct, though related, parts – a *world state* and a *blockchain*.

The world state is a database that holds the current state of facts about all the business objects in the system. It eases the access and retrieval of the current value of data by removing the need to calculate it by traversing the entire transaction log. The world state changes every time an object's state is created, updated, or deleted.

The blockchain is the history of all transactions and state changes that were ever done in the system. The world state is the current result of all those operations. The blockchain data structure, unlike the world state, is immutable as it can only receive new transaction blocks and append them to the chain. The following example (see Figure 4) shows all the possibilities of a blockchain network implemented using Hyperledger Fabric. A network N, offering a service O, contains nodes (peers P1, P2, P3). There are client applications (A1, A2, A3) that have permissions to execute transactions on specific channels (C1, C2). Nodes usually contain the ledger (L1, L2) associated with the channels. Each channel has a set of policies (CP1, CP2) that are established by a consortium (RARB, RBRC) or association between organizations. Each association is served by a Certificate Authority (CA1, CA2, etc.).

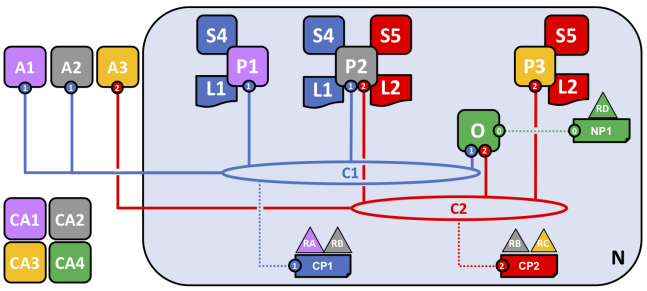

Figure 4: Hyperledger Fabric network example. Reprinted from https://hyperledger-fabric.readthedocs.io/en/release-1.2/network/network.html

Regarding Hyperledger Fabric technology, two relevant concepts are crucial to understand the workflow of the Fabric blockchain authentication process:

Identity — In a blockchain setting, every actor or node who wishes to interact with the network needs a digital identity issued by an authority trusted by the system. Each actor needs a digital identity encapsulated in an X.509 digital certificate [6]. These identities determine the exact permissions over resources and access to information that actors have in a blockchain network.

Certificate Authority — A Certificate Authority (CA) issues certificates based on Public Key Infrastructure (PKI) to network member organizations and their users. In this setting, one or more CA can be used to define the organization's members from a digital perspective. It is the CA that provides the basis for the organization's actors to have a verifiable digital identity. Certificate Authorities also have a

certificate, which they make widely available. This allows the consumers with identities issued by a given CA to verify them by checking that the certificate could only have been generated by the holder of the corresponding private key (the CA).

4 CROWD JOURNALISM DECENTRALIZED ECOSYSTEM

The main objective of this work is the creation of an ecosystem for Crowd Journalism where citizens, amateur journalists and news organizations can collaborate in the news video creation process through a secure and transparent transaction system. The approach we propose involves the creation of a video marketplace through an organizational-level decentralization running over a permissioned blockchain network. It means that, unlike cases such as Bitcoin or Ethereum-based applications, decentralization does not occur on the users.

In our model, as illustrated in Figure 5 (for two organizations), the blockchain network is composed of a set of peers, and each peer is a node in the network that actively participates in the consensus algorithm and form altogether the organization consortium. In this case, each peer represents a news agency or media group, which is independent of the blockchain infrastructure. The news agencies deploy and maintain this network, acting as content aggregators since they are responsible for collecting the live videos captured by the citizen journalists through a platform designed for video creation (as represented on the top left of Figure 5). They also act as independent entities, increasing the network reputation.

On the other hand, the news companies represent buyers for crowd content. Therefore, they need to have a marketplace platform available to complete the system's flow. Each organization can have its own instantiation of this marketplace platform (as it can be seen at the top right of Figure 5) in which its users — employees, journalists, columnists, and other interested individuals — can watch, rate, and purchase the videos using the blockchain. For this purpose, a Marketplace Module was designed which enables the creation of multiple customizable views of the marketplace outside the blockchain setting, one for each media group. Moreover, this module bridges the connection with the blockchain network, so media groups do not have to concern about the complexity of blockchain connection, authentication, and communication.

Also, there is a Certificate Authority (CA) associated with each peer, which is the entity that certifies a user identity and its permissions within the blockchain network. Ultimately, for each organization, there is a Blockchain Connection Provider (BC Provider) that has permissions to register new user identities associated with the respective organization and to provide the Marketplace Module with the blockchain connection details.

In the following subsections, we explain in more detail each of these components, as well as the business objects included in the smart contract that was defined to implement the Crowd Journalism ecosystem business rules.

4.1 Blockchain Connection Provider

The BC Provider is an auxiliary module whose main goal is to facilitate the communication between the Marketplace Module and the blockchain network, providing the former with the information necessary to access the latter, such as available peers, their locations (IP and port) and public keys. It is also equipped with administrative-level credentials, granted by the consortium, giving it the ability to generate new identities for its organization, whenever requested by the Marketplace Module. An instance of a BC Provider has to be deployed and managed by each organization, along with the CA.

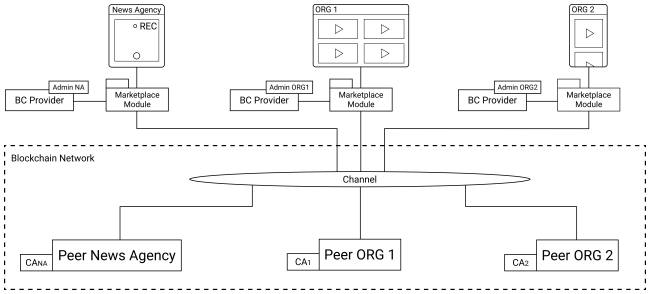

Figure 5: Video marketplace on a blockchain environment

4.2 Marketplace Module

This module is responsible for all the communications with the blockchain network, allowing each media group to easily develop and customize its marketplace platform.

Firstly, it ensures that the connection details with the blockchain network are always up-to-date. Secondly, it deals with the blockchain authentication protocol, allowing the media groups' platforms to register new user identities in the blockchain environment. Finally, it implements a set of operations that provide all the services that a marketplace platform may need.

4.2.1 Connection. The Marketplace Module often interacts with the blockchain environment to retrieve information or to perform new operations. For that, it needs to ensure that the connection details are always up-to-date, in what concerns the communication channel, active peers, CA addresses, etc. Therefore, it uses the BC Provider to get the correct connection details, and stores that information to be reused in further operations. Whenever the connection goes down, or some blockchain connection detail changes (new peers, updated addresses, etc.), the Marketplace Module calls the BC Provider again to refresh the connection information.

4.2.2 Authentication. The authentication process used in our approach is the one of a permissioned blockchain network, which is substantially different from the typical client-server architectures. This kind of system controls the authentication process via certificates. Each user adding transactions to the blockchain has a certificate — a public key, and a private key — which are used to validate the operation requests. In this blockchain environment, there is no login or username-password authentication; only a register phase where the user is created and enrolled, and the certificates are generated, returned, and preserved (Figure 6, steps 1 – 7) to be used later, on every performed operation (Figure 6, steps 8 – 9).

The authentication process starts with a user being created, which includes verifying if the new user id is not already registered. The Marketplace Module executes a request to the respective

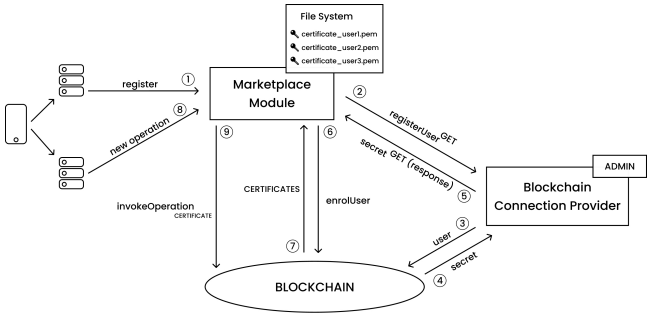

Figure 6: Authentication protocol

BC Provider, which addresses the registration request to the correspondent CA. Then, a secret is returned, and it is used by the Marketplace Module to enrol the user in the blockchain network. Finally, the new identity is created and its certificates are stored in the Marketplace Module's file system, ready to be used to perform a new operation.

4.2.3 Operations. Apart from the connection and authentication processes, all operations in the Marketplace Module are key actions to operate every aspect of a marketplace platform.

An operation is the invocation of a specific method within the blockchain smart contract, whether to perform a change in the system or to retrieve information from the world state. To better understand the main operations involved in this Crowd Journalism ecosystem, it is necessary to identify what are our business objects and explain the defined smart contract.

4.3 Business objects and the smart contract

Hyperledger Fabric uses transactions to store facts and state changes about a set of business objects, as mentioned in Section 3.2. These objects are the entities included in each smart contract transaction. In our approach, the created objects for the Crowd Journalism ecosystem are Video, Classification, Purchase, Event, and Tag (see the UML class diagram in Figure 7).

Video objects collect the information associated with a new live video, created by a user through the video recorder platform. A Video instance holds information such as the video id, the associated event, the related tags, price, and geolocation.

Classification objects hold information about the act of a user rating a video, and it contains the user id, the video id, and the rating value.

Purchase object is created when a user purchases a video, and it contains the transaction value and the respective video and user id.

Event object is an accident, natural disaster, conference, interview, debate, or any newsworthy incident of a journalistic nature (in the Crowd Journalism business model) that may be liable to have associated multimedia content. It contains the event id and the associated tags.

Tag object categorizes the different videos and events in the system.

Lastly, the **User** class is considered for conceptual design and diagram-completeness purposes. Hyperledger employs the user's information through its credentials and wallet, but it is not a blockchain business object.

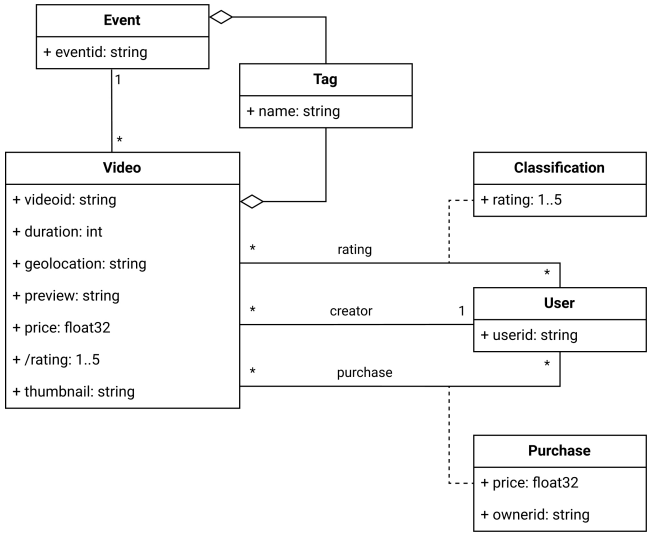

Figure 7: Blockchain data structure

As mentioned in Section 3.1.2, after identifying the business objects, it is necessary to define the smart contracts that implement the business rules. In our model, there is only one smart contract which has to be replicated in each network peer. Each transaction operation must be processed by every peer and validated before being added to the ledger.

In our ecosystem, every transaction has one of the following operations associated with it:

Event creation — when a news agency creates a breaking news event, a new Event object is created in the blockchain containing its name, associated tags, geolocation range, and event duration.

Start a video recording — when a crowd entity starts recording a live video, a new blockchain transaction is registered, creating a new Video object containing information about its ownership, geolocation, associated event, and tags.

Stop a video recording — when a video capture is finished, a set of media operations are performed to generate the video assets (such as thumbnails, watermarked version, and the original video asset) and the corresponding Video object in the blockchain is updated with this new information. Also, a price is defined based on the video duration and its resolution. After this process, the video is ready to be purchased in the marketplace platform.

Search videos — when a user wants to look for videos matching specific criteria, a query to the blockchain is made, and the world state is used to retrieve that specific subset of objects. The videos can be filtered by ids, events, tags, geolocation, and date.

Purchase a video — when a marketplace user wants to purchase a video, a new transaction request is made to the

blockchain where this user's balance is validated to ensure that he has enough credit to buy the video. If the transaction is authorized, the user will gain access to the video asset and a new Purchase object is registered into the blockchain ledger, indicating that a user purchased a video by a predefined price. This transaction goes through the consensus mechanism to guarantee its truthfulness.

Classify a video — when a user classifies a video in the marketplace, a transaction is submitted to the ledger with information about this rating. For each user ↔ video pair, there is at most a Classification object, which corresponds to the latest classification provided by that user.

Having such a blockchain infrastructure configured by a news agency, a media group can easily develop and customize its marketplace platform using the Marketplace Module and the BC Provider here described. In the next section, a validation prototype is presented and detailed.

5 VALIDATION PROTOTYPE

The validation prototype is a proof of concept implementation demonstrating an organization-side deployment of the Crowd Journalism ecosystem. In this section, we show the deployment architecture, the involved technologies, the application of the Marketplace Module and the instantiation of several marketplace features.

At the organization side, the developed prototype is a Web application consisting of the BC Provider and three deployment components (see Figure 8): a React UI client, where the user navigates through the platform; a Node API Server, with the business logic; and a MySQL database server, to store auxiliary data.

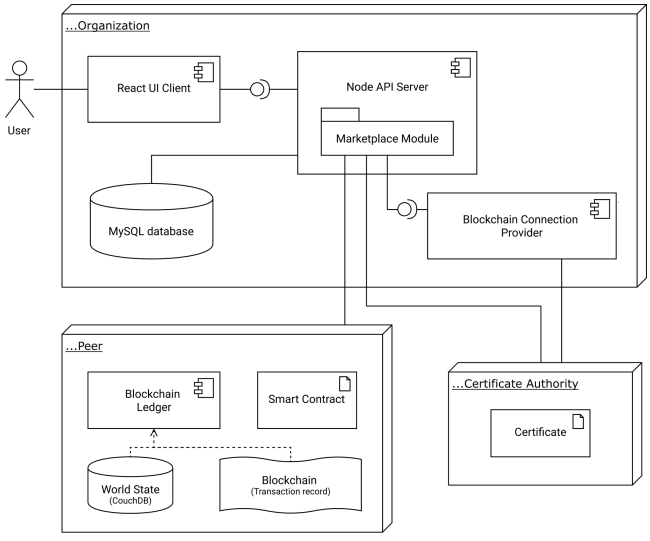

Figure 8: Component architecture diagram

This marketplace prototype has a deployment structure of a client-server system. The client is a React application, consisting of pages, components, layouts, and assets that together create the marketplace front-end that interacts with an Express NodeJS server via HTTPS. The server holds a Rest API with the necessary endpoints to fully operate a marketplace platform. Together with the server, there is a MySQL relational database to store user credential details, the shopping cart data, and any other auxiliary information. Also, there is an instance of a BC Provider to execute the authentication protocol and to provide the blockchain configuration data upon requested. Finally, the server imports the Marketplace Module to query and to operate the blockchain and to communicate with the BC Provider, in the authentication protocol.

The prototype implements the following features: user authentication, video listing, filtering, viewing, rating, and purchasing.

5.1 Authentication

The prototype's authentication is designed as a regular username-password verification. However, as explained in Section 4.2.2, the blockchain authentication is done through digital identities and certificates. To register a new user (both locally and in the blockchain), it is necessary to provide the authentication credentials which, in this case, are: the user's email, a username, and a password. In Figure 9, a UML diagram is presented which represents the steps followed in the registration process of a new user. As it can be seen, the registration process takes place in two phases. First, the Marketplace Module is invoked to create a new identity for that user: if there are no errors such as "duplicated user-id" or any connection issues, the user is created and registered in the blockchain system, and its certificates are saved in the organization's file system. Then, the user is created locally, that is, the user information and login credentials are stored in the local database.

In this way, only the username (taken as user id for the digital identity creation) is needed to register a user in the blockchain, meaning that the email and password are used only for the local authentication system. As a consequence, each organization is responsible for authenticating users with their credentials.

5.2 Video list and filter

The main section of the marketplace is the video gallery. The gallery provides a set of videos that have been posted by the users of the news agency's recorder platform, or by any other means that ingest content in the blockchain. Each video has relevant associated metadata such as thumbnails, the id of the video creator, the associated tags, the current rating, and the video duration. This marketplace section has also filtering mechanisms where videos can be grouped by *tag*, *event*, *geolocation*, *rating*, etc. The various retrieval operations are performed by the Marketplace Module which uses the certificates, stored in the module's file system, and invokes the smart contract to read the world state.

5.3 Video preview and purchase

Videos, in the video gallery, are provided with metadata that can be accessed by the user in another section of the marketplace, where there is also an adaptable and easy-to-use video player that displays a preview of the original video. This preview is generated from the blockchain smart contract when the video is created, with a watermark defined by the consortium, and it guarantees the integrity and flawlessness of the marketplace system, as the interested user can only access the original video after it is successfully purchased.

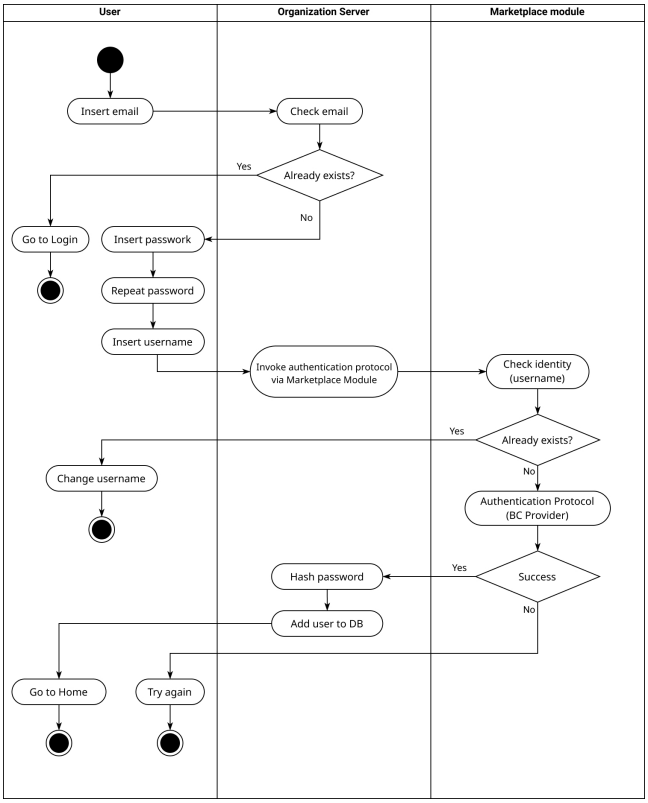

Figure 9: Activity diagram for register

To purchase a video, the user can add it to its cart and then proceed to check-out.

The operations to retrieve the video preview and original versions, and to purchase a video are defined in the Marketplace Module.

As explained in Section 4, the Marketplace Module and BC Provider define an abstraction layer over the blockchain that allows news organization to setup different online views of the available videos without the need to have an in-depth knowledge of the blockchain itself. In this section we detailed the necessary steps to create one of these individual views, namely through the usage of common web technologies and their interaction with the developed components.

6 CONCLUSIONS

Designing a new approach to Crowd Journalism using a blockchain infrastructure combines the advantages of crowdsourcing as a way to acquire knowledge and content, with the power of a decentralized blockchain network, which guarantees privacy, traceability, and content ownership. This work meets these two factors, while making the software adaptable and customizable, and the network decentralized in a media consortium.

In this paper, we presented a new approach to crowdsourced journalism based on a decentralized paradigm where citizens are rightfully rewarded for the content that they produce, through a video marketplace, and news organizations, with different permissions, form the blockchain consortium. In this model, news agencies act as content aggregators and as a liaison channel with crowd entities while media groups are potential buyers of the videos. We proposed and defined a marketplace module that act as a bridge between the blockchain and the different web applications of the organizations allowing the creation of customized views and instantiations of the marketplace. This will facilitate the development of online marketplaces for each of the participating organizations as it abstracts much of the blockchain inherent complexity. We presented how to establish the connections, the authentication process and other operations using the blockchain. Business objects and the corresponding smart contracts are also defined. Finally, a validation prototype was implemented and described, as a proof of concept of the proposed approach.

ACKNOWLEDGMENTS

This work was partially funded by the project ARTICONF which receives funding from the European Union's Horizon 2020 research and innovation program under grant agreement number 825134.

REFERENCES

[1] Tanja Aitamurto. 2016. Crowdsourcing as a Knowledge-Search Method in Digital Journalism: Ruptured ideals and blended responsibility. *Digital Journalism* (2016). https://doi.org/10.1080/21670811.2015.1034807

[2] Elli Androulaki, Yacov Manevich, Srinivasan Muralidharan, Chet Murthy, Binh Nguyen, Manish Sethi, Gari Singh, Keith Smith, Alessandro Sorniotti, Chrysoula Stathakopoulou, Marko Vukolić, Artem Barger, Sharon Weed Cocco, Jason Yellick, Vita Bortnikov, Christian Cachin, Konstantinos Christidis, Angelo De Caro, David Enyeart, Christopher Ferris, and Gennady Laventman. 2018. Hyperledger fabric. In *Proceedings of the Thirteenth EuroSys Conference on - EuroSys '18*. Association for Computing Machinery (ACM), 1–15. https://doi.org/10.1145/3190508.3190538

[3] Bogdan Batrinca and Philip C Treleaven. 2015. Social media analytics: a survey of techniques, tools and platforms. *Ai & Society* 30, 1 (2015), 89–116.

[4] By Shayne Bowman and Chris Willis. 2003. We Media: How audiences are shaping the future of news and information. *Media* 10, 03 (2003). https://doi.org/10.1177/1461444809336551

[5] Citizen. 2017 (accessed July 10, 2020). *Introducing: Citizen.* https://medium.com/citizen/introducing-citizen-a8d2f3fabf03.

[6] David Cooper, Stefan Santesson, Stephen Farrell, Sharon Boeyen, Russell Housley, W Timothy Polk, et al. 2008. Internet X. 509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile. *RFC* 5280 (2008), 1–151.

[7] Manoj Kumar Danthala and Siddhartha Ghosh. 2015. Bigdata Analysis: Streaming Twitter Data with Apache Hadoop and Visualizing using BigInsights. *International journal of engineering research and technology* 4 (2015).

[8] Josep Lluis De La Rosa, Victor Torres-Padrosa, Andrés El-Fakdi, Denisa Gibovic, O Hornyák, Lutz Maicher, and Francesc Miralles. 2017. A survey of blockchain technologies for open innovation. In *Proceedings of the 4th Annual World Open Innovation Conference*. 14–15.

[9] Dean Forrester. 2016. Crowdsourced journalism: a new democratic platform? *POLIS: journalism and society at the LSE* (2016).

[10] Adriana Gallardo. 2017. Do You Know Someone Who Died or Nearly Died in Childbirth? Help Us Investigate Maternal Health. https://www.propublica.org/getinvolved/help-propublica-and-npr-investigate-maternal-mortality

[11] Stuart Haber and W. Scott Stornetta. 1991. How to time-stamp a digital document. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 537 LNCS.

[12] Jeff Howe. 2006. The Rise of Crowdsourcing. *Wired Magazine* 14, 06 (2006). https://doi.org/10.1086/599595

[13] Matthew Iles. 2018 (accessed July 10, 2020). *The Civil White Paper.* https://civil.co/.

[14] Bernat Ivancsics. 2019. Blockchain in Journalism. (2019).

[15] Romeo Kienzler. 2016. Hyperledger – eine offene Blockchain Technologie. In *Blockchain Technology*. De Gruyter. https://doi.org/10.1515/9783110488951-005

[16] Byeowool Kim and Yong Yoon. 2018. Journalism Model Based on Blockchain with Sharing Space. *Symmetry* 11 (12 2018), 19. https://doi.org/10.3390/sym11010019

[17] Gabriela Motroc. 2019. Hyperledger Fabric 1.4 marks a very important milestone: First LTS release. https://jaxenter.com/hyperledger-fabric-1-4-first-lts-release-154332.html

[18] Karthika Muthukumaraswamy. 2010. WHEN THE MEDIA MEET CROWDS OF WISDOM. *Journalism Practice* 4, 1 (feb 2010), 48–65. https://doi.org/10.1080/17512780903068874

[19] Satoshi Nakamoto. 2008. *Bitcoin: A Peer-to-Peer Electronic Cash System.* Technical Report. Satoshi Nakamoto Institute. 1–9 pages. https://doi.org/10.1007/s10838-008-9062-0

[20] Nicholas Negroponte. 1996. *Being digital.* Vintage.

[21] Rabia Noor. 2016. Citizen Journalism vs. Mainstream Journalism: A Study on Challenges Posed by Amateurs. *Athens Journal of Mass Media and Communications* 3, 1 (2016). https://doi.org/10.30958/ajmmc.3.1.4

[22] Mats Nylund and Asha Gopalkrishnan. 2014. Crowdsourcing in Media. (2014).

[23] Esteban Ordano. 2017. *Po.et (White paper).* Technical Report. https://assets.website-files.com/5a0c978e0d22aa0001464356/5a7796662b07370001ace7a1{_}whitepaper.pdf

[24] Ned Scott and Dan Larimer. 2017. Steem: An incentivized, blockchain-based, public content platform. https://steem.com/SteemWhitePaper.pdf

[25] Z. Shae and J. Tsai. 2019. AI Blockchain Platform for Trusting News. In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. 1610–1619.

[26] Jochen Spangenberg. 2014. News from the Crowd : Grassroots and Collaborative 3 . GRASSROOTS JOURNALISM - A NEW. *Www* (2014), 765–768. https://doi.org/10.1145/2567948.2579327

[27] Steem. 2017 (accessed July 10, 2020). *An incentivized, blockchain-based, public content platform.* https://steem.com/SteemWhitePaper.pdf.

[28] Nick Szabo. 1994 (accessed July 10, 2020). *Smart Contracts.* http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart.contracts.html.

[29] Luis Von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum. 2008. recaptcha: Human-based character recognition via web security measures. *Science* 321, 5895 (2008), 1465–1468.

[30] Zibin Zheng, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang. 2017. An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. In *Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017*. Institute of Electrical and Electronics Engineers Inc., 557–564. https://doi.org/10.1109/BigDataCongress.2017.85

[31] Zibin Zheng, Shaoan Xie, Hong-Ning Dai, Weili Chen, Xiangping Chen, Jian Weng, and Muhammad Imran. 2020. An overview on smart contracts: Challenges, advances and platforms. *Future Generation Computer Systems* 105 (2020), 475 – 491. https://doi.org/10.1016/j.future.2019.12.019