

Sistema de Reputação Baseado em Blockchain para Detecção de Fake News*

Bruno H. Lippert¹, Bruno F. Scheltzke², Roben C. Lunardi^{1,2}, Regio A. Michelin^{1,2},
Avelino F. Zorzo¹

¹Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)
Porto Alegre, Brasil

²Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS)
Porto Alegre, Brasil.

Abstract. *Currently, fake news have become a big challenge for the society. Consequently, different organizations are searching for solutions to reduce the impact of fake news. Therefore, this paper advances the solution for this problem proposing a system, based on blockchain and user reputation, that registers fake news in a distributed ledger. The architecture and implementation of the system, called “No Fake News”, is still under development. This paper presents the proposed solution and a discussion about its adoption.*

Resumo. *Atualmente, a identificação de notícias falsas se tornou um grande desafio para a sociedade. Desta forma, diferentes organizações buscam por soluções para reduzir o impacto das ditas fake news. Portanto, este trabalho avança sobre este problema, propondo um sistema, baseado em blockchain e reputação do usuário, que registra notícias falsas em uma blockchain. A arquitetura e implementação do sistema, chamado “No Fake News”, ainda está em desenvolvimento. Este artigo apresenta a solução proposta e discussões sobre sua adoção.*

1. Introdução

O termo *fake news*, ou, “notícia falsa” em português, trata de uma notícia que passa uma informação não verídica intencionalmente com o objetivo de criar uma polêmica e/ou denegrir pessoas ou organizações [Shu et al. 2018]. *Fake news* tornou-se um assunto recorrente na mídia, dado que, com o avanço da tecnologia da informação qualquer pessoa com acesso a Internet pode disseminar informações falsas e maliciosas. Sendo assim, nenhuma notícia de fonte pouco conhecida pode ser considerada verdadeira até que seja confirmada sua veracidade [Shu et al. 2018]. Uma notícia falsa possui vários indicadores de sua falsidade, tais como: ausência de fonte, sensacionalismo, erros de formatação ou ortografia. Contudo, usualmente, o trabalho de busca de veracidade é realizado por pessoas especializadas, demandando tempo considerável para ser finalizado [Krishnan and Chen 2018].

Por ter um potencial de atingir milhões de usuários em pouco tempo, *fake news* tornou-se uma das principais armas políticas, através de manchetes falsas e tendenciosas. Por exemplo, essas notícias podem influenciar diretamente no resultado de uma eleição

*O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal Nível Superior – Brasil (CAPES) – Código de Financiamento 001, pelo INCT em Ciências Forenses financiado pelo CNPq, pela HP Brasil usando incentivos da Lei de Informática, e também da Totvs S/A.

[Hyman 2017]. Desta forma, *fake news* tornou-se uma ferramenta de difamação que possui potencial para influenciar usuários *online*.

Após o surgimento do Bitcoin, diversas propostas de *blockchain* surgiram para garantir a integridade, autenticidade e disponibilidade da informação em diferentes contextos. *Blockchains* armazenam a informação em blocos ligados de forma encadeada utilizando o *hash* do bloco anterior. Este conjunto de blocos encadeados é distribuído e validado pelos demais nodos. Desta forma, alterações na *blockchain* são dificultadas devido as características da mesma [Tschorsch and Scheuermann 2016]. Sendo assim, caso um destes nodos tente inserir uma informação que não condiz com um certo conjunto de regras predefinidas ou tente alterar um dado que foi armazenado, os valores *hash* armazenados em cada bloco não serão satisfeitos, e o bloco não será aceito [Orman 2018].

Tendo em vista os problemas gerados por informações falsas e a proposta que a tecnologia *blockchain* oferece, este artigo propõe uma solução para determinar a veracidade de notícias a partir de avaliações de usuários cadastrados. Com o objetivo de fornecer integridade e auditabilidade das informações, o sistema desenvolvido armazena opiniões sobre notícias utilizando a plataforma Speedy-Chain [Lunardi et al. 2018] [Michelin et al. 2018]. Quando a veracidade de uma notícia é requisitada, o sistema calcula a reputação dos usuários que votaram na notícia em questão, para assim determinar sua veracidade. Consequentemente, o sistema se propõe a alertar usuários de portais de notícia sobre a veracidade da informação. Portanto, diminuindo consideravelmente o impacto que *fake news* têm sobre o cotidiano das pessoas.

2. Trabalhos Relacionados

A área de verificação de notícias tem recebido atenção da comunidade científica nos últimos anos [Hyman 2017] [Buntain and Golbeck 2017] [Krishnan and Chen 2018] [Shu et al. 2018]. Desta forma, sistemas para identificação de informações falsas foram propostos [Benevenuto 2018] [Chaves 2018].

Buntain e Golbeck [Buntain and Golbeck 2017] desenvolveram um método para identificar notícias falsas na rede social Twitter. Os autores propõem um algoritmo que pesquisa, em dois bancos de dados, assuntos que potencialmente podem ser usados para criar uma *fake news*. A partir disto, o algoritmo analisa a estrutura das informações inseridas na rede social, com base em diversos critérios pré-definidos, para concluir se a informação é verdadeira ou não. Este mecanismo tem uma taxa de sucesso na identificação de *fake news* de aproximadamente 69%. Apesar de um índice elevado para a identificação automatizada de *fake news*, a solução proposta está limitado a plataforma Twitter.

Eleição Sem Fake [Benevenuto 2018] é um projeto capaz de monitorar notícias no Facebook e identificar “bots” nas redes sociais. Apesar de ser uma ferramenta para auditar dados, ela não permite que o usuário decida quais notícias pretende analisar e não permite a análise de notícias de temas variados. Como a ferramenta possui um enfoque no cenário político, o sistema se limita a dados demográficos referentes a distribuição do público que os portais de notícias possuem, assim como páginas de Facebook.

O Detector de Fake News [Chaves 2018] é um software onde os usuários podem informar se determinada notícia é uma *fake news* ou “click bait” (notícia com título

tendencioso para que o usuário clique no *link* da notícia). Apesar de ser uma proposta promissora, analisa apenas notícias do Twitter e Facebook, além de não permitir que o usuário realize buscas por notícias específicas na base de dados.

3. Solução Conceitual

Buscando a garantia de resiliência, disponibilidade, não repúdio e integridade dos dados a proposta do sistema No Fake News utiliza a tecnologia *blockchain*, dado que esta garante as propriedades necessárias ao sistema. Além destas propriedades, é possível que qualquer usuário audite os dados disponíveis. Esta proposta, além de possibilitar o envio de notícias para avaliação, permite que o usuário busque - nos dados armazenados na *blockchain* - notícias específicas. Desta forma, calcula-se a reputação dos usuários que interagem com o sistema e os atributos que possibilitam definir um índice de confiabilidade de uma notícia (ver Seção 4). Portanto, esta proposta torna-se, também, uma base de dados distribuída para pesquisa de *fake news* com garantia de integridade dos votos.

De forma a prover os atributos necessários ao sistema No Fake News, adotou-se uma arquitetura modular baseado no modelo REST para realizar a comunicação entre o módulo da SpeedyChain (*blockchain*), módulo para representação da API, e Front-End para interação com usuário. Os atributos relacionados a veracidade de uma notícia são calculados em tempo real a cada interação do usuário. A Figura 1 apresenta uma definição estrutural da proposta, bem como os principais módulos desenvolvidos para o projeto.

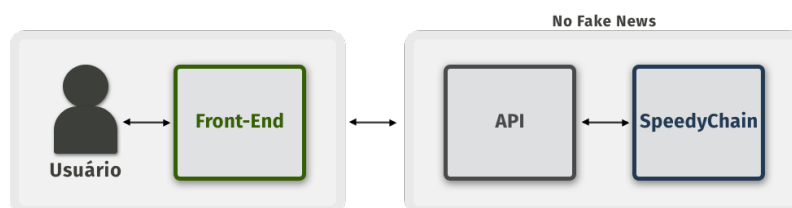


Figura 1. Arquitetura do No Fake News.

No sistema proposto, o usuário pode realizar operações de voto em uma notícia específica, e verificar atributos de confiabilidade de notícias através do componente de Front-End. Através dos votos de usuários, a API pode, enfim, calcular a veracidade das notícias (ver Seção 4). Além disso, o módulo de Front-End pode futuramente ser responsável por recolher o cadastro de notícias por parte da mídia e apresentar a reputação dos usuários, portais e autores.

A API é responsável por expor métodos que fornecem as funcionalidades para o Front-End. Este módulo do sistema provê a possibilidade de registrar a opinião sobre uma notícia, e encaminhar para inserção na *blockchain*. Além disso, a API tem função de receber o pedido e conferir a veracidade de uma determinada notícia. Desse modo, cabe a esse módulo coletar as informações necessárias diretamente da SpeedyChain. A veracidade de uma notícia é definida pela reputação dos usuários que opinaram sobre ela. A API solicita informações desses usuários para a SpeedyChain, que entrega todas as chaves públicas de identificação desses usuários.

A reputação de um usuário é definida pelas suas opiniões em notícias. A API coleta da SpeedyChain as notícias que o usuário votou, e, ao passo que existam votos su-

ficientes para determinar atributos de veracidade dessas notícias, o sistema pode comparar esses resultados com a opinião do usuário e assim determinar sua reputação.

Após ter todas as informações necessárias, a API é capaz de calcular as reputações dos usuários que opinaram na notícia requerida, e assim determinar sua veracidade para o usuário que originou a solicitação.

4. Implementação

Tendo como base a definição conceitual do sistema apresentada na Seção 3, a sua implementação foi feita utilizando tecnologias atuais. Dentre estas tecnologias buscou-se uma opção de implementação de blockchain que fosse capaz de atender as necessidades do sistema, e dadas as opções disponíveis SpeedyChain mostrou-se uma escolha adequada, visto que a mesma consiste de uma *blockchain* composta de blocos que podem continuar recebendo transações após serem inseridos na *blockchain* [Lunardi et al. 2018, Michelin et al. 2018]. Desta forma, todos o votos passados e os futuros votos de um usuários são armazenados em um mesmo bloco. Cada novo voto (transação) é inserido mantendo um *link* com os demais através da função *hash* [Michelin et al. 2018].

Para um usuário autenticar no sistema, primeiro este deve acessar o módulo de Front-End (conforme Figura 1). Após a confirmação de que a chave pública está registrada na SpeedyChain (caso não esteja, é preciso criar um novo usuário), trocam-se as chaves entre o sistema e o usuário para realizar a criptografia dos dados comunicados. Desta forma, estabelece-se a comunicação inicial com o sistema. A cada interação de inserção do usuário, a SpeedyChain requer que o Front-End seja capaz de assinar a informação com sua chave privada e cifra-la com sua chave de comunicação para que haja um maior nível de confiabilidade no sistema. Caso um usuário tente inserir algum voto e sua comunicação não foi estabelecida, ou há erro na criptografia ou na assinatura dos dados, nenhum dado é enviado para outros nodos da *blockchain* para sua inserção e uma mensagem de erro é retornada para o Front-End.

Para dar suporte para o usuário determinar se a notícia é ou não verdadeira, o sistema No Fake News (no módulo API da Figura 1) possui um conjunto de indicadores: voto de um usuário para a notícia sendo verdadeira (*VT*); voto de um usuário para a notícia sendo falsa (*VF*); a reputação do usuário votante na notícia (*R*); índice de veracidade da notícia (*T*); índice de falsidade da notícia (*F*); diferença de índice de veracidade e falsidade (*Df*); média da reputação dos votantes como verdadeiro (*MT*); média da reputação dos votantes como falso (*MF*).

O usuário pode incrementar a reputação (*R*) inicial no sistema de acordo com sua atividade no sistema. Um usuário nunca terá uma reputação mais baixa do que a inicial. O contrário pode acarretar em usuários maliciosos com baixa reputação criando novas contas no sistema, conforme [Zacharia et al. 2000].

Para cálculo dos índices *T* e *F*, o No Fake News separa os usuários que interagiram com a notícia pelas suas opiniões (verdadeiro/falso), e agrega o valor do peso de suas reputações: $T = \sum_{i=1}^n (VT_i * RT_i)$ e $F = \sum_{j=1}^m (VF_j * RF_j)$. Conforme apresentado nas fórmulas para *T* e *F*, o usuário pode verificar os parâmetros de veracidade e falsidade alcançados pela notícia. Quanto maior o índice, maior o peso para ser considerado.

O índice Df é o índice de confiança no resultado com maior peso (entre T e F), onde: $Df = \left| \frac{(T - F)}{(T + F)} \right|$. Desta forma, considera-se o percentual do peso de veracidade ou falsidade (o que for maior) da diferença de pesos em relação ao total opinado.

Finalmente, o sistema determina a qualidade da análise. Para isto o sistema calcula a média das reputações dos usuários que opinaram: $MT = \frac{\sum_{i=1}^n RT_i}{n}$ e $MF = \frac{\sum_{j=1}^m RF_j}{m}$. Portanto, uma notícia com muitos votos de baixa reputação tem um peso menor na análise, mesmo que tenha um número significativo de votos.

Atualmente foram implementadas duas versões do *Front-End*: uma versão como um *plugin* para o Google Chrome; e uma versão como aplicativo para iPhone. Ambas realizam a interação do usuário com o sistema. O sistema atual fornece parâmetros para facilitar a decisão por um usuário, consultando uma base resiliente e confiável. Todavia, a indicação automatizada se uma notícia é falsa (ou não) ainda está em desenvolvimento.

5. Comparação com soluções existentes

O No Fake News possibilita a validação de qualquer notícia disponível na Internet, independente da plataforma publicada. O que difere dos sistemas comparados, que utilizam informações do Facebook ou Twitter conforme apresentado na Tabela 1.

Sistema	Público alvo	Origem da notícia	Método de validação de notícias	Pesquisa por uma notícia
No Fake News	Qualquer usuário	Qualquer website	Reputação dos usuários que opinaram	Disponível a qualquer pessoa
Eleição Sem Fake	Facebook	Facebook	Monitoramento de notícias anunciadas	Notícias políticas da região do usuário
Detector de Fake News	Facebook ou Twitter	Facebook ou Twitter	Índices de veracidade baseado em reputação	Não disponível

Tabela 1. Comparação de soluções existentes com o No Fake News

Na Tabela 1 pode-se observar que o algoritmo de validação das notícias do sistema proposto, gerará resultados automatizados (atualmente gera apenas índices) sobre a veracidade de uma notícia. O No Fake News é baseado na reputação dos votantes, diminuindo assim custo e tempo com a utilização de agências de checagem de fatos.

O sistema proposto apresenta transparência, descentralização e integridade das informações. Os índices sobre veracidade das notícias podem ser acessados por qualquer pessoa ou sistema, sem requerer cadastro, diferente das outras soluções (Tabela 1).

6. Conclusões e Trabalhos Futuros

Na solução proposta, a veracidade de uma notícia é dependente da reputação dos usuários que votaram, e suas reputações são da mesma forma dependentes do resultado das notícias que votaram. Devido ao fato de que a reputação de um usuário pode variar, o sistema não pode dar um veredito final sobre uma notícia. Desse modo, adotou-se uma abordagem de cálculo de reputação e veracidade em tempo real mesmo com o eminente detrimento

da performance do sistema. Sobrecarga de requisições e tempo de espera podem ocorrer dependendo do tamanho do número de notícias e usuários do sistema. Portanto, questões de escalabilidade devem ser discutidas em trabalhos futuros.

A estrutura do sistema de reputação e os diferentes atributos calculados para a veracidade de uma notícia trazem algumas precauções contra possíveis usuários maliciosos que possam votar em uma certa opinião em massa [Zacharia et al. 2000]. Entretanto, está em análise inserções manuais de veracidade ou falsidade de notícias por órgãos confiáveis. Ainda pretende-se desenvolver um modelo de cache temporário de reputação no usuário. Dessa forma, o tempo de resposta do sistema pode ter melhor desempenho. No sistema proposto neste artigo, o módulo da API calcula atributos de uma notícia e encaminha aos usuários que podem tomar uma decisão sobre a veracidade da notícia. Pretende-se, como trabalho futuro, desenvolver uma solução automatizada para calcular o índice de confiabilidade, baseado nos índices apresentados na Seção 4.

Referências

- Benevenuto, F. (2018). Eleições sem fake. <http://www.eleicoes-sem-fake.dcc.ufmg.br/>. Acesso em: 25 set. 2018.
- Buntain, C. and Golbeck, J. (2017). Automatically identifying fake news in popular twitter threads. In *2017 IEEE International Conference on Smart Cloud (SmartCloud)*, pages 208–215.
- Chaves, R. (2018). Detector de fake news. <https://fakenewsdetector.org/pt>. Acesso em: 25 set. 2018.
- Hyman, J. (2017). Addressing fake news: Open standards amp; easy identification. In *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, pages 63–69.
- Krishnan, S. and Chen, M. (2018). Identifying tweets with fake news. In *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, pages 460–464.
- Lunardi, R. C., Michelin, R. A., Neu, C. V., and Zorzo, A. F. (2018). Distributed access control on iot ledger-based architecture. In *16th IEEE/IFIP Network Operations and Management Symposium (NOMS 2018)*, pages 1–8, Taipei, Taiwan.
- Michelin, R. A., Dorri, A., Lunardi, R. C., Steger, M., Kanhere, S. S., Jurdak, R., and Zorzo, A. F. (2018). SpeedyChain: A framework for decoupling data from blockchain for smart cities. *ArXiv e-prints*.
- Orman, H. (2018). Blockchain: the emperors new pki? *IEEE Internet Computing*, 22(2):23–28.
- Shu, K., Wang, S., and Liu, H. (2018). Understanding user profiles on social media for fake news detection. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 430–435.
- Tschorsch, F. and Scheuermann, B. (2016). Bitcoin and beyond: A technical survey on decentralized digital currencies. *IEEE Communications Surveys Tutorials*, 18(3):2084–2123.
- Zacharia, G., Moukas, A., and Maes, P. (2000). Collaborative reputation mechanisms for electronic marketplaces. *Decision Support Systems*, 29(4):371 – 388.