

Gedecentraliseerd en transparant versiebeheer aan de hand van blockchain principes en IPFS: een praktische toepassing voor open source bedrijven.

Onderzoeksvoorstel Bachelorproef 2019-2020

Michiel Schoofs¹

Samenvatting

Versiebeheer is een inherent deel van het software ontwikkelingsproces. Toch wordt dit beheer vaak uitbesteed aan derde partijen waaronder Microsoft. Zeker voor de open source beweging is dit in strijd met hun ideologische principes.

Er bestaan manieren om zelf aan versiebeheer te doen die evenwel vaak omslachtig blijken en veel middelen vergen. In deze bachelorproef wordt door middel van Blockchain principes en IPFS aan versiebeheer gedaan zodat bedrijven met minimale overhead hun versie beheer kunnen decentraliseren.

Er wordt vertrokken vanuit verschillende onderzoekspapers om de context en verschillende aspecten van blockchain en decentralisatie uit te leggen. Vervolgens zullen deze principes worden toegepast op een concreet software project dat uitmondt in een werkend gedecentraliseerd versiebeheer systeem. De bedoeling is om bedrijven aan te zetten en te motiveren om meer open source te gebruiken in hun software ontwikkelingsproces.

Sleutelwoorden

Applicatieontwikkeling, blockchain — versiebeheer — decentralisatie — opensource

Co-promotor

Geen

Contact: ¹ michiel.schoofs@student.hogent.be;

Inhoudsopgave

1	Introductie	1
2	State-of-the-art	2
3	Methodologie	2
4	Verwachte resultaten	2
5	Verwachte conclusies	2
	Referenties	2

1. Introductie

De onderzoeksvraag waaruit wordt vertrokken luidt als volgt: "Hoe kan een bedrijf zijn versiebeheer decentraliseren aan de hand van blockchain principes en IPFS?" Binnen deze onderzoeksvraag zijn er drie begrippen:

- **Versiebeheer:** Git -een grote speler op het gebied van Versiebeheer- hanteert volgende definitie van het begrip "Versiebeheer is het systeem waarin veranderingen in een bestand of groep van bestanden over de tijd wordt bijgehouden, zodat je later specifieke versies kan opvragen." Deze definitie werd gepubliceerd in het boek *Pro git* (Chacon & Straub, 2014).

- **IPFS:** Interplanetary File System of (IPFS) werd in de paper "IPFS - Content Addressed, Versioned, P2P File System" geïntroduceerd door Benet (2014). Hierin stelde hij zijn technologie voor als een peer-to-peer gedistribueerd bestandssysteem waarin alle computers gaan werken met dezelfde bestandsindeling. Hiermee wordt bedoeld dat bestanden kunnen worden opgedeeld in verschillende delen en vervolgens worden opgeslagen op verschillende computers op een gezamenlijk netwerk. Vervolgens kunnen bestanden worden opgevraagd door middel van dit gezamenlijk netwerk te gaan aanspreken. Er is dus geen gecentraliseerd aanspreekpunt. IPFS leent zich ook tot versiebeheer.

- **Blockchain:** Blockchain is een manier om data op te slaan aan de hand van blokken. Deze blokken bevatten verschillende gegevens. Deze gegevens worden omgezet door middel van wiskundige functies die ook wel hashfuncties worden genoemd. Door de blokken onderling aan elkaar te koppelen en wiskundige functies te gebruiken bij het verifiëren van de integriteit van de blokken ontstaat er een veilige en volledig gedecentraliseerde manier van dataopslag.

Versiebeheer wordt vaak uitbesteed aan derden of zelf gedaan aan de hand van een centrale server. Het nadeel hiervan is dat er één centrale plek is waar het kan mislopen. Stel bijvoorbeeld dat de centrale server gegevens verliest is men alles kwijt. Een ander probleem is dat er binnen versiebeheer een aantal monopolies ontstaan waaronder Microsoft die GitHub kocht in 2018. Dit staat haaks op de open source beweging die streeft naar een transparante en democratische manier van software ontwikkeling. Door het introduceren van de bovengenoemde technologieën kunnen zowel de bestanden als de nodige informatie voor versiebeheer worden verspreid waardoor er geen centraal punt is en er ook geen commercieel bedrijf bij betrokken is.

2. State-of-the-art

De manier van werken is gebaseerd op het artikel “Decentralized document version control using ethereum blockchain and IPFS.” (Nizamuddin e.a., 2019) In het onderzoek wordt er gebruik gemaakt van smart contracts. Dit is in essentie code die zal uitgevoerd worden als aan bepaalde voorwaarden wordt voldaan. Deze smart contracts worden gebruikt om de verschillende aspecten van versiebeheer en data vast te leggen en uit te voeren. Voor de bestanden binnen het project wordt gekozen voor IPFS om op een gedecentraliseerde manier deze te kunnen opslaan.

De paper vormt een zeer goede aanzet en ook de werkmethode is uitvoerig beschreven. Toch blijft het zeer abstract. Belangrijke aspecten van versiebeheer worden kort of niet aangehaald waaronder “cloning”, “merging” of “branching”. In de bovengenoemde paper wordt een sterke focus op Ethereum gelegd, ontwikkeling bovenop deze blockchain interpretatie brengt echter significante overhead met zich mee. Zo is de snelheid van het systeem afhankelijk van de capaciteit en belasting van het netwerk op het gegeven moment.

Deze bachelorproef legt de focus op het ontwikkelen van een concrete toepassing. Ook de meer complexe en technische problemen zullen worden behandeld. De algemene principes van blockchain zullen vrijer worden geïmplementeerd en op een lokaal netwerk van enkele computer worden verspreid. In plaats van een grotere architectuur en implementatie te gebruiken

3. Methodologie

Er wordt vertrokken vanuit een literatuurstudie om de verschillende elementen van versiebeheer en de reeds bestaande technologieën te verkennen. Vervolgens komen de aspecten van blockchain en IPFS aan bod door middel van een demo waarin wordt gebruik gemaakt van een Word document met verschillende versies.

Tot slot worden de verschillende aspecten van versiebeheer aan de hand van een demo-applicatie geïllustreerd. Hiervoor zijn er drie hypothetische gebruikers: Mark, Alice en Bob die samen een T-shirt webshop ontwikkelen. Ze zullen

hiervoor gebruiken maken van ASP.Net en Visual Studio. Binnen hun ontwikkelingsproces zullen ze een aantal gekende problemen tegenkomen waaronder “merge conflicten” en verschillende “branches”.

Bij elk van die problemen wordt er gekeken naar hoe Git -een klassiek versiebeheer systeem- dit oplost en hoe er een oplossing kan voorzien worden vanuit de voorgestelde gedistribueerde blockchain benadering. Voor het opstellen van de blockchain wordt gebruik gemaakt van C# en eventueel EthereumJS. De bedoeling is om op het einde van de bachelorproef tot een werkend prototype te komen dat gebruikt kan worden voor verschillende doeleinden.

4. Verwachte resultaten

Het eindresultaat van de Bachelorproef is om op een onderbouwde manier een prototype aan te reiken om op gedecentraliseerde wijze aan versie beheer te gaan doen. De voorgestelde werkwijze wordt grondig vergeleken met Git op de volgende twee punten:

- Snelheid van een transactie: hoe lang duurt het om bewerkingen zoals pull requests en branching toe te passen op een project en/of branch?
- Performantie qua geheugengebruik: hoe efficiënt wordt er binnen de algoritmen van de oplossing omgesprongen met geheugengebruik? Hiermee wordt zowel het extern geheugen (Hardschijf, SSD) als het werkgeheugen bedoelt.

De verwachting is dat de implementatiesnelheid lager zal zijn dan met de klassieke Git-systemen, omdat blockchain van nature vrij omslachtig en intensief is. De performantie van het geheugensysteem zal eveneens slechter scoren ten opzichte van Git. De blockchain implementatie waarborgt echter een hogere mate van data integriteit .

5. Verwachte conclusies

Gedecentraliseerd versiebeheer door middel van Blockchain en IPFS is zeker technisch mogelijk. De voordelen die het biedt zijn niet alleen zuiver ideologisch. De inherente garantie op data integriteit en het weghalen van een centraal “point of failure” is interessant voor grote bedrijven met een groot aantal aan verschillende projecten. Er zijn echter een aantal nadelen verbonden waaronder de omslachtige procedure en het intensief gebruik van computertechnische middelen. Dit maakt het voor kleine bedrijven minder interessant.

Referenties

- Benet, J. (2014). IPFS - Content Addressed, Versioned, P2P File System.
- Chacon, S. & Straub, B. (2014). *Pro Git* (2nd). Berkely, CA, USA, Apress.

Nizamuddin, N., Salah, K., Azad, M. A., Arshad, J. & Rehman, M. (2019). Decentralized document version control using ethereum blockchain and IPFS. *Computers & Electrical Engineering*, 76, 183–197. <https://doi.org/10.1016/j.compeleceng.2019.03.014>