Security and Privacy Preservation in Mobile Advertising

by

Dongxiao Liu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Vojislav B. Mišić Professor, Ryerson University
Supervisor	Xuemin (Sherman) Shen University Professor, University of Waterloo
Internal Member	Xiaodong Lin Adjunct Associate Professor, University of Waterloo
Internal Member	Mark Crowley Assistant Professor, University of Waterloo
Internal–external Member	Jun Liu Associate Professor, University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Mobile advertising is emerging as a promising advertising strategy, which leverages prescriptive analytics, location-based distribution, and feedback-driven marketing to engage consumers with timely and targeted advertisements. In the current mobile advertising system, a third-party ad broker collects and manages advertisements for merchants who would like to promote their business to mobile users. Based on its large-scale database of user profiles, the ad broker can help the merchants to better reach out to customers with related interests and charges the merchants for ad dissemination services. Recently, mobile advertising technology has dominated the digital advertising industry and has become the main source of income for IT giants. However, there are many security and privacy challenges that may hinder the continuous success of the mobile advertising industry. First, there is a lack of advertising transparency in the current mobile advertising system. For example, mobile users are concerned about the reliability and trustworthiness of the ad dissemination process and advertising review system. Without proper countermeasures, mobile users can install ad-blocking software to filter out irrelevant or even misleading advertisements, which may lower the advertising investments from merchants. Second, as more strict privacy regulations (e.g. European General Data Privacy Regulations) take effect, it is critical to protect mobile users' personal profiles from illegal sharing and exposure in the mobile advertising system.

In this thesis, three security and privacy challenges for the mobile advertising system are identified and addressed with the designs, implementations, and evaluations of a blockchain-based architecture. First, we study the anonymous review system for the mobile advertising industry. When receiving advertisements from a specific merchant (e.g. a nearby restaurant), mobile users are more likely to browse the previous reviews about the merchant for quality-of-service assessments. However, current review systems are known for the lack of system transparency and are subject to many attacks, such as double reviews and deletions of negative reviews. We exploit the tamper-proof nature and the distributed consensus mechanism of the blockchain technology, to design a blockchain-based review system for mobile advertising, where review accumulations are transparent and verifiable to the public. To preserve user review privacy, we further design an anonymous review token generation scheme, where users are encouraged to leave reviews anonymously while still ensuring the review authenticity. We also explore the implementation challenges of the blockchain-based system on an Ethereum testing network and the experimental results demonstrate the application feasibility of the proposed anonymous review system. Second, we investigate the transparency issues for the targeted ad dissemination process. Specifically, we focus on a specific mobile advertising application: vehicular local advertising,

where vehicular users send spatial-keyword queries to ad brokers to receive location-aware advertisements. To build a transparent advertising system, the ad brokers are required to provide mobile users with explanations on the ad dissemination process, e.g., why a specific ad is disseminated to a mobile user. However, such transparency explanations are often found incomplete and sometimes even misleading, which may lower the user trust on the advertising system if without proper countermeasures. Therefore, we design an advertising smart contract to efficiently realize a publicly verifiable spatial-keyword query scheme. Instead of directly implementing the spatial-keyword query scheme on the smart contract with prohibitive storage and computation cost, we exploit the on/off-chain computation models to trade the expensive on-chain cost for cheap off-chain cost. With two design strategies: digest-and-verify and divide-then-assemble, the on-chain cost for a single spatial keyword query is reduced to constant regardless of the scale of the spatial-keyword database. Extensive experiments are conducted to provide both on-chain and off-chain benchmarks with a verifiable computation framework. Third, we explore another critical requirement of the mobile advertising system: public accountability enforcement against advertising misconducts, if (1) mobile users receive irrelevant ads, or (2) advertising policies of merchants are not correctly computed in the ad dissemination process. This requires the design of a composite Succinct Non-interactive ARGument (SNARG) system, that can be tailored for different advertising transparency requirements and is efficient for the blockchain implementations. Moreover, pursuing public accountability should also achieve a strict privacy guarantee for the user profile. We also propose an accountability contract which can receive explanation requirements from both mobile users and merchants. To promote prompt on-chain responses, we design an incentive mechanism based on the pre-deposits of involved parties, i.e., ad brokers, mobile users, and merchants. If any advertising misconduct is identified, public accountability can be enforced by confiscating the pre-deposits of the misbehaving party. Comprehensive experiments and analyses are conducted to demonstrate the versatile functionalities and feasibility of the accountability contract.

In summary, we have designed, implemented, and evaluated a blockchain-based architecture for security and privacy preservations in the mobile advertising. The designed architecture can not only enhance the transparency and accountability for the mobile advertising system, but has also achieved notably on-chain efficiency and privacy for real-world implementations. The results from the thesis may shed light on the future research and practice of a blockchain-based architecture for the privacy regulation compliance in the mobile advertising.

Acknowledgements

I would like to pay sincere gratitude to all the people who lend support and help for my Ph.D. studies.

I wish to express special thanks to my supervisor, Dr. Xuemin (Sherman) Shen for his always kind support and guidance. Dr. Shen helps me develop research skills, motivates me to work hard and pursue excellence, and encourages me to embrace and deal with challenges. My Ph.D. training cannot be completed without his valuable suggestions and advice.

I would like to thank Prof. Vojislav B. Mišić, Prof. Xiaodong Lin, Prof. Mark Crowley, and Prof. Jun Liu for being the committee members of my thesis. Their insightful discussions and comments help me identify the key technical challenges and improve the quality of the thesis.

I wish to show sincere regards to all the members of the Broadband Communications Research (BBCR) group. It is their sincere supports that carry me through the challenging times of my Ph.D. studies. It is an unforgettable experience to work with the group members, who not only provide valuable comments on my research, but also make me feel like living at home. Special thanks go to Dr. Kuan Zhang, Dr. Jianbing Ni, Dr. Haibo Zhou, Dr. Ning Zhang, Dr. Qiang Ye, Dr. Shan Zhang, Dr. Wenchao Xu, Dr. Wen Wu, Dr. Nan Chen, Dr. Weisen Shi, Dr. Junling Li, Cheng Huang, and Kaige Qu, who have helped me adapt to the studying and living at Waterloo. Special thanks go to Dr. Nan Cheng, Dr. Feng Lyu, Dr. Peng Yang, Dr. Jie Gao, Dr. Qihao Li, Dr. Yujie Tang, Dr. Wei Wang, and Dr. Anjia Yang, who have lent me their kind guidance through my Ph.D. studies. Sincere thanks go to all the BBCR alumni 2016–2020 for all the treasured moments we have spent together!

I would like to thank very much to the faculty members and the staff at the Faculty of Engineering, University of Waterloo. It is their professional and efficient work that makes my studying at UW an amazing journey.

Finally, I wish to show special thanks to my parents. It is their unconditional support and patience that help me fulfill the career goal.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

OBA	Online Behavior Advertising
PIR	Private Information Retrieval
GDPR	General Data Protection Regulation
P2P	Peer to Peer
UTOX	Unspent Transaction Output
PoW	Proof of Work
PoS	Proof of Stake
IoT	Internet of Things
KP-ABE	Key Policy Attribute-based Encryption
QAP	Quadratic Arithmetic Program
zk-SNARK	Zero-Knowledge Succinct Non-interactive Argument of Knowledge
IIoT	Industrial Internet-of-Things
IDM	Identity Management Entity
JPBC	Java Pairing based Cryptography
PoA	Proof of Authority
VC	Verifiable Computation
LDA	Latent Dirichlet Allocation
TA	Trusted Authority
SAN	Smart Advertising Network
SNARG	Succinct Non-interactive ARGument
DLog	Discrete Logarithm
SMC	Secure Multi-party Computation

Chapter 1

Introduction

1.1 Background

Mobile advertising [1, 2] is a means of advertising technologies in the digital advertising industry. It helps business to promote their products or services by disseminating customized advertisements to mobile users through multiple mobile channels. For example, a passenger can use her/his mobile phones to search nearby gas stations for trip planning [3, 4]. A fitness woman/man can receive recommendations of training facilities and nutrition plans from their smartwatches [5]. People staying at home may occasionally receive pushes from the social media that recommends nearby restaurants.

With the rapid developments of mobile devices and applications, mobile services have dominated people's daily life. From the statistics [6], mobile traffic contributes 57 percent of overall online traffic in U.S., 2017, which makes the mobile advertising the dominating advertising technology compared with tradition advertising technology, e.g., newspaper, TV, billboard [7]. Specifically, mobile advertising has its own characteristics:

- ▷ Personalized Advertising: Mobile advertising enables user personalization [8, 9] from user interactions with mobile applications. For example, a user's view and search history of a hockey game may result in an advertisement of nearby NHL events. At the same time, merchants can specify their advertisements with diversified advertising content, e.g., message, flyer, in-app push, and different advertising strategies [10], e.g., exact keyword matching or broad keyword matching strategy in Google. By doing so, advertising efficiency is greatly enhanced by disseminating ads only to targeted mobile users.

- ▷ Multi-channel Advertising: Mobile users are usually equipped with multiple mobile devices. Through a universal Personally Identifiable Information (PII) [11], e.g., a primary email address, user information can travel cross boundaries of different devices and applications. For example, a Google search at the tablet can result in a Youtube recommendation at the smart phone. Therefore, mobile advertising is able to better reach users anytime and anywhere.

- ▷ Location Awareness & Timeliness: Mobile devices are installed with location-based sensors, e.g., Bluetooth [12] and Global Positing System (GPS). As a result, mobile advertising can provide users with location-aware recommendations [13, 3] that may lead to more in-store visits. Advertisers can also dynamically adjust their ad contents based on context information, which can increase the user experience with the timely and latest information, such as coupons and flyers.

As a result, mobile advertising is the dominating technology in the advertising industry [14], that contributes 72 percent of overall digital advertising budget in U.S. by 2019 [6]. In the following, the commercialized mobile advertising model will be discussed in details.

1.1.1 Mobile Advertising Model

To achieve location-aware, timely, multi-channel, and personalized mobile advertising, it is critical to develop a novel business model [3] for the success of both merchants and mobile users. First, it is cost-ineffective for merchants, especially small businesses, to manage their user databases and ad deliveries. Second, it is shown that mobile users spend most of their time on the most popular apps, e.g., Google, Youtube, Facebook, WhatsApp. The above characteristics of mobile ecosystem have resulted in a unique mobile advertising model, which consists of three entities in its simplest form [10, 15]: mobile users, merchants, and third-party ad brokers (denoted as ad network, alternatively).

- ▷ Mobile users are equipped with a wide range of mobile devices. They enjoy various mobile services and can receive advertisements of related interests through multiple channels, e.g., search engine, in-app pushes, and in-video display ads.

- ▷ Merchants would like to promote their businesses by disseminating targeted advertisements. In the mobile advertising, merchants rely on third-party ad brokers to manage user preferences and personalized advertising strategies.

▷ Ad brokers are third-party companies that serve critical roles in the mobile advertising system. Ad brokers, e.g., Google and Facebook, build their own databases of user profiles, manage ad contents and strategies for merchants, disseminate ads to targeted mobile users and charge them for advertising dissemination services.

Merchants pay for the advertising services based on the per-click/view model [9]. Specifically, ad brokers will charge merchants when mobile users view the ad contents or click links in the ad. The third-party mobile advertising model has found many benefits. First, merchants are free from managing their own ad campaigns. Second, ad brokers can fully utilize their wealth of user data and provide effective and efficient advertising services for merchants. As a result, it has been reported that 87 percent of advertising revenue of Facebook comes from the mobile advertising services [6].

Specifically, the thesis mainly considers the following three processes of the mobile advertising system: user profiling, targeted ad dissemination, and feedback-driven marketing.

▷ User Profiling [16] is a means of collecting and managing user preference profiles. To achieve personalized mobile advertising for more efficient ad dissemination, ad brokers aim at profiling users into different advertising groups of interests. From users' interactions with mobile apps, e.g., web visits, shopping history and search activities, ad brokers can assign mobile users with keyword tags. Keyword tags can include user age, gender, interest, and locations. For example, mobile users can be assigned with 'basketball', 'beauty & fitness' in Google Ads [17], if you view an NBA video on Youtube or search a gym on Google map. Keywords can be categorized into different subsets, such as sports and technology.

  The reason behind the user profiling across different mobile applications is the utilization of PII. Mobile users can use single-sign-on authentication across many applications. For example, mobile users can access multiple services, e.g., Youtube, Google Map, Chrome with a universal Google account identified by an email address. If mobile users turn on the ad personalization option [17], the ad brokers are able to record user account activities to build ad preference files for mobile users.

▷ Targeted ad dissemination [18, 19] refers to the process of disseminating ads from merchants to related mobile users. Similar to user preference profile, merchants also choose a set of keywords as their targeting policies. Based on the merchant policy and user profile, ad brokers can run the ad campaigns with keyword matching strategies [17]. Specifically, merchants can adopt the broad matching strategy that uses fuzzy matching techniques; or exact keyword/phase matching strategy that only counts

exact matchings. Merchants can also choose where (applications) and how (banner, video, in-app pushes) the ads are dissimilated to mobile users, e.g. display ads at Google map or video ads at Youtube. Ad brokers collect ad impressions [20] from mobile users and charge merchants based on per-click/view model.

- ▷ Feedback-driven marketing can help merchants to adjust their advertising strategy and advertising budgets. For example, from ad impression reports of mobile users, merchants can track down the number of mobile users who interact with their advertisements. Ad brokers can also provide merchants with aggregated statistics [21] of their customers for better marketing decision makings. At the same time, mobile user's feedback also provides valuable reviews of the purchased services or products, which plays an important role for merchants to promote their images and increase sales.

The third-party mobile advertising model has been successful for many years, which lies the foundation of the free Internet service model.

1.1.2 Advanced Use Cases

Recently, technical advances in the next generation wireless networks (5G) and Internet-of-Things (IoT) are reshaping the mobile advertising [5] industry. With the developments of 5G and IoT [22], smart devices are seamlessly connected anywhere and anytime. IoT services, such as smart home and electronic healthy, are generating a large volume of personal data, which are of great value to merchants and ad brokers for better user profiling and ad dissemination. Ultra high-speed and low-latency capabilities in 5G also enable advertisers to provide mobile users with fruitful ad contents. For example, the emerging multicasting services [23] can increase the overall bandwidth utilization for video broadcasting advertisements [24]. The device-to-device (D2D) communication technology in 5G networks [25] realizes efficient and reliable direct content sharing among peer users [26], which can enhance the ad dissemination when the core network traffic is high.

Vehicular local advertising is a typical example of advanced mobile advertising services. Modern smart vehicles are equipped with position sensors, communication modules [27, 28], and in-vehicle multimedia systems. Vehicular users, e.g., drivers or passengers, can enjoy location-based advertising services [29, 7] for effective trip planning and entertainment. With the D2D technology, smart vehicles can also serve as the ad brokers [30] to directly dissimilate ads via D2D channels or vehicle-to-vehicle communication technology.

1.2 Security and Privacy Challenges

While the third-party mobile advertising model is enjoying its great economic success, there have also been increasing concerns over the security and privacy issues in the system, which may hinder the developments of the mobile advertising if without proper countermeasures. First, there are multiple stakeholders in the mobile advertising system: ad brokers, mobile users, merchants, application developers, and technology vendors [31]. The stakeholders, especially ad brokers and merchants, are usually profit-driven and would like to maximize their revenues as much as possible. Second, there are multiple processes in the mobile advertising system, from user profiling to feed-back driven marketing, which makes it complicated to manage and audit. Third, multiple end devices are involved in the mobile advertising, especially for the Android ecosystems. As a result, there is no universal ad library [32] for developers, which increases the risk of software vulnerability. In the following, security and privacy threats of the mobile advertising system are summarized.

1.2.1 Security Threats

- ▷ Fake advertisement. Dishonest merchants (advertisers) may not provide correct information about their services or products. In the social media advertising [33], no mandatory requirement is enforced for advertisers to register with true identities. At the same time, there is a lack of an efficient mechanism for checking the ad content. As a result, mobile users often receive annoying advertisements that are inaccurate, biased, or even misleading.

- ▷ Malvertising. The ad broker is usually a multi-sector profit-driven company, which may not always follow the pre-determined advertising strategy or provide insufficient scrutiny for malicious mobile applications. As a result, malvertising may embed links to scam webpages or cause downloads of malware [34]. This significantly increases the financial and privacy risk for mobile users in the mobile advertising system.

- ▷ Ad-fraud attack. Ad brokers collect ad impressions to charge merchants for advertising services. For third-party applications, developers can get revenue for displaying ad contents for ad brokers [15]. Therefore, it is highly motivated for developers to fake ad impressions for more revenues [35]. At the same time, dishonest or Sybil mobile users could also generate fake ad impressions. Ad-fraud attacks can greatly affect merchant confidence in the mobile advertising system, thereby resulting in a decrease of advertising investments.

- ▷ Dishonest review. The mobile advertising relies on the review system to provide user feedback to improve service quality. In some cases, user reviews can serve as the purpose of advertisements. For example, when users search a restaurant in Google map or Yelp, users are likely to view the reviews before dropping by the restaurant. Dishonest users may collude with merchants to conduct various attacks [36] against the review system, such as whitewashing and fake review [37], which may jeopardize the trustworthiness of the marketing place.

1.2.2 Privacy Threats

- ▷ User profiling requires statistics gathering of personal user activities [38], which may contain user location, personal interest, and search history. Mobile users are also tracked cross devices by device identifiers and global IDs [39]. The gathered statistics are considered as user private information that may reveal user habits, daily routines, and health status.

- ▷ Ad impression reporting requires ad brokers to collect mobile user view/click history of the advertisements for charging purposes [20]. The ad impressions may also contain user browsing history that should be prevented from being exposed to the adversaries.

- ▷ User feedbacks and reviews may reveal their shopping history and locations, which may result in the leakage of personal information. At the same time, users may be reluctant to leave negative feedback for the fear of the consequences [36].

1.3 Research Motivations and Challenges

Security and privacy threats are hindering the developments of the mobile advertising system. Ad-fraud attacks are costing money, which is predicted to be 43 billion US dollars in 2019 by TrafficGuard/Juniper [40]. Fake advertisements are annoying mobile users and driving them to turn off the ad personalization while dishonest reviews are reducing user confidence in the marketing place. For the privacy threats, as European General Data Protection Regulation (GDPR) [41] takes effects, strict legal requirements on collecting, storing, and sharing user personal data are enforced. That is, mobile users are granted the legal rights to fully control their personal data, which drives the mobile advertising industry to update its privacy policies.

Extensive research efforts have been directed to the solutions to the security and privacy threats of the mobile advertising. The thesis investigates the mobile advertising system from an architecture perspective and addresses the fundamental reason behind the mentioned threats: lack of system transparency [42]. Specifically, the advertising transparency requires advertising policies and processes are open and verifiable to the involved parties, e.g., mobile users and merchants.

- ▷ Mobile users should be aware of their preference profiles: how they are tracked cross different devices and why they are assigning the keyword tags.
- ▷ Advertising policy should be transparent to mobile users [43]. Mobile users are annoyed when they receive biased or even misleading ads, which may result in the increasing popularity of the ad-blocking software [44].
- ▷ Ad dissemination process should be verifiable to both the mobile users and merchants. Mobile users should know why they receive a specific ad [33] while merchants should know their ads reach targeted mobile users.
- ▷ Review system should be open and transparent to the public, where the accumulation process of each merchant should be verifiable by mobile users.

By increasing the advertising transparency, it motivates the involved parties to follow the pre-determined advertising protocol and build a reliable advertising system. The public awareness of advertising activities can also be enhanced, which pushes the stakeholders to take effective and efficient actions against any advertising misconduct. Therefore, it is urgent to design a transparent advertising system to address the security and privacy threats at the architecture level. Ad brokers are making their efforts to increase the system transparency by providing users with more explanations [43, 33]. At the same time, many research efforts have been put into increasing the transparency of user profiling, ad dissemination, and marketing, which may rely on web extensions [42], independent auditors [45], and trusted hardware [46]. The above solutions assume the trust of a single authority, which can sometimes be untrustworthy. For example, major ad brokers are reported to pay for blocking software to be put into the white list [47].

A decentralized solution that relies on distributed consensus is more promising for a transparent and reliable mobile advertising system [36, 48]. Originated in its success from electronic cash [49], blockchain is a distributed ledger maintained by mutually distrustful peer nodes, that is transparent, immutable, and open. There have been discussions on utilizing the blockchain to serve the role of ad broker [48, 50] or a marketing place [48] with

early industrial attempts [51], e.g., AdChain [52]. Specifically, a blockchain-based mobile advertising system has found the following benefits: (1) transparent and controllable user profiling. (2) verifiable ad dissemination and charging. (3) open and reliable review system [53, 51]. However, building a blockchain-based mobile advertising system also faces non-trivial obstacles:

- ▷ Efficiency. The distributed consensus of the blockchain comes at the cost of system efficiency, in terms of transaction throughput and confirmation time. In a public blockchain, e.g. Bitcoin and Ethereum, every transaction needs to be verified and stored at each miner node. In a consortium blockchain [54], transactions are verified by validating nodes. Considering the large volume of advertising data, a straightforward approach to utilize the blockchain may pose prohibitive computation and communication costs, which motivates the first design challenge in this thesis: a blockchain-based mobile advertising system with practical designs.

- ▷ Privacy. The open nature of the blockchain makes it suitable to increase the advertising transparency. At the same time, user profiles and ad impressions are required to be kept private by the emerging privacy regulations, which creates a conflict with the blockchain transparency. Therefore, how to design a blockchain-based mobile advertising system that strikes a notable balance between system transparency and user privacy is another challenge that the thesis aims to resolve.

1.4 Research Contributions

The thesis focuses on addressing the security and privacy threats in the mobile advertising system by increasing the advertising transparency. Specifically, the thesis aims at building a blockchain-based architecture that is compatible with the current mobile advertising system and resolves the design challenges of the efficiency and privacy requirements with practical design and feasible implementations. The main contributions of this thesis are summarized as follows:

- ▷ The thesis proposes an anonymous reputation system that preserves consumer identities and individual review confidentialities. To increase system transparency and reliability, the thesis further exploits the tamper-proof nature and the distributed consensus mechanism of blockchain technology. With system designs based on various cryptographic primitives and a Proof-of-Stake (PoS) consensus protocol, the

proposed blockchain-based reputation system is more efficient to offer high levels of privacy guarantees compared with existing ones. Finally, the thesis explores the implementation challenges of the blockchain-based architecture and present a proof-of-concept prototype system by Parity Ethereum. The thesis measures the on/off-chain performance with the scalability discussion to demonstrate the feasibility of the proposed reputation system.

- ▷ The thesis develops a transparent and accountable vehicular local advertising system by utilizing the blockchain technology. Considering the prohibitive cost of directly implementing a large-scale advertising system on the blockchain, the thesis introduces two design strategies, *digest-and-verify* and *divide-then-assemble*. In specific, a large-scale spatial keyword database is digested and stored on the blockchain with succinct cryptographic authenticators. The ad dissemination is then conducted with modular executions of two off-chain spatial keyword query functions, the results of which are assembled and verified in an advertising smart contract using the stored authenticators. By doing so, expensive on-chain computation and storage overheads are significantly reduced at a cost of acceptable off-chain overheads. The thesis formalizes the security requirements of the vehicular local advertising system as *Auditing Security* and achieves the notion with thorough security analysis. Extensive experiments are conducted to demonstrate the practicality of the blockchain-based vehicular local advertising system.

- ▷ The thesis proposes a blockchain-based Smart Advertising Network with Privacy-preserving Accountability (*SANPA*). Specifically, the thesis designs a composite Succinct Non-interactive Argument (*SNARG*) system, that commits advertising policies as cryptographic authenticators in a smart contract. By doing so, *SANPA* is compatible with the existing *SAN* without posing prohibitive implementation cost over the blockchain architecture. Users or retailers can require explanations of an advertising activity by sending a challenge to the smart contract. With the succinctness and privacy preservation of *SNARG* system, the contract can efficiently verify whether the challenged advertising activity follows committed policies without exposing user profile privacy. If the misconduct is identified, the contract enforces public accountability on the involved parties by confiscating their cryptocurrency deposits. The thesis conducts extensive experiments to provide both on-chain and off-chain benchmarks, which demonstrates the application feasibility of *SANPA*.

1.5 Outline of this Thesis

The organization of this thesis is as follows. Chapter 2 presents a comprehensive survey of related literature on the security and privacy issues of the mobile advertising system and the blockchain techniques with applications to public accountability infrastructure. Chapter 3 presents the first research contribution: Anonymous Reputation System for IIoT-enabled Retail Marketing atop PoS Blockchain. Chapter 4 presents the second research contribution: Transparent and Accountable Vehicular Local Advertising with Practical Blockchain Designs. Chapter 5 presents the third research contribution: Exploiting Blockchain for Transparent Mobile Advertising with Privacy-preserving Accountability. Chapter 6 summaries the thesis with future research directions.

Chapter 2

Literature Review

In this chapter, we review the state-of-the-art literature regarding security and privacy challenges and solutions in mobile advertising. Specifically, we first review the solutions based on trusted single authorities. Then, we discuss the basics of the blockchain technologies with their applications to enhancing trust and transparency in mobile advertising systems. Finally, we review the non-interactive cryptographic argument techniques.

2.1 Security and Transparency for Mobile Advertising

Extensive research efforts have been put into increasing the security and transparency for the mobile advertising, including comprehensive surveys [43, 33], intelligent web extensions [42], third-party explainers [45], and trusted hardware [46].

Son *et al.* [39] carefully studied the implementations of ad libraries in Android systems. The authors identified specific interfaces that can be utilized by third-party developers to infer user personal information during the advertising. Their findings highlighted the essentiality for the current mobile advertising system to increase security measures against malicious developers. Gui *et al.* [55] collected and analyzed a large volume of ad complaints from advertising users. Their results revealed the users' lack of confidence in the content and relevance of the received ads, which could lead to the installation of ad-blocking software. Chen et al. [34] conducted an extensive analysis of the existing mobile apps and ads. Their results showed the prevalence of fraud or irrelevant ads and an urgent need for a more transparent and accountable advertising system. Andreou *et al.* [43] investigated

the transparency explanations of Facebook advertising, regarding how a user is labeled with attributes and why the user receives a specific ad. By collecting a large amount of explanation data from different users, the authors concluded that the transparency explanations by the brokers are often incomplete, vague, and even misleading. Later, Andreou *et al.* [33] investigated the behavior of advertisers in social network advertising. The results demonstrated the need for an effective and efficient mechanism to publicly audit advertiser activities and enforce accountability against advertising misconduct.

Many research works have been directed to develop countermeasures against *ad-fraud* attacks. *Ad-fraud* attacks are conducted by malicious application developers who fake user click/view for charging merchants for advertising display. Crussell *et al.* [32] made the efforts to adopt machine-learning based traffic analysis techniques and identify fake ad impressions. Specifically, the proposed scheme automatically investigated two types of *ad-fraud* attacks: ad impression reporting with hidden app running or without user interaction. Shao *et al.* [56] investigated the interface between mobile apps and web pages to study the hidden mobile attacks and frauds. The proposed system helps mobile users identify the accountability of publishers or the ad broker. The study presented an overall overview of malicious applications, which demonstrated the need of pursuing joint accountability of application developers, web owners, and the ad broker. Dong *et al.* [57] identified various types of mobile *ad-fraud* attacks and designed an automatic fraud detection framework based on UI state transitions and android application analysis techniques. Specifically, the proposed framework dynamically evaluated the UI state transitions to detect hidden ad displays with scalable and accurate detection rates. Jin *et al.* [58] conducted a comprehensive study of ad libraries in existing ad broker, which identifies the main characteristics of ad libraries that may have impacts on the mobile ad-fraud attacks. The proposed scheme investigated the ad libraries in an API level to provide a comprehensive classification of ad types and API interfaces. Chen *et al.* [34] studied the relationship between click frauds and malvertising and proposed an integrated framework that captures ad traffics and detect malicious ads.

At the same time, researchers have made their efforts on increasing the transparency of the current advertising system to regain user and merchant confidence. Li *et al.* [46] used trusted hardware techniques (ARM trusted zone) to propose a verifiable advertisements click and display framework on mobile applications. The proposed method not only increased the advertising transparency but enhanced the prevention of *ad-fraud* attacks. Jin *et al.* [18] designed a web extension to manage the flow charts, that explains to users the ad selection and profile control processes. The use of web extensions to manage the advertising system provides mobile users with easy-to-implement choices for advertising transparency explanations. Parra-Arnau *et al.* [42] developed a detection system that looks into users'

web browser profile with a measurement scheme for user profile uniqueness, which enables configurable and flexible transparency options for web users. The designed web extension also provides users with selective transparency and ad blockading options. Venkatadri *et al.* [45] utilized a third-party organization as a transparency explainer to manage the transparency explanations for users. Specifically, the third-party explainer will respond to user request of transparency explanations by collecting advertising policies and investigating the targeting process on the ad network without learning additional user information.

To summarize this subchapter, the security and transparency issues of the mobile advertising system have attracted research works from different perspectives. Different from the existing methods that rely either on a single authority or software, this thesis explores a different path to investigate the challenges and solutions of a distributed blockchain-based solution for the mobile advertising system. By doing so, the public transparency could be significantly increased for the advertising system.

2.2 Privacy-preserving Mobile Advertising

The subchapter reviews recent research activities on designing privacy-preserving advertising systems for mobile users, including the user profiling, the targeted advertising [19], and the private ad impression report. Moreover, the subchapter investigates the privacy regulations and their impacts on the mobile advertising system.

Private user profiling and targeted advertising have attracted extensive research efforts. Guha *et al.* [59] investigated the privacy-preserving targeted advertising issue by introducing a new entity, i.e., dealer, between the users and the ad brokers. The dealer anonymizes the users by breaking the linkability between users and their profiles and enabling private reporting of ad impressions. Backes *et al.* [60] studied the privacy issues in Online Behavior Advertising (OBA). Specifically, the proposed scheme utilized Private Information Retrieval (PIR) [61] for private ad dissemination and oblivious token techniques for ad impression reporting. From the cryptographic primitives, the proposed scheme also achieved provably security for the advertising system. Hardt *et al.* [38] exploited the differential privacy technique [62] for private advertising statistics aggregation and reporting to identify the critical tradeoffs between user privacy and ad personalization accuracy. Davidson *et al.* [63] increased the user profile privacy by shifting the computations of ad disseminations to the client side, therefore decreasing the amount of data that leaves users' devices. Jiang *et al.* [10] proposed a private targeted advertising system based on Private Stream Searching [64] to protect user profile privacy. The proposed scheme also utilized homomorphic

encryptions to break the role of the ad brokers in ad impression reporting to protect user impression privacy.

Toubiana *et al.* [9] pried into the private ad impression issue if there is a malicious user that does not honestly report ad click/view statistics. Based on the zero-knowledge proof technique, the proposed scheme enables the users to prove the correctness of their impression reporting without leaving individual statistics. Green *et al.* [20] further investigated the scalability issue for private ad impression reporting based on an efficient zero proof technique [65] with some optimization techniques. Qian *et al.* [66] also studied the selective ad impression aggregation issue with novel designs from differential privacy and homomorphic encryptions. As the General Data Protection Regulation (GDPR) takes effect in 2018 [41], strict restrictions on mobile applications to share personal user data with third-party ad brokers are enforced. Specifically, *GDPR* defines rights of personal users, e.g. right to be informed and right to be forgotten, that must be protected in the advertising system. Urban *et al.* [31] further discussed the impacts of *GDPR* on the data sharing and usage of cookies on the advertising industry.

To summarize this subchapter, state-of-the-art literature has explored a wide range of techniques in real-word practices to achieve a privacy-preserving advertising system. Different from the existing literature, this thesis utilizes the blockchain-based architecture for the mobile advertising with inherent transparency and openness nature. Thus, how to enjoy the benefits of the blockchain while preserving user profile and impression privacy is one of the main challenges in this thesis.

2.3 Privacy-preserving Reputation System

Trust and reputation management is becoming prevalent for the success of a global marketing system [67, 68]. People are used to referring to reviews of a product or a shop before their shoppings. This subchapter reviews the research advances on building a trusted and transparent reputation system [69, 70, 37, 71, 72].

Blomer *et al.* [70] proposed a reputation system based on group signature technique. The proposed scheme enables users to leave an anonymous review for a product or service. Motivated by [70], Blomer *et al.* [37] further proposed a feedback-driven reputation system with public linkability. The main goal of the proposed system [37] is to preserve consumer anonymity while preventing the double-review attack. The proposed scheme formalized the security requirement of the anonymous reputation system with comprehensive proofs. Bag *et al.* [73] proposed a personalized reputation system taking into consideration of

the trustworthiness of consumers. Specifically, the proposed scheme utilized homomorphic encryptions with non-interactive zero-knowledge proof technique to preserve the individual review of each user. Bazin *et al.* [72] designed a feedback-driven reputation system with secure rating aggregations. Non-interactive zero proof technique [74] was combined with blind signature in [72] to achieve consumer anonymity. Zhai *et al.* [69] proposed a tracking-resistant anonymous reputation system by leveraging an anonymity provider with mix-net technology. By doing so, the proposed scheme can break the links between the original user review and mixed ones, therefore concealing the individual review statistics. Azad et al. [71] utilized a homomorphic cryptographic system and the non-interactive zero-knowledge proof to design a decentralized reputation system with individual rating score confidentiality.

To summarize this subchapter, existing literature for blockchain-based reputation systems has achieved a variety of properties such as anonymity, decentralization, and system transparency. Due to the lack of mutual trust among mobile advertising system and profit considerations, a distributed reputation system with verifiable transparency is preferred. This thesis takes advantage of the blockchain technology for a trustworthy reputation system in a fully distrustful environment. At the same time, the thesis addresses the implementation challenges in the design of the reputation system to achieve compatibility with existing blockchain platforms.

2.4 Blockchain and Smart Contract

The blockchain is a distributed ledger maintained by a peer-to-peer (P2P) network, where each participant can send transactions to another without going through a centralized authority. Specifically, blockchain is a chain of blocks that are chained together by secure hash functions [75]. Each block consists of a number of P2P transactions that directly transfer cryptocurrencies from the sender to the receiver(s). The blockchain adopts consensus protocols among participants to help them maintain a consistent view of the distributed ledger. Compared with traditional distributed system, the blockchain motivates the participants to join the maintenance of the blockchain by providing reward mechanisms [76, 77].

The blockchain is first introduced as the underlying infrastructure of the digital currency Bitcoin [49] in 2009. Digital currency [78], also known as electronic cash, is the divisible and unforgeable virtual currency with exchange values. Bitcoin is the digital currency network with the following characterizes:

- ▷ Distribution: There is no centralized authority in the pure P2P Bitcoin network.

- ▷ Openness: The Bitcoin network is an open-source project. Any network peer can join the Bitcoin network.
- ▷ Anonymity: A participant of the Bitcoin network only provides a pseudonym and a unique public/secret key pair [79].
- ▷ Transparency: Any transaction in the Bitcoin is publicly verifiable to the whole network.
- ▷ Immutability: Any transaction that is confirmed in the Bitcoin network cannot be later modified, which makes the Bitcoin an append-only ledger.
- ▷ Consistency: Every participant of the Bitcoin network should maintain the same view of the shared ledger.
- ▷ Double-spending prevention: No malicious participant can spend the same digital currency at different transactions.

In each block of the Bitcoin, there are a few transactions included with a compact Merkle proof [80]. Each block also consists of a hash digest of the previous block. Since there is no centralized authority to record financial information (remaining balance, transaction history, etc.), the Bitcoin adopts the unspent transaction output (UTOX) to manage P2P transactions [49]. Specifically, a sender can prove to the public that she is the owner of a few UTOX with a digital signature. Then, the sender includes the UTOX and the address of the receiver into a transaction, digitally uses the private key to sign the transaction. The sender finally sends the signed transaction to the blockchain network [49].

There may be different participants proposing transactions at the same time. Therefore, it is critical to determine the next valid transactions for the ledger. This is determined by the consensus protocol in the Bitcoin, e.g. Proof of Work (PoW) [76] and Proof of Stake (PoS) [77]. A special participant in the Bitcoin is called miner, who is responsible for collecting, verifying, and proposing the next block of the Bitcoin network. In PoW, different miners compete to solve a hash puzzle. In PoS, a set of slot leaders are randomly determined from miners, each of which will collect and propose the block at her time slot. The Bitcoin rewards the miner who proposes the next block with 25 bitcoins. By doing so, miners are motivated to maintain the correctness and integrity of the Bitcoin network.

Etherem [81] is another distributed blockchain network that supports its own digital currencies, i.e. Ether. It shares many common features with the Bitcoin, e.g. anonymity, openness, transparency, etc. However, Ethereum implements an account-based model [81] instead of the UTXO model in the Bitcoin. It is observed that the blockchain can be

regarded as a state machine, where each successful transaction changes the blockchain from one state to another. Ethereum first introduced the smart contract [82] to the blockchain network, which is a trusted computer program. The smart contract specifies the terms and conditions of involved parties (blockchain participants) and can take actions (transferring cryptocurrencies) when conditions are met. In Ethereum, a smart contract is a special address on the blockchain network with data and codes stored. Each Ethereum node call contract functions by sending transactions with the data to the contract address. Compared with a traditional contract, the Ethereum smart contract has the following features [83]:

- ▷ Trustworthy witness: Benefiting from the transparency and immutability nature of the blockchain, the integrity and authenticity of a smart contract is ensured.
- ▷ Automatic execution: Ehtereum ensures that a valid transaction will be confirmed in the ledger within a threshold time. Therefore, efficient and automatic contract execution is ensured in the smart contract.

Smart contract is suitable for many applications, such as supply chain management [84, 85], and digital goods exchange [86].

Both the Bitcoin and Ethereum have been successful by building trust among distrustful peer nodes. At the same time, the underlying public blockchain infrastructure also raises concerns over its efficiency and scalability [87]. For example, the block time (time between two consecutive confirmed blocks) in Bitcoin is around 10 minutes and average of about 12,000 transactions are confirmed per hour [88]. In contrast to the public blockchain, consortium or private blockchain [54, 89] is proposed for industrial applications, where there exists a certain degree of trust between industrial partners [84]. Compared with the public blockchain, consortium or private blockchain enforces identity and access management for the network participants. For example, a group of financial institutions can utilize the consortium blockchain for efficient reconciliations.

To realize the promises of the blockchain technology, extensive research efforts have been directed to increase its scalability while addressing security and privacy issues [90, 91]. Blockchain sharding [92] was proposed to enable parallel processing of the ledgers. Another research line focused on separating on-chain and off-chain blockchain computations [93, 94]. For example, off-chain payment channels [95] could be established to process a number of transactions before uploading the transactions to the ledger. At the same time, privacy is also a primary concern for many blockchain-based applications. For example, users would not like to expose their transaction history to the public since it may contain some sensitive information [96]. Terms and conditions in a smart contract could also be sensitive

and should be concealed from the public view, which motivated the research of privacy-preserving smart contract [97].

To summarize this subchapter, blockchain is a promising technology that promotes trusted and transparent consensus in distributed systems. However, there still remain some unresolved privacy and efficiency issues. This thesis aims at adopting the blockchain technology for a more reliable and transparent mobile advertising system while addressing two challenges at the same time: practical designs and privacy-preserving yet public accountability.

2.5 Blockchain-based Accountability Infrastructure

The practices of building a blockchain-based accountability architecture have been explored in a wide range of application scenarios, such as personal data sharing [98, 99], transparent legal systems [100], vehicular forensics [101], Internet of Things (IoT) [102], and protocol accountability verifications [103].

Frankle *et al.* [100] investigated the accountability issues in a secret process of a court system, such as the surveillance of a criminal target. The authors utilized the public ledger and multi-party computation techniques to achieve the system transparency and target privacy at the same time. Li *et al.* [101] constructed a blockchain-based vehicular forensics framework that enforces accountability and fine-grained access control over the forensics data. The authors built a distributed Key Policy Attribute-based Encryption (KP-ABE) scheme that prevents dishonest agencies to abuse the power. Neisse *et al.* [98] studied the General Data Protection Regulation (GDPR) for personal data usages and proposed a blockchain-based framework that realized personal data accountability and provenance tracking. Wu *et al.* [99] further pried into the detailed policy of GDPR in terms of transparency and compliance requirements. The authors proposed an offline channel and business relationship model for a blockchain-based architecture to boost the system scalability. Boudguiga *et al.* [102] adopted the blockchain to achieve the secure updates of the objects or nodes in the IoT. The proposed scheme increased the availability and accountability of the IoT system with a transparent ledger.

Li et al. [104] proposed a blockchain-based carpooling system, which boosted the information exchange among carpooling companies and ensured the carpooling user privacy. The authors adopted anonymous credential technique in the blockchain to achieve transparent and privacy-preserving collaborative carpooling system. Li et al. [30] proposed a blockchain-assisted ad dissemination scheme, which motivated the vehicles in road networks to act as ad brokers with fair rewards for ad dissemination services. Zhang et al.

[105] proposed a blockchain-based public-key searchable encryption scheme. The proposed scheme utilized the blockchain as a trusted storage and addressed the keyword guessing attacks for the searchable encryption. Li *et al.* [106] developed a transparent crowdsourcing framework based on the Ethereum smart contract. The designed and implemented framework provided comprehensive benchmarks for future research in the blockchain-based crowdsourcing applications. Lu *et al.* [107] designed a privacy-preserving crowdsourcing framework based on the blockchain and the anonymous credential technique, which addressed the conflicts between blockchain transparency and user privacy. Dorri *et al.* [108] proposed a blockchain-based framework for automobile services in vehicular networks. Liu *et al.* [109] designed a blockchain-based solution for distributed network provenance. The proposed scheme ensures the integrity and correctness of cross-domain provenance queries between network administrators. Frey *et al.* [110] proposed a blockchain-based recommendation system for the e-commerce system. Nguyen *et al.* [111] utilized the blockchain to construct a geo-marketplace for trading location data in a transparent manner.

To build a more transparent marketplace [112, 113], blockchain technologies have been exploited for reputation system construction [114, 36]. Schaub et al. [114] proposed a fully decentralized reputation system atop a public blockchain with the blind signature technique to achieve consumer anonymity. Soska et al. [36] proposed an anonymous reputation system based on the ring signature and the robust transaction chain property of the blockchain technology. There have been recent advances on building blockchain-based searchable encryption schemes. Hu *et al.* [115] utilized the smart contract to construct a symmetric searchable encryption scheme. The proposed scheme achieves the verifiability of search results from the transparency of the Ethereum blockchain. Xu *et al.* [116] designed a compact authenticated structure for searchable indexes that supports rich search functionalities on the blockchain. The proposed scheme also explored the implementation challenges by desiring cross-block authenticated indexes.

To summarize this subchapter, extensive research works have been directed to utilize the blockchain technology to increase the transparency and accountability of the existing systems. At the same time, practical implementation issues of the blockchain technologies. e.g. scalability and privacy, pose significant challenges when applying the blockchain to the mobile advertising system.

2.6 Non-interactive Argument

Non-interactive Argument *NArg* [117, 118] enables a prover to prove to a verifier that an instance (x, w) with a statement x and a witness w satisfies a (public) relation $\mathcal{R}$. *NArg*

is widely used in many applications, e.g. digital signature [79], anonymous credential [119, 120, 121], and verifiable computations [122]. Recently, *NArg* systems are lively research areas due to the booming trends of the blockchain. This thesis focuses on a research line of *NArg* systems in bilinear groups, that recognize the Quadratic Arithmetic Program (QAP)-based relations and the low-degree algebra relations. A *NArg* scheme should achieve *soundness*, *complexness* and *non-interactive*. Informally, *soundness* ensures that a prover cannot forge an instance (x', w') that is not in $\mathcal{R}$ to pass the vesication. *Completeness* guarantees that an honest verifier always accepts correct instances (x, w). *Non-interactive* requires *NArg* to be an one-move proof system. *NArg* can achieve *zero-knowledge* by concealing instances (x, w) in the verification.

2.6.1 Low-degree Polynomial Relations

NArg for low-degree polynomial relations can be efficiently instantiated in elliptic curve groups equipped with bilinear pairings [123, 124]. This thesis only considers knowledge in the exponent with Pedersen commitment schemes [125] and single discrete logarithms [74, 126]. If we denote $\mathbb{G} = (\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T)$ with a prime order p and generators $g_1, g_2, ..., g_n \in \mathbb{G}_1^n$. A few examples of such relations are shown as follows:

- ▷ Given public parameters g_1^a, g_2^a and secrets $a \in \mathbb{Z}_p$, a *NArg* system proves that g_1^a and g_2^a contain the same secret a.
- ▷ Given public parameters g_1^a, g_2^b, g_3^c and secrets $a, b, c \in \mathbb{Z}_p^3$, a *NArg* system proves that $c = a + b$.
- ▷ Given public parameters $\prod_{i=1}^n g_i^{a_i}, \prod_{i=1}^n g_i^{b_i}$, a *NArg* system proves that $(b_1, b_2, ..., b_n) \in \mathbb{Z}_p^n$ is a permutation of $(a_1, a_2, ..., a_n) \in \mathbb{Z}_p^n$.

With the homomorphic property of the Pedersen commitment, additions in the exponent can be efficiently implemented. With a bilinear pairing, an multiplication in the exponents can be efficiently implemented. Generally, a three-move public-coin Sigma protocol with pre-image based proof techniques [74] can be converted to the non-interactive setting with Fiat-Shamir heuristics [127]. In the following, the thesis reviews how the *NArg* systems for different polynomial relations in the discrete logarithm setting are realized.

Pointcheva and Sanders [128] constructed a novel group signature scheme with applications to anonymous credentials, which is a proof of knowledge of a signature in the discrete logarithm setting. The proposed scheme [128] achieves a short proof size, which makes it

more suitable for blockchain-based applications. Later, the construction of the PS signature [128] was improved in [119] to achieve security properties based on non-interactive assumptions. Bootle *et al.* [129] constructed an efficient *NArg* system for arithmetic circuit evaluations in bilinear groups. At the heart of their design is a zero-knowledge proof system for inner product evaluations. With the inner-product technique, a polynomial evaluation is constructed for evaluating a circuit with logarithmic complexities. Subsequently, Bunz *et al.* [130] constructed an efficient range proof technique that convinces the verifier that a commitment lies in a certain range. The proposed Bulletproof system also supports proof aggregations, which makes it suitable for e-cash applications. Bootle *et al.* [131] further improve the previous work [129] for low-degree polynomials with batch techniques. Damgaard *et al.* [132] developed a zero-knowledge proof system for Hamming weight evaluations for a set of Pedersen commitments. The novel polynomial evaluation technique in [132] achieves a proof system with a compact proof size. Lai *et al.* [133] constructed a set of proof systems for vector commitments from various assumptions, e.g. Root assumption or CDH. The proposed scheme supports opening a vector commitment at given positions and can also evaluate linear combinations of the vector commitments.

The proof system in the discrete logarithm requires no trusted setup of the public parameters. At the same time, the complexity of such systems, in terms of prover and verifier, is often linear or logarithmic with the size of the polynomial relations.

2.6.2 Quadratic Arithmetic Program-based Relations

Arithmetic circuit evaluation can be recognized by a Quadratic Arithmetic Program (*QAP*). A *QAP* consists of three sets of polynomials constructed from the original arithmetic circuit. Given the assigned values of each I/O and each intermediate gate of the circuit, the *QAP* evaluates the three sets of polynomials and conducts a divisibility check with a targeted polynomial. This results in the evaluations in bilinear groups with parings, which is also denoted as *Zero-Knowledge Succinct Non-interactive Argument of Knowledge (zk-SNARK)* in the literature. In the following, the thesis summarizes the research progress of *QAP*-based argument systems with applications to veritable computations.

Gennaro *et al.* [134] exploited the *QAP* theory for arithmetic circuit evaluations, which converts the circuit evaluation to a low-degree divisibility check in bilinear groups. The proposed *SNARG* system is non-interactive and succinct, which results in the notable storage and computation efficiency at the verifier and thus makes it suitable for blockchain-based applications where on-chain storage and computation are expensive. Subsequently, Parno *et al.* [122] constructed a more efficient *QAP* scheme, that reduced the degree

of compiled programs, the key size and prover computation overhead. Parno *et al.* also proposed a comprehensive framework that can compile a subset of C programs into a *QAP*, which resulted in practical verifiable computations for general relations. Ben-Sasson *et al.* [135] designed a framework for verifiable C program executions from linear probabilistically-checkable proofs. Ben-Sasson *et al.* [136] also proposed a set of optimization techniques to improve the computation and storage efficiency of the Pinocchio framework [122]. Costello *et al.* [137] extended the *QAP* to multi-*QAP* that enables multiple uses of the same data across different loops, which reduced the circuit size and prover computation overhead. Later, Groth *et al.* [138] constructed new instantiations of the *QAP*-based proof system, which greatly reduced the proof size and verifier computation overheads with only a few group elements in the proof. Kosba *et al.* [139] further designed a JAVA programming framework that improved the program-to-circuit interface with more compact circuit size. The proposed framework utilized the *libsnark* library [140] as the backend for proof systems.

Compared with the *NArg* system in the discrete logarithm setting, the main feature of the *QAP*-based *NArug* system is the verifier efficiency. For example, succinct proof sizes and verification costs are achieved in [138]. However, *QAP*-based *NArug* system requires a trusted setup of the common reference strings since a trapdoor secret s is used in the setup phase. Moreover, *QAP*-based *NArug* system utilizes general circuit constructions, which makes it less efficient for specialized relations, e.g. proof of knowledge of a signature in elliptic curve groups.

2.6.3 Composite Arguments

Since different relations can be efficiently instantiated with different *NArg* systems, it is promising to carefully design a campsite argument system that combines the advantages of different techniques.

Fiore *et al.* [141] utilized the multi-exponentiation component in the proof of *zk-SNARK* systems, that is actually an extended Pedersen commitment of the circuit input and can be reused multiple times for different instances. A hash-and-prove scheme was proposed in [141] for outsourced verifiable computations. Chase *et al.* [142] proposed a zero-knowledge proof system that combines traditional Sigma protocol with the *zk-SNARK* techniques. The proposed scheme finds many interesting applications, e.g. proof of knowledge of a signature if the message is first hashed. Later, Agrawal *et al.* [143] formalized composite arguments, that utilized *zk-SNARK* as the core computing component with committed inputs/outputs based on traditional Sigma protocols in the discrete logarithms. Campanelli *et al.* [144] proposed a framework with modular utilizations of *NArg* systems from *QAP*-based relations and polynomial relations.

To summarize this subchapter, non-interactive argument systems are powerful techniques that enable verifiable computations and the proof of knowledge for polynomial relations. At the same time, the complexity of the current mobile advertising system, e.g. various privacy and transparency requirements for different processes, raises many challenges in designing a versatile and efficient argument system. Therefore, the thesis carefully tailors the design of non-interactive argument systems to achieve the rich functionalities in the mobile advertising system with strong security, privacy, and efficiency guarantees.

2.7 Summary

This chapter reviews the existing works that are closely related to this thesis. First, the chapter discusses the security and transparency issues in current mobile advertising systems with state-of-the-art technical solutions. Second, the chapter presents the privacy issues and review the research lines of privacy-preserving advertising systems. Third, the chapter studies the basics of the blockchain and smart contract technologies with applications to trusted accountability infrastructures for distributed systems. Finally, the chapter investigates the *NArug* techniques in the discrete logarithms and QAP-based relations from bilinear pairings with different use cases.

Chapter 3

Anonymous Reputation System for IIoT-enabled Retail Marketing atop PoS Blockchain

3.1 Background

Industrial Internet-of-Things (IIoT) [145], which consists of a global network of smart objects, is reshaping and revolutionizing the retail industry [146]. In a global retail ecosystem, suppliers, manufactures, and retailers are adopting IIoT to improve manufacturing operational efficiency and reduce supply-chain management cost [147, 85]. Leveraged with cloud computing and big data technologies, IIoT is also envisioned to benefit the retail marketing that speaks to the needs of competitive market globalization and consumer demand diversification [147]. With the help of IIoT technology, retailers are able to collect massive feedbacks from various sources and devices, which can help them better manage their business. In particular, consumer feedbacks play a critical role for retailers to establish reputations among industrial partners and build consumer confidence [148]. Specifically, consumers are allowed to leave feedbacks (usually a rating score and/or a review message) for their experiences with retailers [36]. These feedbacks accumulate over time and can be enumerated by other entities in the retail industry.

However, there are still some challenging issues that could hinder the development of a reliable retail reputation system. Firstly, the process of leaving feedbacks may reveal much personal consumer information, which can be used to track and profile consumers [37]. Moreover, consumers may be reluctant and compelled while leaving a negative review to

a specific retailer in the fear of related consequences [71]. Simply leveraging pseudonyms for rating anonymity cannot resolve this concern, which can suffer from de-anonymization attacks [149]. Secondly, current reputation systems mainly utilize a centralized marketplace that collects and accumulates consumer reviews. However, it has been evidenced that the current centralized marketplace may fail to keep their promise of a desired trust level due to the leak of private consumer information and lack of system transparency [36].

The shortcomings of current reputation systems motivate the research efforts on designing a reputation system that provides strong consumer anonymity guarantees [69, 70, 37, 71] without relying on a centralized marketplace [114, 36, 72]. Besides anonymity, reputation systems are required to resist to various attacks (such as self-rating and Sybil attacks [150]), which becomes more challenging in a decentralized marketplace [72]. However, existing decentralized solutions for reputation systems provide insufficient system transparency, which is essential for the IIoT-enabled retail marketing due to lack of mutual trust among the involved entities. To realize a more open and transparent reputation system, extensive research efforts have been directed to the design of a blockchain-based architecture [36, 114]. In their designs, blockchain serves as an immutable ledger where the review generation and reputation accumulation process can be publicly verified and traced. The underlying consensus and incentive mechanisms of blockchain technology also contribute to the boost of mutual trust among consumers and retailers. Although these attempts [36, 114] have exploited blockchain technologies for building up a promising reputation system, the proposed systems pay insufficient attention to the efficiency and scalability issues of the blockchain technology [84]. Moreover, implementation challenges of a blockchain-based architecture have not been well investigated.

To address the issues, we propose an **A**nonymous **R**eputation **S**ystem atop a **P**roof-of-**S**take blockchain (ARS-PS). The proposed ARS-PS allows retailers to establish reputations by accumulating feedbacks from consumers. Meanwhile, the ARS-PS ensures that retailer reputation accumulation process is transparent to the public while providing strong anonymity to consumers. The contributions of this chapter are summarized as follows.

- ▷ We design an efficient and anonymous reputation system by leveraging a randomizable signature [119, 37] with non-interactive zero-knowledge proof technique [127, 130]. The proposed system preserves the identity and the individual review confidentiality of the consumer. Only the aggregated review statistics for retailers is revealed to the public.
- ▷ We exploit the consensus protocol in [77] and design a blockchain-based architecture to integrate the proposed anonymous reputation system. Our designs enhance

the system transparency for review generation and reputation accumulation while preserving the consumer anonymity and accountability at the same time.

- ▷ We explore the implementation challenges of the blockchain-based architecture: (1) compatibility with current blockchain platforms; and (2) insufficient support for cryptographic primitives. We develop a proof-of-concept prototype system based on Ethereum Parity [151]. We build a testing blockchain network and the experimental results demonstrate efficiency and feasibility of the proposed ARS-PS.

The remainder of this chapter is organized as follows. In Section 3.2, the system model, security model, and design goals are presented. In Section 3.3, we present the building blocks in this chapter. In Section 3.4 and Section 3.5, we propose the anonymous reputation system and the efficient integration with a PoS blockchain. We analyze the security of the proposed ARS-PS in Section 3.6, and evaluate its performance in Section 3.7. Finally, we conclude this chapter in Section 3.8.

3.2 Problem Formulation

In this section, we formulate the system model, security model, and design goals of this chapter.

3.2.1 System Model

In Fig. 3.1, there are three entities in our system: consumers, retailers, and an identity management entity (IDM).

- ▷ **Consumer.** A consumer, uniquely identified by C_i, can make purchases from retailers and later leave a numeric rating score for the retailer.
- ▷ **Retailer.** A retailer, uniquely identified by R_j, can sell products to consumers and establish reputations from consumer feedbacks. Retailers also act as stakeholders and collaboratively maintain a public leger (denoted as $\mathcal{L}$) based on a PoS consensus protocol [77].
- ▷ **IDM.** IDM is a government agency that is in charge of issuing and managing identities and credentials of consumers and retailers.

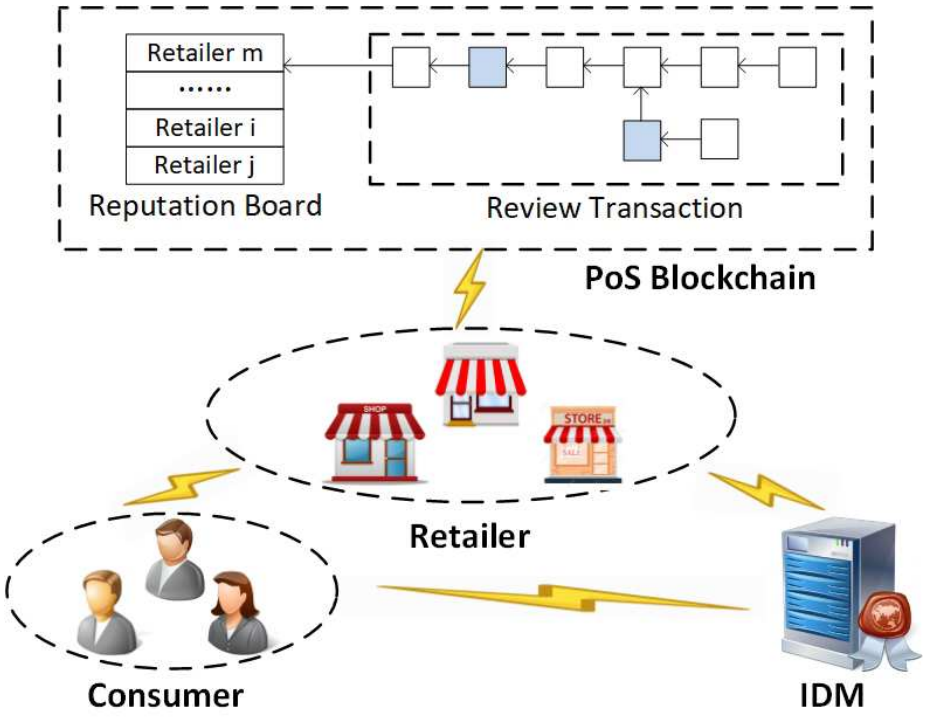


Figure 3.1: Anonymous Reputation System

At a high level, the ARS-PS works as follows. Consumers and retailers first register themselves to IDM. Each consumer obtains an anonymous identity credential from IDM. Afterwards, consumers can make purchases from retailers and obtain an anonymous rating token. Later, a consumer can leave a review (a rating score) for a retailer by making a review transaction to $\mathcal{L}$ and privately tie the review to a previous purchase. Finally, review transactions for the same retailer accumulate as a numeric score in the reputation board. Note that IDM in ARS-PS can be extended to a distributed identity management system [152].

3.2.2 Security Model

We assume IDM to be fully trusted. This is reasonable since the behavior of IDM is a government agency responsible for the administration of the citizens. Some consumers and retailers can be malicious and may launch a bunch of attacks to the system such as Sybil attacks, and whihte/bad mouthing attacks [37]. For the security of public leger $\mathcal{L}$, we borrow the assumptions from [77, 153]. In particular, the stake in the PoS consensus protocol is associated with the reputation of retailers in the ARS-PS. We require that an adversary cannot control the majority of the stake (reputation) in the system. Meanwhile, we assume that a rational retailer (stakeholder) with high reputation (stake) is more willing to maintain the correctness of the ledger $\mathcal{L}$. This is reasonable since the cost for a high-scored retailer to behave maliciously is huge [77].

3.2.3 Design Goals

Under the security assumptions, we summarize the design goals of the ARS-PS.

- ▷ Bounded Confidentiality. A consumer's individual review statistics (rating scores) should be kept private. Only the aggregated retailer review statistics is revealed to the public. However, individual rating scores should have upper and lower boundaries. Consumers cannot submit rating scores that exceed the boundaries.
- ▷ Conditional Anonymity. Obtaining a rating token or leaving a review on a public ledger will not expose a consumer's true identity. However, IDM should be able to recover the true identity of an anonymous review in case of consumer misbehavior.
- ▷ Unforgeability. The anonymous identity credential and rating token cannot be forged. Without the credential and the token, consumers cannot submit a valid review to the public ledger.

▷ Confined Unlinkability. The public cannot determine if two valid reviews for different retailers are from the same consumer. However, the reviews are linkable if a consumer leaves multiple reviews for the same retailer.

▷ Transparency. Review generation and reputation accumulation process should be transparent and publicly verifiable to all retailers and consumers.

▷ Blockchain Security. The public transaction ledger should be robust and on-chain transactions should be immutable.

3.3 Building Blocks

In this section, building blocks in this chapter are presented. We denote three cyclic groups $\mathbb{G}_1$, $\mathbb{G}_2$, and $\mathbb{G}_T$ with a prime order p and a Type III bilinear pairing $e : \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$. $g, h \in \mathbb{G}_1$ and $\tilde{g} \in \mathbb{G}_2$ are generators of $\mathbb{G}_1, \mathbb{G}_2$. $\mathcal{H}$ is a collision-resist hash function that maps a string of arbitrary length to $\mathbb{Z}_p$. We denote $u \in_R \mathbb{Z}_p$ as randomly choosing a number from $\mathbb{Z}_p$.

3.3.1 Zero-Knowledge Proof

Zero-knowledge proof technique enables one party (prover) to prove to another party (verifier) that she knows some secret s for a public verifiable relation without exposing the secrets. In this chapter, we use the notation [154] for proof statement in the discrete-logarithm setting [22]. A typical example can be written as follows.

$$\mathbf{PK}\{(r_1, r_2) : Y_1 = h^{r_1} g^{r_2} \wedge Y_2 = g^{r_1}\}. \tag{3.1}$$

$r_1, r_2 \in \mathbb{Z}_p$ are the secrets that need to be proven and $Y_1, Y_2, h, g \in \mathbb{G}_1$ are the public parameters. The above proof can be instantiated using sigma protocol with Fiat-Shamir heuristic [127] as follows.

1. The prover chooses two random numbers $k_1, k_2 \in_R \mathbb{Z}_p$ and computes commitments $T_1 = h^{k_1} g^{k_2}$ and $T_2 = g^{k_1}$.
2. The prover computes $c = \mathcal{H}(Y_1, Y_2, T_1, T_2)$ and $z_1 = k_1 + cr_1$, $z_2 = k_2 + cr_2$.
3. For a given proof T_1, T_2, z_1, z_2, the verifier computes $c = \mathcal{H}(Y_1, Y_2, T_1, T_2)$ and checks $T_1 \stackrel{?}{=} Y_1^{-c} h^{z_1} g^{z_2}$ and $T_2 \stackrel{?}{=} Y_2^{-c} g^{z_1}$. The verifier accepts the proof if all the conditions hold.

3.3.2 PS Signature

Proposed by David Pointcheval and Olivier Sanders [119], PS signature is a signature scheme with a short signature size. The secret parameter $\mathcal{S}$ for the signature scheme is x, y, where $x, y \in_R \mathbb{Z}_p$. The public parameters $\mathcal{P}$ is $(g, \tilde{g}, \tilde{X}, \tilde{Y})$, where $g \in \mathbb{G}_1, \tilde{g} \in \mathbb{G}_2$, and $\tilde{X} = \tilde{g}^x, \tilde{Y} = \tilde{g}^y$. PS signature can be utilized to sign on committed messages, and the signature of the committed message is randomizable. In the following, two detailed techniques that are used to construct anonymous identity credentials and rating tokens are presented.

Sign on Committed Messages

We define a function $\mathbf{SigCom}(T, \mathcal{P}, \mathcal{S}, u)$ that takes as input the commitment $T = g^m$ of a message $m \in \mathbb{Z}_p$, public/secret parameters $\mathcal{P}/\mathcal{S}$, and a random number $u \in_R \mathbb{Z}_p$. The function outputs σ as the PS signature of the message m as follows.

$$\sigma = (\sigma_1, \sigma_2) = (g^u, (g^x \cdot T^y)^u). \tag{3.2}$$

Prove Knowledge of a Signature

Suppose that we have a signature tuple $\sigma = (\sigma_1, \sigma_2)$ of a message m. The prover first chooses $t \in_R \mathbb{Z}_p$ to randomize the signature as $(\sigma'_1, \sigma'_2) = (\sigma_1^t, \sigma_2^t)$. Then, the prover needs to prove that:

$$\mathbf{PK}\{(m, \sigma) : \sigma \textit{ is a PS signature on } m\}. \tag{3.3}$$

In specific, the prover chooses $k \in_R \mathbb{Z}_p$ and computes $R = e(\sigma'_1, \tilde{Y})^k$. The prover then obtains a random challenge $c \in \mathbb{Z}_p$ using Fiat-Shamir heuristic and computes $s = k + c \cdot m$. Given $(\sigma'_1, \sigma'_2, c, s)$, a verifier can compute $R' = (e(\sigma'^{-1}_1, \tilde{X})e(\sigma'_2, \tilde{g}))^{-c} e(\sigma'^s_1, \tilde{Y})$ and checks if the random challenge c is correctly computed.

3.3.3 Bulletproof System

Bulletproof [130] is an efficient zero-knowledge proof system for range proof on committed values with compact proof size. An instance of bulletproof can be written as follows.

$$\mathbf{PK}\{(a, r) : Y = h^r g^a \wedge a \in [0, 2^n]\}. \tag{3.4}$$

$Y = h^r g^a$ is a Pedersen commitment of the integer $a \in \mathbb{Z}_p$ using randomness r. The above proof will convince the verifier that the secret in the commitment Y lies in the range $[0, 2^n]$. Bulletproof can be instantiated in the discrete logarithm setting and made non-interactive with Fiat-Shamir heuristic. We refer the readers to [130] for the detailed construction.

3.3.4 *Ouroboros* - A PoS Blockchain

Blockchain is a public ledger maintained by a peer-to-peer network that provides immutable and transparent list of transaction records [155]. It contains an increasing list of blocks of transactions shared by network peers. Network peers rely on consensus protocols to reach consistency on the shared public ledger. In this chapter, a state-of-art Proof-of-Stake (PoS) based blockchain *Ouroboros* [77] is adopted due to its efficiency and rigorous security guarantees. In the following, we summarize the concepts and design principles of *Ouroboros* [77].

- ▷ *Stakeholder.* A stakeholder is the miner in *Ouroboros*. Each stakeholder is assigned with a certain amount of stake and the amount of stake can change overtime.
- ▷ *Epoch/Slot.* An epoch is a set of equal time slots. The *Ouroboros* assumes global clock is divided into discrete epochs and all the stakeholders maintain a roughly synchronized clock.
- ▷ *Users.* Users are the participants of the blockchain network. Users can make transactions to transfer crypto currencies and change the state of the public ledger.
- ▷ *Block/Ledger.* A block is a collection of transactions. A sequence of blocks constitutes a leger.

In *Ouroboros*, a stateholder is elected as the slot leader for each time slot. The role of the slot leader is to collect transactions and issue only one block for the time slot. The core of the *Ouroboros* is a leader selection function that elects the slot leader proportionally to stakeholder's stake. That is, the more stake a stakeholder has, the more likely she will be elected as a slot leader.

3.4 Anonymous Reputation System

In this section, we propose an anonymous reputation system based on PS-signature [119], Bulletproof system [130] and non-interactive zero-knowledge proof technique. We assume

secure and authenticated channels are established among entities.

3.4.1 System Setup

Given λ as the system security parameter, IDM chooses a set of bilinear groups $\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T$ with a prime order p and an asymmetric bilinear pairing e [119]. $\mathcal{H} : \{0,1\}^* \rightarrow \mathbb{Z}_p$ denotes a hash function. IDM chooses non-identical generators $g_1, g_2 \in \mathbb{G}_1$ and $\tilde{g} \in \mathbb{G}_2$. IDM randomly chooses $\mathcal{S} = (x, y) \in_R \mathbb{Z}_p^2$ to compute the system public key $\tilde{X} = \tilde{g}^x$, $\tilde{Y} = \tilde{g}^y$. Finally, IDM sets the system public parameter $\mathcal{P}$ as follows:

$$\mathcal{P} = \left\{ \ \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, p, g_1, g_2, \tilde{g}, \tilde{X}, \tilde{Y}, \mathcal{H}, e \ \right\}. \tag{3.5}$$

3.4.2 Consumer Registration

A consumer C_i first registers herself at IDM using her true identity. After that, C_i interacts with IDM to obtain an anonymous identity credential as follows.

1. C_i chooses a secret $cs_i \in_R \mathbb{Z}_p$ and computes $(T_{i,1}, T_{i,2}) = (g_1^{cs_i}, \tilde{Y}^{cs_i})$. Then, C_i generates π_{cs_i}, a zero-knowledge proof of cs_i as follows.

$$\mathbf{PK}\{(cs_i) : T_{i,1} = g_1^{cs_i} \wedge T_{i,2} = \tilde{Y}^{cs_i}\}. \tag{3.6}$$

   C_i sends $(T_{i,1}, T_{i,2}, \pi_{cs_i})$ to IDM.

2. IDM first checks the validity of π_{cs_i} and $e(T_{i,1}, \tilde{Y}) \stackrel{?}{=} e(g_1, T_{i,2})$. If either of the equations does not hold, IDM aborts. Otherwise, IDM chooses $u \in_R \mathbb{Z}_p$ and computes a PS signature on the committed message $T_{i,1}$ for consumer C_i as follows.

$$\begin{aligned} \sigma_i &= \mathbf{SigCom}(T_{i,1}, \mathcal{P}, \mathcal{S}, u) \\ &= (\sigma_{i,1}, \sigma_{i,2}) = (g_1^u, (g_1^x \cdot T_{i,1}^y)^u). \end{aligned} \tag{3.7}$$

   IDM stores $(C_i, T_{i,1}, T_{i,2}, \sigma_i)$ and sends σ_i to C_i.

3. Upon receiving σ_i from IDM, C_i checks $\sigma_{i,1} \neq 1_{\mathbb{G}_1}$ and

$$e(\sigma_{i,1}, \tilde{X}\tilde{Y}^{cs_i}) \stackrel{?}{=} e(\sigma_{i,2}, \tilde{g}). \tag{3.8}$$

   If the equation holds, C_i stores (cs_i, σ_i) as her anonymous identity credential.

3.4.3 Retailer Registration

Retailers register themselves at IDM as follows.

1. A retailer R_j chooses $\tilde{g}_j \in_R \mathbb{G}_2$, $x_j, y_j, sk_j \in_R \mathbb{Z}_p^3$, and computes $\tilde{X}_j = \tilde{g}_j^{x_j}$, $\tilde{Y}_j = \tilde{g}_j^{y_j}, pk_j = g_2^{sk_j}$. The secret parameter of R_j is $\mathcal{S}_j = (x_j, y_j, sk_j)$, and the public parameter is $\mathcal{P}_j = (\tilde{g}_j, \tilde{X}_j, \tilde{Y}_j, pk_j)$.

2. Then, R_j generates a proof π_{R_j} as follows.

$$\mathbf{PK}\left\{ (x_j, y_j, sk_j) : \tilde{X}_j = \tilde{g}_j^{x_j} \wedge \tilde{Y}_j = \tilde{g}_j^{y_j} \wedge pk_j = g_2^{sk_j} \right\}. \tag{3.9}$$

   R_j sends its public key P_j and π_{R_j} to IDM.

3. IDM checks the validity of proof π_{R_j}. IDM aborts when the proof is invalid. Otherwise, IDM stores $(R_j, \mathcal{P}_j)$.

3.4.4 Rating Token Generation

Consumers can make purchases from retailers via anonymous payment channels, such as zerocash [96]. After making a purchase from R_j, C_i can obtain an anonymous rating token as follows.

1. C_i chooses $g_{i,j} \in_R \mathbb{G}_1$ and $t \in_R \mathbb{Z}_p$ to compute $(\sigma'_{i,1}, \sigma'_{i,2}) = (\sigma_{i,1}^t, \sigma_{i,2}^t)$, $Y = g_{i,j}^{-cs_i}$ using σ_i. C_i constructs a proof as follows.

$$\mathbf{PK}\left\{ \begin{array}{c} (cs_i, \sigma_i) : \\ \sigma_i \textit{ is a PS signature on } cs_i \wedge \\ Y = g_{i,j}^{-cs_i} \end{array} \right\}. \tag{3.10}$$

2. In specific, C_i chooses $k \in_R \mathbb{Z}_p$ and computes:

$$\begin{aligned} R &= e(\sigma'_{i,1}, \tilde{Y})^k = e(\sigma_{i,1}, \tilde{Y})^{kt}, \\ T &= g_{i,j}^k, \\ c &= \mathcal{H}(\sigma'_{i,1}, \sigma'_{i,2}, R, Y, T, g_{i,j}), \\ s &= k + c \cdot cs_i. \end{aligned} \tag{3.11}$$

   The proof is the combination of the general pre-image zero-knowledge technique with the proof-of-knowledge of signature technique by re-using the response s. C_i sends $(\sigma'_{i,1}, \sigma'_{i,2}, Y, g_{i,j}, c, s)$ to R_j.

3. R_j computes R', T' and checks:

$$\begin{aligned} R' &= (e(\sigma'^{-1}_{i,1}, \tilde{X})e(\sigma'_{i,2}, \tilde{g}))^{-c} e(\sigma'^{s}_{i,1}, \tilde{Y}), \\ T' &= Y^c g^s_{i,j}, \\ c &\stackrel{?}{=} \mathcal{H}(\sigma'_{i,1}, \sigma'_{i,2}, R', Y, T', g_{i,j}). \end{aligned} \tag{3.12}$$

If the equation holds, R_j will generate an anonymous rating token $\sigma_{i,j}$ for C_i using x_j, y_j:

$$\begin{aligned} \sigma_{i,j} &= \mathbf{SigCom}(Y, \mathcal{P}_j, \mathcal{S}_j, u') \\ &= (\sigma_{i,j,1}, \sigma_{i,j,2}) = (g^{u'}_{i,j}, (g^{x_j}_{i,j} \cdot Y^{-y_j})^{u'}). \end{aligned} \tag{3.13}$$

where $u' \in_R \mathbb{Z}_p$. R_j sends the anonymous rating token $\sigma_{i,j}$ to C_i via a secure channel.

4. Upon receiving $\sigma_{i,j}$, C_i checks $\sigma_{i,j,1} \neq 1_{\mathbb{G}_1}$ and

$$e(\sigma_{i,j,1}, \tilde{X}_j \tilde{Y}^{cs_i}_j) \stackrel{?}{=} e(\sigma_{i,j,2}, \tilde{g}_j). \tag{3.14}$$

If the equation holds, C_i stores $\sigma_{i,j}$ as her rating token for retailer R_j.

3.4.5 Anonymous Review Generation and Verification

IDM chooses a set of retailers to form a committee $\mathcal{L}_C$. A consumer C_i can leave a rating score for the retailer R_j using the rating token $\sigma_{i,j}$ and the identity credential σ_i as follows.

1. C_i chooses a rating score $s_{i,j}$, where $s_{i,j} \in \mathbb{Z}_p$ can be an integer in $[1, 10]$. C_i obtains the public keys pk_j of all the committee members and computes $pk_C = \prod_{R_j \in \mathcal{L}_C} pk_j$. C_i chooses $r \in_R \mathbb{Z}_p$ and encrypts $s_{i,j}$ as follows.

$$r_{i,j} = (r_{i,j,1}, r_{i,j,2}) = (pk^r_C g_2^{s_{i,j}}, g^r_2). \tag{3.15}$$

C_i constructs a proof $\pi_{i,j}$ to prove that $r_{i,j}$ is a valid encryption of $s_{i,j}$ that lies in $[1, 10]$:

$$\mathbf{PK} \left\{ \begin{array}{c} (s_{i,j}, r) : r_{i,j,1} = pk^r_C g_2^{s_{i,j}} \wedge \\ r_{i,j,2} = g^r_2 \wedge s_{i,j} \in [1, 10] \end{array} \right\}. \tag{3.16}$$

The above proof can be instantiated via sigma protocol and bulletproof system.

2. C_i chooses random numbers $r_1, r_2 \in \mathbb{Z}_p$ and computes:

$$\begin{aligned} &\beta_1 = \sigma_{i,1}^{r_1}, \quad \beta_2 = \sigma_{i,2}^{r_1}, \quad \beta_3 = \sigma_{i,j,1}^{r_2}, \\ &\beta_4 = \sigma_{i,j,2}^{r_2}, \quad \beta_5 = g_1^{\mathcal{H}(R_j)cs_i}. \end{aligned} \tag{3.17}$$

C_i needs to prove the knowledge of a valid rating token and an identity credential by constructing the proof as follows.

$$\mathbf{PK} \left\{ \begin{array}{c} (cs_i, \sigma_i, \sigma_{i,j}) : \\ \sigma_i, \sigma_{i,j} \ are \ PS \ signatures \ on \ cs_i \wedge \\ \beta_5 = g_1^{\mathcal{H}(R_j)cs_i} \end{array} \right\}. \tag{3.18}$$

3. In specific, C_i chooses a random number $k_{ep} \in \mathbb{Z}_p$ and computes:

$$\begin{aligned} &\alpha_1 = e(\beta_1, \tilde{Y})^{k_{ep}}, \quad \alpha_2 = e(\beta_3, \tilde{Y}_j)^{k_{ep}}, \\ &\alpha_3 = g_1^{\mathcal{H}(R_j)k_{ep}}, \\ &ch = \mathcal{H}(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \alpha_1, \alpha_2, \alpha_3, R_j, r_{i,j}, \pi_{i,j}), \\ &s_i = k_{ep} + ch \cdot cs_i. \end{aligned} \tag{3.19}$$

C_i sets $\sigma = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, ch, s_i)$ and sends the anonymous review $(\sigma, r_{i,j}, \pi_{i,j}, R_j)$ to the committee members.

4. Upon receiving the ratings from C_i, the committee members check the validity of the anonymous review. The committee members first compute the following equations using system public parameters $\mathcal{P}$ and retailer R_j's public key $\mathcal{P}_j$.

$$\begin{aligned} &\alpha_1' = e(\beta_1, \tilde{X})^{ch} e(\beta_2, \tilde{g})^{-ch} e(\beta_1, \tilde{Y})^{s_i}, \\ &\alpha_2' = e(\beta_3, \tilde{X}_j)^{ch} e(\beta_4, \tilde{g}_j)^{-ch} e(\beta_3, \tilde{Y}_j)^{s_i}, \\ &\alpha_3' = \beta_5^{-ch} \cdot g_1^{\mathcal{H}(R_j)s_i}. \end{aligned} \tag{3.20}$$

The committee members check the validity of proof $\pi_{i,j}$ as specified in [130] and whether $ch \stackrel{?}{=} \mathcal{H}(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \alpha_1', \alpha_2', \alpha_3', R_j, r_{i,j}, \pi_{i,j})$. If both of the conditions hold, the committee members accept the anonymous review.

3.4.6 Review Aggregation

Committee members aggregate the valid encrypted rating scores for each retailer. For retailer R_j, committee members compute $s_j = (s_{j,1}, s_{j,2}) = (\prod r_{i,j,1}, \prod r_{i,j,2})$ for all the

valid encrypted rating scores $r_{i,j}$. For retailer R_j, a committee member C_m computes a partial decryption token $p_{j,m} = s_{j,2}^{sk_m}$, where sk_m is the secret key of C_m. The committee member constructs a proof $\pi_{j,m}$ that the partial decryption token is correctly constructed as follows.

$$\mathbf{PK}\{(sk_m) : p_{j,m} = s_{j,2}^{sk_m} \wedge pk_m = g_2^{sk_m}\}. \tag{3.21}$$

The final decryption $\mathcal{RS}_j$ of the aggregated rating score for retailer R_j should be:

$$\mathcal{RS}_j = \frac{s_{j,1}}{\prod_{C_m \in \mathcal{L}_C} p_{j,m}} = g_2^{\sum s_{i,j}}. \tag{3.22}$$

It should be noted that the final aggregated rating score $\sum s_{i,j}$ is at the exponent of g_2. All retailers and consumers can efficiently pre-compute a table that contains g_2^l, where l can range from 0 to a few thousands.

3.4.7 Linking and Tracing

For all the valid reviews, committee members will check if there exist the same β_5. If committee members find the same β_5 from different reviews, it indicates that a consumer submitted multiple reviews for the same purchase. The committee members will report the anonymous review of the misbehaving consumer to IDM. To recover the true identity of the misbehaving consumer, IDM checks the following equation for each $(T_{i,1}, T_{i,2})$ stored in its storage:

$$e(\beta_2, \tilde{g}) \cdot e(\beta_1, \tilde{X})^{-1} \stackrel{?}{=} e(\beta_1, T_{i,2}). \tag{3.23}$$

IDM publishes $T_{i,1}$ and $T_{i,2}$ that matches the above equation as the misbehaving consumer.

In this section, we propose a reputation system that enables consumers privately make purchases and leave reviews. In the next section, we will present the details on implementing the proposed system on a PoS blockchain to improve system transparency and reliability.

3.5 Anonymous Reputation System atop PoS Blockchain

In this section, we integrate our anonymous reputation system atop a PoS blockchain - *Ouroboros* [77]. The operations proposed in the previous section 3.4 are classified into

two categories: on-chain and off-chain operations. The off-chain operations include consumer/retailer registration and rating token generation that require interactions between IDM, retailers, and consumers via secure channels.

Review generation, verification, and aggregation are on-chain operations that happen over a public ledger $\mathcal{L}$. We adopt a hybrid blockchain model in the ARS-PS. Retailers act as stakeholders based on the PoS protocol in *Ouroboros* with their reputations associated with the stake. Retailers need to obtain permissions from the IDM before they can serve as stakeholders. Consumers act as blockchain users who can freely join the blockchain network. Consumers can leave reviews and enumerate accumulated retailers' reputation scores by making different types of transactions to the ledger. The reasons that we adopt *Ouroboros* are threefold.

- ▷ A PoS blockchain is more suitable for constructing a consortium network.
- ▷ A PoS blockchain offers qualitative efficiency and scalability compared with a PoW blockchain.
- ▷ Committee member management in the ARS-PS can be realized via the consensus protocol in [77].

The blockchain-based anonymous reputation system consists of the following steps. Notations from Section 3.5 are re-used.

3.5.1 Genesis Block Generation

IDM runs the *System Setup* of Section 3.5, generates and publishes the system parameters $\mathcal{P}$. Consumers and retailers can obtain $\mathcal{P}$ via secure channels, such as TLS. IDM also defines $T_{\mathcal{A}}$ as the size of the anonymity set, which indicates the privacy level of the system. Retailers interact with the IDM to register their public keys $\mathcal{P}_j$. IDM creates a global reputation board $\mathcal{B}$ that contains the global reputation scores $\mathcal{RS}_j$ for each retailer. Consumers register themselves at IDM to obtain anonymous identity credentials σ_i. Both retailers and consumers can join the blockchain network to obtain their blockchain accounts with a public/private key pair to sign on the transactions. Retailer blockchain account information is publicly associated with their identities, while consumer blockchain accounts remain anonymous.

IDM sets the global clock of the system and divides the clock into epochs of equal time slots. Each epoch is divided into three stages: *Accumulation*, *Aggregation* and *Revelation*.

The number of time slots for each stage is $\mathcal{K}_1, \mathcal{K}_2, \mathcal{K}_3$, respectively. At the beginning of each epoch, IDM runs a committee selection function [77] to select a committee of retailers with high reputation scores, which is responsible for the slot leader selection and review revelation process. Afterwards, IDM generates a genesis block of the ledger $\mathcal{L}$ consisting of system parameters $\mathcal{P}$, retailer parameters $\mathcal{P}_j$, retailers blockchain account information, and the list of committee members $\mathcal{L}_C$ in this epoch. Committee members run a leader selection function [77] to select slot leaders for time slots in this epoch.

3.5.2 Review Accumulation

For each registered retailer R_j, IDM creates a review smart contract SC_j. The smart contract SC_j records the reviews for the retailer R_j. In particular, the contract SC_j has two functions *Update* and *GetReview*. The *Update* function takes into the anonymous reviews from consumers. The anonymous reviews can later be accessed by the *GetReview* function. In specific, consumer C_i can make purchases from retailer R_j in an off-chain manner and obtain a valid rating token $\sigma_{i,j}$. C_i can generate an anonymous review transaction T_r including the anonymous review $(\sigma, r_{i,j}, \pi_{i,j}, R_j)$ to the smart contract SC_j by calling the *Update* function. The smart contract SC_j records the anonymous review in its storage for future reputation aggregation and revelation.

3.5.3 Review Aggregation

In the *Aggregation* stage, each slot leader is responsible for the review aggregation task of $1/\mathcal{K}_2$ of overall retailers. Slot leaders aggregate the encrypted reviews for each retailer in the following steps.

- ▷ A slot leader queries the current state of contracts SC_j in her management scope. The slot leader will report double-reviews for the same retailer to the IDM to recover the true identity of the misbehaving consumer.
- ▷ For retailer R_j, the slot leader checks the number of valid received reviews. If the number exceeds $T_{\mathcal{A}}$, the slot leader aggregates the valid encrypted rating scores to obtain an aggregated rating score s_j.
- ▷ The slot leader constructs a reveal smart contract $\mathcal{R}$. The contract $\mathcal{R}$ includes the aggregated rating scores for retailers in her management scope with a counter C_{R_j} that records the number of reviews received for the retailer. The reveal contract also

provides a function *UpdateToken* to receive partial decryption tokens from committee members.

After all the slot leaders in this stage publish the $\mathcal{R}$ contracts, the system proceeds to the final *Revelation* stage.

3.5.4 Review Revelation

In the *Revelation* stage, committee members first check the reveal contracts $\mathcal{R}$ generated from the previous stage. For the aggregated rating scores, committee members update their partial decryption tokens to the reveal contracts using the *UpdateToken* function. After obtaining all the partial decryption tokens for the reveal contracts, IDM verifies the correctness of the partial decryption tokens and decrypts the aggregated scores using Equation 3.22. Finally, IDM updates the reputation scores in the global reputation board for retailers.

3.5.5 Epoch Update

For the next epoch, retailers interact with IDM to generate a new set of retailer public keys $\mathcal{P}_j$ for each retailer R_j. IDM runs the committee selection function for the new epoch. New committee members then run the leader selection function for this epoch according to the updated global reputation scores. For the encrypted reviews that are not aggregated in the previous epoch, consumers generate new review transactions with the updated committee encryption parameters.

3.6 Security Analysis

In this section, we give the security analysis of the proposed ARS-PS based on the design goals.

3.6.1 Bounded Confidentiality

Consumers encrypt their rating scores with committee members' public keys. Committee members will check the validity of the reveal contracts and only publish their partial

decryption tokens for the valid aggregated rating scores. That is, an adversary can obtain the individual review statistics only if he can solve the **DDH** problem in $\mathbb{G}_1$ [156] or he can control the whole committee members to recover the decryption key. At the same time, consumers need to prove that the encrypted rating scores lie in a correct range. Due to the *Soundness* and *Completeness* property of Bulletproof [130], the verifier will accept the range proof if it is correctly constructed. That is, the bounded confidentiality is preserved in our system.

3.6.2 Conditional Anonymity

The consumer C_i first registers herself at IDM to obtain an anonymous identity credential σ_i. To obtain an anonymous rating token, consumer C_i chooses a random generator $g_{i,j}$ for each purchase and proves to the retailer that the committed message $Y = g_{i,j}^{-cs_i}$ contains the same consumer secret with the identity token in a zero-knowledge manner. Then, retailers can sign on the committed message $Y = g_{i,j}^{-cs_i}$. When leaving an anonymous review, C_i needs to prove the knowledge of a valid rating token and an anonymous identity credential using the sigma protocol [127]. Thus, the anonymity of obtaining a rating token and leaving a review can be reduced to the *Zero-knowledge* property of the underlying sigma protocol in the discrete logarithm setting. When a consumer misbehavior is detected, slot leaders report the anonymous reviews to IDM to recover the identity of the consumer. Retailers cannot recover the identity of a consumer since consumers do not generate the $\tilde{Y}_j^{cs_i}$ when obtaining the rating token. That is, conditional anonymity is preserved in the ARS-PS.

3.6.3 Unforgeability

To generate the anonymous identity credential, IDM needs to sign on the committed message $g_1^{cs_i}$ using PS signature. Similarly, the retailer needs to sign on the committed message $g_{i,j}^{cs_i}$ to generate a rating token for consumer C_i. That is, the unforgeability of the identity credential and rating token can be reduced to the unforgeability of the PS signature [119], which can be further reduced to ***q-MSDH-1*** assumption in the non-interactive setting [119]. To generate the anonymous review σ and $\pi_{i,j}$, the consumer needs to prove the knowledge of an identity credential and a rating token at the same time. Thus, the consumer cannot forge the anonymous review if the underlying sigma protocol [127] is sound.

3.6.4 Confined Unlinkability

The unlinkability requires that retailers and consumers cannot determine if two reviews are from the same consumer. This property comes from two folds. First, a consumer can choose different random generators to require a rating token. Second, the consumer can further randomize the rating token by choosing a random number r_2 when generating an anonymous review and prove the knowledge of consumer secret in a zero-knowledge manner. That is, the unlinkability can be reduced to the security of underlying sigma protocol. When generating a review, C_i needs to construct β_5 and prove to the public that β_5 contains the same secret cs_i with $\beta_1, \beta_2, \beta_3, \beta_4$. If C_i leaves multiple reviews for the same retailer, the β_5 in the anonymous review is publicly identical. The combination of conditional anonymity and confined unlinkability helps the system mitigate Sybil attacks.

3.6.5 Transparency

The review accumulation, aggregation and revelation are implemented by the review and reveal contracts on the public ledger. Consumers can make review transactions to change or query the state of the contracts. Since the transactions and ledger state changes are open to the public's view, transparency of reputation system is guaranteed [107].

3.6.6 Blockchain Security

As a public transaction ledger, the blockchain security is formally defined as **Persistence** and **Liveness** [77]. Specifically, we borrow the definitions from [77]. **Persistence** preserves the stability of the public ledger. **Liveness** means that a valid transaction is guaranteed to be included in the ledger after a certain time. If the adversary cannot control the most stakes in the system, *Ouroboros* is proven to achieve the above properties [77]. The ledger is maintained by registered retailers and the retailer's reputation in our system is associated with the stake in the PoS consensus protocol of *Ouroboros*. A retailer with a higher reputation score is less likely to behave distrustfully since the cost for the misbehavior is expensive. As a result, the public transaction ledger is robust in the ARS-PS. We then discuss the security of the review and reveal contracts.

In the *Accumulation* stage, consumers make transactions to the review contracts. Based on the ledger robustness, the transactions will finally be confirmed after certain number of slots with a high probability. Prorogation delays could happen such that some reviews

may not be included on the ledger in this epoch. In this case, consumers can update their reviews in the next epoch.

In the *Aggregation* phase, slot leaders verify the correctness of the reviews and aggregate the encrypted rating scores. That is, the security in this stage (i.e. the correctness of the aggregated rating scores) depends highly on the trustworthiness of the slot leaders. If a slot leader does not fulfill his task (e.g. aggregate incorrect reviews or purposely exclude some reviews), his misbehavior may not be discovered immediately. However, since the historical reviews and aggregated rating scores are open to the public, anyone in the system can check the correctness in the future and makes a complaint if the misbehavior of a slot leader is detected. By properly setting the punishment for misbehaving slot leaders, a rationale slot leader is motivated to correctly fulfill the task. Moreover, blockchain accounts of consumers remain anonymous in the ARS-PS. A malicious consumer may generate a large number of invalid reviews to use up the slot leader's computational capacities. To prevent this attack, the review contracts can require consumers to deposit currencies to the contract and only returns the currencies to the consumer when the review is verified. Secure and anonymous payment channels (such as zerocash [96]) can be utilized to preserve consumer anonymity and unlinkability in this process.

In the *Revelation* stage, committee members verify the correctness of reveal contracts and update their partial decryption tokens to the reveal contract. The correctness of the tokens is ensured by the zero-knowledge proof $\pi_{j,m}$. The public cannot decrypt the aggregated rating scores unless all the committee members have successfully submitted their tokens to the ledger. Compared with communication overhead in the *Accumulation* stage, only finite transactions are required in this stage. To mitigate the impact of communication delay among committee members, we can set a larger number of $\mathcal{K}_3$ to ensure the ledger robustness at this stage. For the committee member that fail to submit the token, IDM can directly contact the committee member. We can also implement a threshold encryption scheme [157] to improve system robustness.

3.7 Performance Evaluation

In this section, we evaluate the performance of the proposed ARS-PS. We first compare the ARS-PS with existing schemes in terms of functionalities. Then, we present a proof-of-concept implementation based on Parity Ethereum, and demonstrate the implementation feasibility. Finally, we discuss the scalability of the ARS-PS.

3.7.1 Functionality

Table 3.1: Overview of Functionalities

Proposal	Architecture	C-Anonymity	B-Confidentiality	C-Unlinkability	Transparency
Blomer[37]	Centralized	√	√	√	
Zhai[69]	Decentralized	√	√	√	
Azad[71]	Decentralized		√		√
Schaub[114]	Blockchain	√		√	√
Soska[36]	Blockchain	√		√	√
ARS-PS	Blockchain	√	√	√	√

In Table 3.1, we summarize the recent advances in reputation systems in terms of architectures and desired functionalities. C-Anonymity means conditional anonymity, B-Confidentiality means bounded confidentiality and C-Unlinkability means confined unlinkability. Compared with a centralized architecture [37], a decentralized architecture [69, 71] is preferred for its advantage in eliminating a single trusted marketplace. Blockchain-based solutions [114, 36] and the ARS-PS further increase system transparency. As we discussed in the security analysis section, versatile functionalities are achieved in the ARS-PS by integrating a PoS blockchain with a set of cryptographic primitives.

3.7.2 Implementation Overview

We present a proof-of-concept implementation of the ARS-PS as shown in Fig. 3.2. We simulate IDM, consumer, and retailer with JAVA clients on a laptop with 2.40 GHz Intel Core i5 processors and 8 GB memory. We implement an MNT curve with an embedding degree 6 based on Java Pairing based Cryptography (JPBC) [158]. We instantiate Bulletproof system with a range of 3 bit with the implementation of the linear size arguments.

We set up a testing Ethereum Proof of Authority (PoA) blockchain network [159]. In particular, two kinds of Parity nodes are implemented in Parity PoA network.

- ▷ Authority nodes serve as retailers that can be selected as slot leaders to validate transactions and issue blocks.
- ▷ User nodes serve as consumers that can make anonymous review transactions to the blockchain.

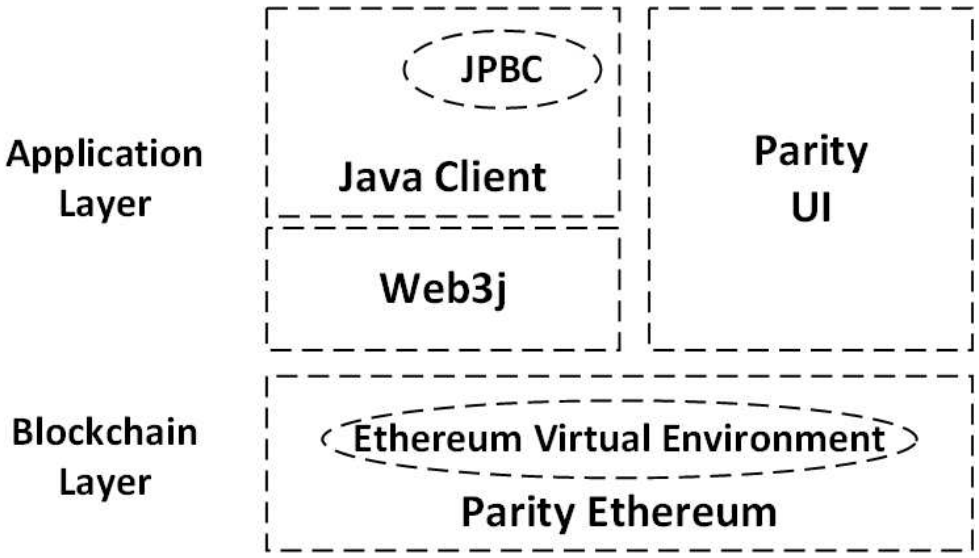


Figure 3.2: Implementation Overview

For illustrative purposes, a few authority nodes and user nodes are deployed in our experiments. Slot leaders are statically specified and written as configurations in the chain specification file. We increase the block gas limit in our testing network for storing the reviews. JAVA clients communicate with the associated Parity nodes via web3j [160] to send transactions and interact with smart contracts. Moreover, we encode the public parameters of the system and authority nodes into Java clients. A review smart contract written in Solidity [161] is deployed via Parity UI, that provides an *Update* function and a *GetReview* function.

We evaluate the system efficiency in terms of on-chain and off-chain performance. On-chain operations denote the review transaction generation/verification. Off-chain operations denote the registration and token generation phases.

3.7.3 Off-chain Performance

We evaluate the off-chain performance including consumer/retailer registration, rating token generation among entities. In Table 3.2, experimental results show that the computation incurs a few milliseconds.

Table 3.2: Off-chain Overhead

Operations	Involved Entities	Time (ms)
Consumer Registration	Consumer/IDM	487
Retailer Registration	Retailer/IDM	263
Rating Token Generation	Consumer/Retailer	259

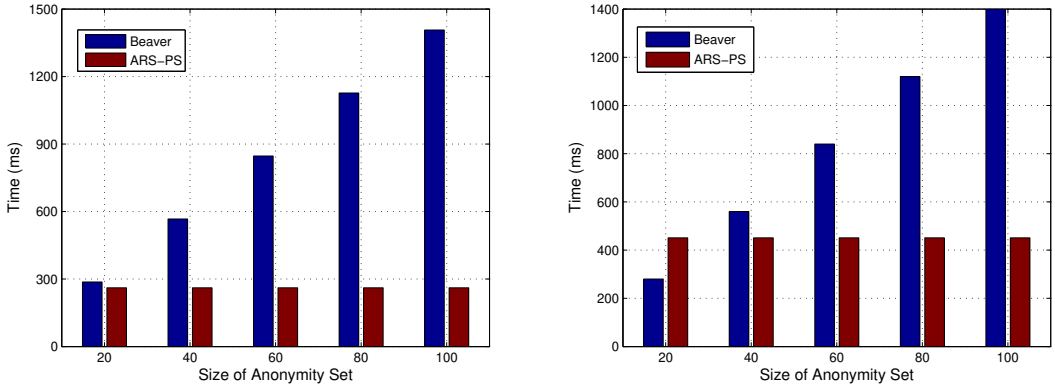
3.7.4 On-chain Performance

We simulate an epoch of the ARS-PS. In particular, consumers with rating tokens and identity credentials leave anonymous reviews by calling the *Update* function in the review contract. Then, the slot leader retrieves all the reviews from the review contract and verifies the correctness of the proofs. The slot leader creates another reveal contract $\mathcal{R}$ that aggregates the encrypted rating scores of valid reviews and receives partial decryption tokens from committee members.

We move the on-chain proof verifications to be conducted by the slot leader out of the EVM. In Table 3.3, we show the computational cost of generating and verifying an anonymous review. We further compare the ARS-PS with another blockchain-based literature that is based on ring signature [36] for review generation/verification. A ring-signature based method [36] requires purchase transactions to be also deployed on the public ledger. Consumers collect a set of public keys of previous purchase transactions (anonymity set $T_{\mathcal{A}}$) to generate/verify the anonymous reviews, which results in linearly increasing computational cost as shown in Fig. 3.3(a) and 3.3(b). The review generation/verification may consume a few hundred milliseconds in the ARS-PS. The reasons are twofold: (1) The proof σ consists of an identity proof and rating token proof to achieve conditional anonymity, which results in a double proof of knowledge of PS signature; and (2) paring operations over an MNT curve are expensive in the implemented JPBC library without PBC wrapper.

Table 3.3: Review Generation/Verification

	Rating Score	Proof σ	Proof $\pi_{i,j}$
Generation (ms)	15	183	63
Verification (ms)	N/A	347	110
Size (Bytes)	104	306	565



(a) Review Generation (b) Review Verification

Figure 3.3: Review Computation Cost

3.7.5 Scalability Discussions

In the following, we discuss the system scalability for different stages in one epoch. We define N_C as the number of committee members for the epoch.

Accumulation Stage

In our testing PoA blockchain with optimal network conditions, a consumer that calls the *Update* function will have her review transaction included in the ledger within a few blocks. In real-world implementations [77], the communication delays between consumers and slot leaders may lead to the exclusion of a certain transaction in the epoch. To mitigate this issue, we can increase the number of slots $\mathcal{K}_1$ in this stage and the number of peer connections for the consumer Parity node.

Aggregation Stage

Slot leaders in this stage verify and aggregate the anonymous reviews. The performance is mainly affected by two factors: the number of time slots $\mathcal{K}_2$ and the size of the anonymity set $\mathcal{T}_A$. A larger $\mathcal{K}_2$ reduces the individual computation overhead for slot leaders while increasing the overall epoch time. The quantity of $\mathcal{T}_A$ indicates privacy guarantees for consumers. However, a larger $\mathcal{T}_A$ could also increase the probability that insufficient number of

reviews are received for aggregation in this epoch, which requires consumers to regenerate the reviews in the next epoch.

Revelation Stage

Committee members upload their partial decryption tokens to the reveal contract. The total number of transactions in this stage is $N_C * \mathcal{K}_2$. IDM can choose different N_C for the trade-off between system security strength and efficiency. To further improve the reveal efficiency and prevent decryption failure in case that a committee member does not update her decryption token, a threshold ElGamal encryption system can be adopted [157]. We can also partition the committee into different subgroups to separately manage the review decryption key.

3.8 Summary

In this chapter, we have investigated the privacy and transparency issues in current reputation systems for the IIoT-enabled retail marketing. We have developed an anonymous reputation system that provides a high privacy guarantees for consumers, which can also be efficiently and securely integrated with a PoS blockchain. We have implemented a proof-of-concept prototype system based on Ethereum and the experimental results have demonstrated the feasibility of our proposed system compared with state-of-the-art literature.

Chapter 4

Transparent and Accountable Vehicular Local Advertising with Practical Blockchain Designs

4.1 Background

Vehicular local advertising is a prevalent advertising strategy, that allows roadside retailers (e.g. restaurants or retail stores) to promote targeted advertisements (ads) to nearby vehicular users of related interests. Vehicular users with Global Positioning System (GPS)-enabled devices (e.g. mobile phones) can receive timely and location-aware ads [162] through various vehicular communication channels [163, 164]. For example, vehicular users can find restaurants near their current locations from Google Map, or receive coupons and flyers from Facebook when they drive in a shopping center, which greatly increases their travel efficiency and experience. At the same time, vehicular local advertising also boosts in-store visits and promotes business for roadside retailers. According to statistics [165], an average of 80 percent of local searches convert and 28 percent of local searches lead to an in-store purchase. Therefore, vehicular local advertising is deemed as one of the most promising applications in intelligent transportation systems and has attracted extensive research and practice efforts.

In the vehicular local advertising, retailers rely on advertising brokers to manage their ad disseminations. The third-party advertising model has been prevalent due to the following reasons: (1) It is cost-prohibitive for retailers to manage their advertising network, especially for small-business owners without sufficient IT and marketing expertise. (2)

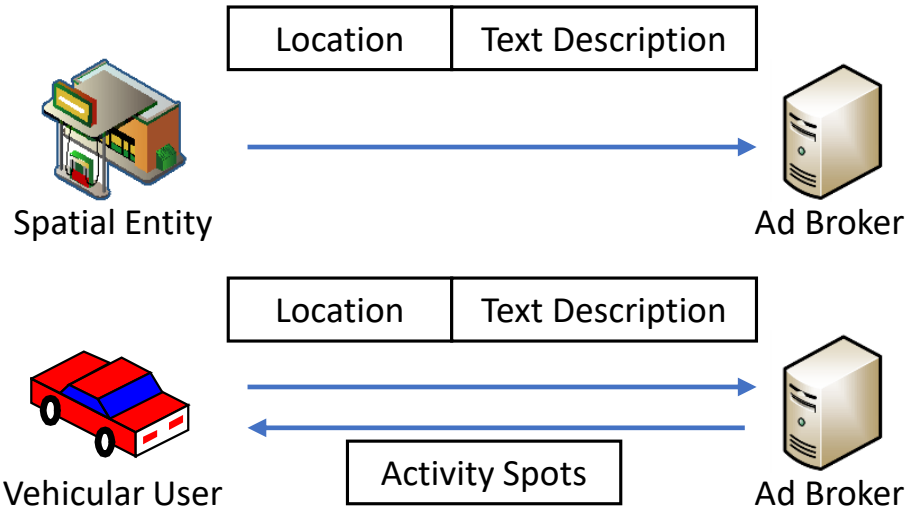


Figure 4.1: Vehicular Local Advertising

Ad brokers can provide retailers with more effective ad dissemination services from their huge user base and extraordinary computing platforms. In the literature, the ad dissemination process of the vehicular local advertising can be modeled as *spatial keyword query* [166, 167, 168], that considers both the spatial description (e.g., the retailer location and vehicle trajectory information) and textual description (e.g., retailer/user-specified keywords). In Fig. 4.1, retailers can specify their locations and a textual description of their business. Retailers then update the location and textual descriptions to the ad broker, e.g., Google, that can send advertisements to vehicular users. Vehicular users can find nearby activity spots by sending a location with textual descriptions of their interests to the ad broker.

The vehicular local advertising system has raised concerns and controversies recently on the lack of system *transparency* and *accountability* [33, 55]. *Transparency* requires the public visibility of the spatial keyword query process, which answers why a vehicular user receives the advertisements of specific spatial entities. It could result in the increasing popularity of ad blockers [55] at the vehicular users if without proper transparency guarantees. *Accountability* refers to two progressive meanings [100]. The first is the public detection of any breach of spatial keyword query protocols, such as the detection of *ad-fraud* attacks [34]. The second is the enforcement of obligations on the misbehaving advertisers or the ad broker. For example, a retailer that sends ads with the discrimination or misinformation should be suspended from the advertising service. While the ad broker is working towards the increase of system transparency and accountability, it is still far from sufficient in the

view of the public [33, 42].

The emerging blockchain technology [81] promises to serve as a public infrastructure to enhance the current vehicular local advertising system [30, 169]. Essentially, a (public) blockchain is a secure and distributed database, that is maintained by an open and peer-to-peer network with mutually distrusted nodes. As a result, it is tempting to migrate the vehicular local advertising system onto the blockchain, such that the spatial keyword query process can be open and verifiable to the public. However, the design and implementation of a blockchain-based vehicular local advertising system faces non-trivial obstacles in terms of the execution efficiency of the spatial keyword query. First, the large number of spatial objects with location and keyword information will result in a large-scale spatial keyword database. Second, the on-chain implementation cost in terms of the storage and computing is prohibitively expensive, since all blockchain nodes must maintain a local copy and verify the correctness of each transaction. As a result, a strawman solution that stores the large-scale spatial keyword database and executes the spatial keyword query on the blockchain is at the doubt of real-world practices. This motivates the main objective of this work: to take advantage of the principles of the blockchain technology for solving a real-world application with the practical designs. In particular, we aim at *practical* designs and implementations of a large-scale vehicular local advertising system that realizes the *transparency* and *accountability* promises of the blockchain technology.

In this chapter, we exploit the tamper-proof and open nature of the blockchain technology and design a blockchain-based Transparent and Accountable Vehicular Local Advertising system (*TAVLA*). From the state-of-the-art cryptographic building blocks, including the verifiable computation (VC) and spatial indexing techniques, *TAVLA* achieves the verifiable spatial keyword query process for the vehicular local advertising. *Verifiable* means that the spatial keyword query process is transparent to the public and any breach of spatial keyword query protocols can be effectively detected. Moreover, we address the scalability and efficiency issue of the blockchain-based advertising system with two design strategies, *digest-and-verify* and *divide-then-assemble*. The contributions of this chapter are as follows:

- ▷ We develop a *verifiable* spatial keyword query scheme with a cryptographic query index, *SKD*-tree, that prunes the spatial keyword search space. Moreover, we construct a *TAVLA* smart contract, that realizes the *SKD*-tree on Ethereum and ensures sufficient public transparency and accountability of the spatial keyword query process.
- ▷ We introduce two design strategies for a practical blockchain-based vehicular local advertising system: *digest-and-verify* and *divide-then-assemble*. The large-scale spa-

tial keyword database is digested and updated onto the blockchain with succinct cryptographic authenticators. The verifiable spatial keyword query scheme is then realized via modular executions of two off-chain verifiable functions. The results are assembled and verified in the *TAVLA* smart contract using the authenticators with limited on-chain storage and computation overheads.

▷ We conduct thorough security analysis to demonstrate that *TAVLA* achieves *Auditing Security* in terms of integrate, correct, transparent and accountable spatial keyword queries. We conduct experiments with the Pinocchio VC framework, which demonstrates that *TAVLA* is practical for implementations.

The organization of this chapter is as follows. In Section 4.2, we revisit the state-of-the-art spatial keyword query technique on road networks. Following definitions in Section 4.2, we formulate the system model of *TAVLA* with formal security model and design goals in Section 4.3. In Section 4.4, we summarize the building blocks of *TAVLA*, including spatial indexing, verifiable computation, and smart contract techniques. In Section 4.5, we propose *TAVLA* in terms of an off-chain spatial keyword query scheme and an on-chain advertising contract. We present the security analysis and performance evaluation in Section 4.6 and 4.7, respectively. Finally, we conclude this chapter in Section 4.8.

4.2 Preliminaries

In this section, we revisit the state-of-the-art *non-verifiable* spatial keyword query technique. First, we introduce the vector space-based probabilistic topic model (**Definition 1**). Second, we formalize the definitions of spatial objects (**Definition 2**) and spatial keyword queries (**Definition 3**).

Definition 1. Probabilistic topic model [170] is a natural language processing mechanism that translates a textual description W of a spatial object o to a topic description T. For example, a textual description for a restaurant can be 'sea food', 'restaurant' and 'hot beverages'. A topic description is a collection of the topic distributions of W. In particular, T is an m-dimension vector. Each item $T[z]$ represents a relevance score of a topic z (intended activity) from an m-dimension topic dictionary D_T.

Latent Dirichlet Allocation (LDA) model [171] is adopted to translate the textual description W to a topic description T in the following equation:

$$T[z] = \frac{N_w + \epsilon}{|W| + |D_T| \times \epsilon} \tag{4.1}$$

N_w represents the number of keywords in W that belong to the topic z. $|W|, |D_T|$ is the size of the keyword, topic dictionary. ϵ is the symmetric Dirichlet prior [167]. In Table 4.1, we show an illustrative example for topic descriptions of 50 spatial entities. Topics from D_T include 'Game', 'Gym', et al.. We can see that a spatial entity can relate to multiple topics (activities). A higher $T[z]$ indicates a larger relevance score of the intended activity z.

To measure the distance between two topic descriptions T and T', we adopt the Euclidean distance metrics as follows:

$$Dist(T, T') = \sqrt{\sum_{z=1}^{m} (T[z] - T'[z])^2} \tag{4.2}$$

Table 4.1: An illustrative example of topic descriptions

	Game	Gym	...	Movie	Shop
T_1	0.67	0.03	...	0.03	0.03
T_2	0.03	0.03	...	0.91	0.03
...	...	...	...	...	...
T_{50}	0.04	0.04	...	0.75	0.04

Definition 2. A spatial object [168, 167] o_i for a spatial entity E_i is formally represented as:

$$o_i = (L_i, T_i)$$

$L_i = (x_i, y_i)$ is a two-dimension coordinate representation. T_i is the topic description of E_i.

Definition 3. A spatial keyword query function $\mathcal{F}_S$ takes into $\mathcal{O} = (o_1, o_2, ..., o_n)$ and a spatial keyword query $Q = (L_q, T_q, k, \lambda)$. It outputs top-$k$ spatial objects R_T from $\mathcal{O}$ that have the most higher relevance scores to the spatial keyword query Q.

$$\mathcal{F}_S(\mathcal{O}, Q) \rightarrow R_T$$

$L_q = (x_q, y_q)$ is a location of interest. T_q is the topic description of intended activities. λ is a preference ratio indicates the importance of the spatial proximity, while $1 - \lambda$ indicates the importance of the topic prolixity. k is the number of spatial entities that will be returned. We define the point distance $Dist(L, L')$ following the Euclidean distance metrics as follows:

$$Dist(L, L') = \sqrt{(L.x - L'.x)^2 + (L.y - L'.y)^2} \tag{4.3}$$

Finally, the relevance score RS between a query Q and a spatial object $o_i = (L_i, T_i)$ is defined in Eq. 4.4 [168, 167], where η is the normalization factor to convert the point distance to a range of (0,1).

$$RS(Q, o_i) = \lambda \times Dist(L_q, L_i)/\eta + (1 - \lambda) \times Dist(T_q, T_i) \tag{4.4}$$

4.3 Problem Formulation

In this section, we formulate the system model, security model, and design goals of *TAVLA*.

4.3.1 System Model

The system model of *TAVLA* is derived from the commercial model of the vehicular local advertising, such as local recommendation or trajectory plan services in the Google map. The main difference is that *TAVLA* introduces the blockchain as a public auditing infrastructure to improve the transparency and enforce accountability for the advertising system. In Fig. 4.2, we abstract four parties in *TAVLA* as spatial entity, vehicular user, ad broker, and the blockchain.

- ▷ A spatial entity is a retailer on a road network, such as a restaurant or a coffee store. The spatial entity would like to promote their business by sending location-aware [172] ads to vehicular users of related interests.

- ▷ Vehicular users refer to moving vehicles on the roads with GPS-based devices. They would like to find spatial entities within a specific geographical area for various activities such as food, movie, gas and shopping.
- ▷ The ad broker (such as Google Ads [17]) provides third-party ad dissemination services for spatial entities. It is responsible for running spatial keyword queries and disseminating location-aware ads to targeted vehicular users.
- ▷ The blockchain in *TAVLA* is a public and appended-only ledger, which supports secure and verifiable transactions among mutually distrustful peer nodes. A smart contract, that specifies involved parties, terms, and obligations, can be deployed on the blockchain.

In *TAVLA*, spatial entities carefully specify spatial objects with their locations and topic descriptions, and send the objects to the ad broker. The ad broker constructs a spatial keyword database and uploads the cryptographic authenticators of the database onto the blockchain. Vehicular users construct the spatial keyword query Q and send it to an advertising contract on the blockchain. The ad broker retrieves spatial keyword queries from the contract, executes the queries over its local spatial keyword database, and sends back results (most relevant spatial entities) to the contract with correctness proof. The advertising contract verifies the correctness of the results and notifies the public if any verification fails. Finally, vehicular users retrieve the correct results from the advertising contract.

In *TAVLA*, the advertising smart contract is running on the Ethereum blockchain. It is critical to motivate the Ethereum miners to join the local advertising process. In the local advertising system, spatial entities will pay the ad broker for managing keywords and delivering ads to related vehicular users, which can be partially directed to the blockchain miners as reward for managing the advertising smart contract.

4.3.2 Security Model

The ad broker and spatial entities are profit-driven commercial organizations. (1) Spatial entities carefully choose their locations and topic descriptions to enjoy effective ad dissemination services and attract more in-store visits. (2) Due to lack of advertising transparency and profit consideration, the ad broker may deviate from the spatial keyword query function in **Definition 3** and promote irrelevant or misleading ads to vehicular users [33]. Vehicular users accept the ad dissemination results if they are correctly generated.

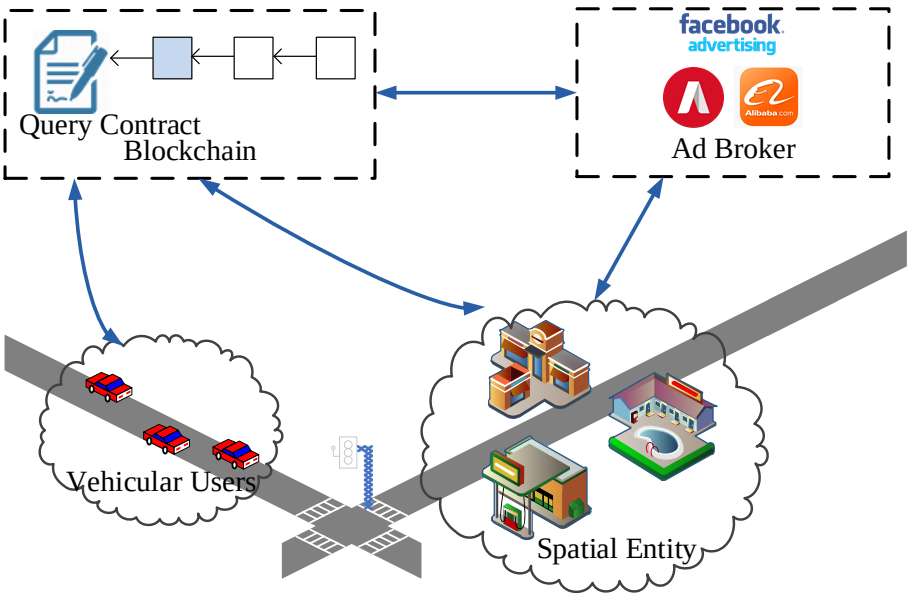


Figure 4.2: System Model

However, vehicular users may reject the results if there is lack of advertising transparency guarantees. Under the assumptions, we formulate the following *Auditing Security*:

Definition 4. Given a spatial keyword query function $\mathcal{F}_S$, that takes spatial objects $\mathcal{O}$ and a spatial keyword query Q as inputs and outputs a query result R_T, the *Auditing Security* has the following properties:

- ▷ Integrity: (1) Any $o_i \in \mathcal{O}$ digested in the on-chain authenticators cannot be maliciously modified in query executions. (2) Any Q generated by a vehicular user cannot be maliciously modified in query executions.
- ▷ Correctness: (1) R_T is a correct output of $\mathcal{F}_S$ with inputs $\mathcal{O}$ and Q. (2) Vehicular users accept the correct results.
- ▷ Transparency: The targeting keywords of spatial entities and the spatial keyword query process is transparent and verifiable to the public.
- ▷ Accountability: The misbehavior of the ad broker is publicly detected if the ad broker breaks either the integrity or correctness properties.

Remark. (1) The proposed blockchain-based architecture increases public transparency of the advertising system, which should contribute to the detection of advertising misconducts. It may also serve as the digital forensic evidence for law enforcement agencies to take actions against misbehaving parties. Moreover, the advertising smart contract can also take deposits from spatial entities and the ad broker and transfer the deposited money to vehicular users in case of any misconducts. (2) *TAVLA* relies on the pseudonym-based anonymity in the blockchain networks to protect vehicular users' query privacy. In specific, vehicular users interact with the advertising contract using anonymous blockchain accounts.

4.3.3 Design Goals

TAVLA aims at realizing the following design goals:

- ▷ Functionality: *TAVLA* should support the expressiveness of the spatial keyword query function in **Definitions 1-3**.
- ▷ Security: *TAVLA* should achieve the *Auditing Security* in **Definitions 4**.
- ▷ Efficiency: *TAVLA* should optimize both the on/off-blockchain computation and storage overheads for practical implementations.

4.4 Building Blocks

In this section, we present the building blocks of *TAVLA*, including the spatial indexing technique, a verifiable computation framework, and the smart contract.

4.4.1 Spatial Indexing

An R-tree [173, 168] is widely used for indexing multi-dimension spatial data. It enables efficiently location searching that returns spatial objects within a given geographical area. In specific, an R-tree T_R in *TAVLA* is a balanced tree that contains the following components:

- ▷ Bounding Rectangle: It is a non-leaf node in the R-tree and represents a geographical area. Each bounding rectangle has n_b children in the R-tree, that represent smaller geographical areas within their parent.

Table 4.2: Notations I

$\mathbb{G}$	Multiplicative groups
E_i	Spatial entity
$\mathcal{O} = (o_1, o_2, ..., o_n)$	Collection of n spatial objects
$o_i = \{L_i, T_i\}$	Spatial object o_i Location $L_i = (x_i, y_i)$ m-dimension topic description T_i
$Q = \{L_q, T_q, k, \lambda\}$	Query location $L_q = (x_q, y_q)$ Query topic description T_q Number of returned spatial objects k Preference ratio λ
T_R	Spatial R-tree
$MB = (L_l, L_r)$	Minimum bounding rectangle MB Lower-left point L_l Upper-right point L_r
n_l	Number of MBs in T_R
n_e	Number of spatial objects in an MB

▷ Minimum Bounding Rectangle (MB): It is a minimum geographical splitting area, specified by a lower-left point L_l, and a upper-right point L_r. Each MB consists of n_e spatial entities.

In Fig. 4.3, the upper figure shows an R-tree example, while the lower figure represents a geographical area. The geographical area is first split into two subareas (bounding rectangles) (R_1, R_2), each of which is then split into smaller subareas (R_3, R_4, R_5, R_6). Recursively, the whole geographical area is split into $n_l = 8$ minimum bounding rectangles (R_7 to R_{14}). Each minimum bounding rectangle consists of $n_e = 50$ spatial entities. We define two algorithms of an R-tree as follows:

Definition 5. Given an R-tree T_R, it consists of two algorithms:

▷ **Search**. The algorithm takes into a location L and recursively query T_R in a top-down manner to find an MB that contains L. The algorithm returns the identifier of the found MB.

▷ **Insert**. The algorithm takes into a spatial object o_i and inserts o_i into T_R. The algorithm finds an MB to place o_i using **Search** algorithm. If there is still room

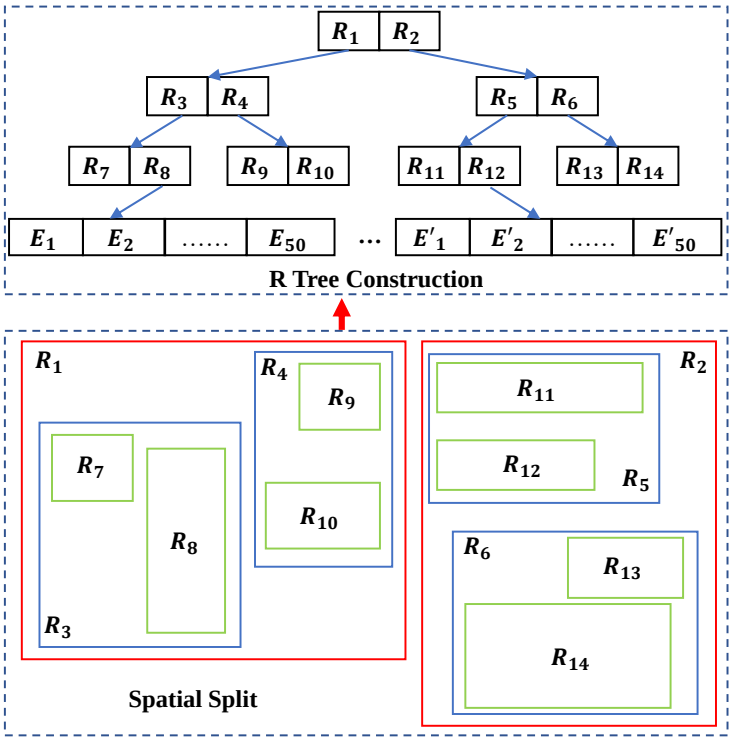


Figure 4.3: An R-tree Example

in MB, the algorithm inserts o_i into MB. Otherwise, the algorithm will *Split* MB into two smaller rectangles and *Adjust* T_R to remain balanced.

We omit the details of *Split* and *Adjust* algorithms. It is a lively research topic in the database area [166] and may be of independent research interests in the design of *TAVLA*.

4.4.2 Verifiable Computation

A VC framework [122] enables a verifier to outsource an execution of an NP-compete relation to a prover and verify the correctness of the execution results. In particular, an NP-complete relation $\mathcal{R}$ is denoted as a public function $\mathcal{F}$. $\mathcal{F}$ takes into I, Q as inputs and outputs a result R. In a VC framework, $\mathcal{F}$ is first converted to a Quadratic Arithmetic

Program (QAP) [134]. Then, the verification of the function execution is equivalent to the divisibility check of the QAP. The divisibility check of the QAP is finally converted to a linear combination check with the Succinct Non-interactive ARGuments ($SNARG$) techniques. We present definitions of the Pinocchio VC framework [122] as follows:

Definition 6. A Pinocchio VC framework consists of three algorithms:

$$VC = \{KeyGen,\ Evaluate, Verify\}.$$

▷ **KeyGen**$(\mathcal{F}, pp)$. The algorithm takes a function $\mathcal{F}$ and system public parameters pp. It outputs a relation-dependent common reference string $CRS = (ek, vk)$, where ek is for evaluations and vk is for verifications.

▷ **Evaluate**$(\mathcal{F}, ek, I, Q)$. The algorithm takes $\mathcal{F}, ek$ and function inputs I, Q. It outputs a function result R along with a correctness proof π_F.

▷ **Verify**(Q, R, π_F, vk). The algorithm takes the input Q, the result R, the proof π_F, and the verification key vk. It outputs *true* if $\mathcal{F}(I, Q) \rightarrow R$. Otherwise, it outputs *false*.

A VC framework should have the following properties:

▷ Succinctness: The length of the proof $|\pi_F|$ is polynomial in the system security parameter α.

▷ Completeness: An honest verifier will always accept (R, π_F), if $\mathcal{F}(I, Q) \rightarrow R$.

▷ Soundness: An computationally-bounded adversary cannot forge an invalid tuple (R', π'_F), where $\mathcal{F}(I, Q) \nrightarrow R$ and $Verify(Q, R', \pi'_F, vk) = true$.

4.4.3 Smart Contract

Smart contract [81, 115] is a computer program executed over the blockchain by all blockchain peer nodes. A smart contract can take into crypto currencies as deposits from blockchain nodes, specify terms for the nodes, and take actions (transfer crypto currencies) if terms are met. In Ethereum, a smart contract is a special blockchain account with codes

written in Solidity [161] on the blockchain. Participating blockchain nodes execute the smart contract by function calls as transactions to the smart contract address. In this way, each contract execution (state transition) is verified by all blockchain peer nodes. Based on the security and openness of the underlying blockchain, the Ethereum smart contract has the following properties. (1) Transparency: contract terms and executions are transparent to the public. (2) Security: contract executions are secure, which means the state transitions must follow the defined terms and cannot be later modified.

4.5 Transparent and Accountable Vehicular Local Advertising

In this section, we present the blockchain-based *TAVLA*. First, we summarize the design strategies of *TAVLA*. Then, we design a verifiable spatial keyword query scheme with three phases: System Setup, *SKD*-tree Construction, and Spatial Keyword Query Processing. Finally, we develop a *TAVLA* smart contract that realizes the verifiable spatial keyword query scheme with on/off chain computation optimizations.

4.5.1 Overview

TAVLA utilizes the public blockchain as an auditing infrastructure, to achieve transparent and accountable ad dissemination process for the vehicular local advertising. To realize the promises of the blockchain technology with feasible storage and computation overheads, we introduce two design strategies as follows:

Digest-and-verify. The ad broker constructs query indexes of the spatial keyword database, digests the indexes as cryptographic authenticators, and uploads the authenticators onto to the blockchain. The on-chain authenticators have multiple functionalities. First, the authenticators are the evidence that the indexes are correctly digested. That is, retailers are able to check that their spatial objects are correctly digested in the indexes. Second, the ad broker can process spatial keyword queries in an off-chain manner. The correctness of the query results can be efficiently verified on the blockchain with the help of the cryptographic authenticators. With this strategy, the on-chain storage and computation overheads can be reduced to *succinct* regardless of the size of the spatial keyword database.

Divide-then-assemble. We identify that off-chain computation and storage overheads are dominated by the input size of $\mathcal{F}_S$, that must be determined at the **KeyGen** phase

in the VC framework. If $\mathcal{F}_S$ takes the whole spatial keyword database as inputs, it would incur prohibitive off-chain overheads. To enhance the off-chain performance, we adopt the probabilistic topic model to prune the keyword dimension of spatial objects. Meanwhile, we divide $\mathcal{F}_S$ into two verifiable functions: spatial search function $\mathcal{F}_R$ and topic matching function $\mathcal{F}_T$. $\mathcal{F}_R$ finds a minimum bounding rectangle MB_j that contains an intended location of a query Q. $\mathcal{F}_T$ finds top-k spatial objects with highest relevance scores in MB_j. The spatial keyword database is also divided into a spatial index and a topic index with distinct on-chain authenticators. As a result, $\mathcal{F}_S$ is realized with modular executions of $\mathcal{F}_R$ and $\mathcal{F}_T$, the results of which will be assembled on-chain with selective authenticators. With this strategy, *TAVLA* achieves a practical off-chain performance by reducing the input size of the off-chain functions.

4.5.2 Verifiable Spatial Keyword Query

In the following subsections, we present the details of the verifiable spatial keyword query scheme. We assume secure and authenticated channels are set up for involving entities. Notations in **Definitions 1-6** are re-used.

Table 4.3: Notations II

pp	Public Parameters
$\mathcal{F}_S = (\mathcal{F}_R, \mathcal{F}_T)$	Spatial keyword query $\mathcal{F}_S$ R-tree search $\mathcal{F}_R$, topic distance $\mathcal{F}_T$
CRS_R, CRS_T	CRS_R for $\mathcal{F}_R$ CRS_T for $\mathcal{F}_T$
$I = (I_R, I_T)$	Spatial index I_R Topic index $I_T = (I_1, ..., I_{n_l})$
$D = (D_R, D_T)$	Spatial index authenticator D_R Topic index authenticator D_T
$A[j]$	j-th element of an array A
π_R	Correctness proof for $\mathcal{F}_R$
π_T	Correctness proof for $\mathcal{F}_T$
π_D	Digest proof for D_R
$\hat{\pi}_D$	Digest proof for D_T
R_T	Spatial keyword query result

System Setup

For illustrative purposes, we introduce a trusted authority (TA) to setup the system. In practice, the role of TA can be replaced by a secure multi-party computation protocol [174]. Specifically, a set of entities (e.g. different ad brokers and randomly selected spatial entities, etc.) can agree on the spatial keyword query algorithm and setup the system. In practice, such setup mechanism has been successfully running for Zerocash system [96].

TA chooses a system security parameter α and sets the bilinear groups $\mathbb{G} = (\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, q, e)$ with a prime order q and a bilinear paring e. TA chooses $g \in \mathbb{G}_1$ and $\tilde{g} \in \mathbb{G}_2$. TA denotes public parameters of the system as $pp = \{\alpha, \mathbb{G}, g, \tilde{g}\}$. TA divides the spatial keyword query function $\mathcal{F}_S$ as $(\mathcal{F}_R, \mathcal{F}_T)$ as follows:

$$\mathcal{F}_R(I_R, Q) \rightarrow MB_j,\ \mathcal{F}_T(I_j, Q) \rightarrow R_T \tag{4.5}$$

I_R is the spatial index of an R-tree T_R. Q is a spatial keyword query $Q = (L_q, T_q, k, \lambda)$. $\mathcal{F}_R$ finds a minimum bounding rectangle $MB_j \in T_R$ that contains L_q and outputs the identifier j of MB_j. I_j is the topic index of spatial objects in MB_j. $\mathcal{F}_T$ finds the top-k spatial objects in I_j that have highest relevance scores with Q and outputs identifiers of the k spatial objects as R_T. We note that algorithms of $\mathcal{F}_R$ and $\mathcal{F}_T$ are determined by TA at the setup phase, which will be discussed in details in *SKD*-tree Construction subsection.

TA computes common reference strings for $\mathcal{F}_R$ and $\mathcal{F}_T$. Since Pinocchio VC framework implements a non-updatable CRS model, TA must determine the size of I_R and I_j to generate the CRS. In specific, T_R is denoted as a balanced tree. n_b is the number of the bounding rectangles in each non-leaf node. n_l is the number of leaf nodes (MB). n_e is the number of spatial objects in each MB. TA runs $\mathbf{KeyGen}(\mathcal{F}_R, pp)$ and $\mathbf{KeyGen}(\mathcal{F}_T, pp)$ to generate $CRS = (CRS_R, CRS_T)$. TA randomly chooses $a, b, c \in \mathbb{Z}_q$ and computes $A = \tilde{g}^a, B = \tilde{g}^b, C = \tilde{g}^c \in \mathbb{G}_2^3$. Then, TA chooses $X = \{X_i\}_{i\in[1,4n_l]}, Y = \{Y_i\}_{i\in[1,4n_l]}$, where X_i, Y_i are randomly chosen from $\mathbb{G}_1$. Similarly, TA chooses $\hat{X} = \{\hat{X}_i\}_{i\in[1,(m+2)n_e]}, \hat{Y} = \{\hat{Y}_i\}_{i\in[1,(m+2)n_e]}$ from $\mathbb{G}_1$. TA computes $Z_i, \hat{Z}_j \in \mathbb{G}_1$ as follows:

$$\begin{aligned} &Z_i = X_i^a Y_i^b F_i^c,\ i \in [1, 4n_l],\ F_i \in CRS_R, \\ &\hat{Z}_j = \hat{X}_j^a \hat{Y}_j^b \hat{F}_j^c,\ j \in [1, (m+2)n_e],\ \hat{F}_i \in CRS_T. \end{aligned} \tag{4.6}$$

m is the dimension of the topic description T_i. $F = \{F_i\}_{i\in[1,4n_l]} \in \mathbb{G}_1^{4n_l}$ are from CRS_R. $\hat{F} = \{\hat{F}_i\}_{i\in[1,(m+2)n_e]} \in \mathbb{G}_1^{(m+2)n_e}$ are from CRS_T. TA sets $Z = \{Z_i\}_{i\in[1,4n_l]} \in \mathbb{G}_1^{4n_l}, \hat{Z} =$

$\{\hat{Z}_i\}_{i\in[1,(m+2)n_e]} \in \mathbb{G}_1^{(m+2)n_e}$. TA denotes $K_D = (K_R, K_T, K_V)$, where $K_R = (F, X, Y, Z)$, $K_T = (\hat{F}, \hat{X}, \hat{Y}, \hat{Z})$, and $K_V = (A, B, C)$.

TA publishes $\{pp, CRS, K_D\}$.

SKD-Tree Construction

Each spatial entity E_i constructs a spatial object $o_i = (L_i, T_i)$ and sends o_i to the ad broker. Upon receiving o_i, the ad broker will return a signature (e.g. ECDCS) on o_i as the proof of receipt. The ad broker collects all received spatial objects as a set $\mathcal{O} = (o_1, o_2, ..., o_n)$. Adopting algorithms in **Definition 5**, the ad broker constructs a spatial index I_R and topic index I_T in Alg. 1.

Algorithm 1: Index Construction

Input: $\mathcal{O} = (o_1, o_2, ..., o_n)$
Output: Spatial index I_R, topic index I_T
Set T_R, I_R, I_T to $\emptyset$
for $i \in [1, n]$ **do**
  Insert o_i into T_R
if *(number of MB in* T_R*)* $< n_l$ **then**
  Pack empty MBs to T_R
for $MB_i \in T_R$ **do**
  Add $(L_l, L_r) \in MB_i$ to I_R
  for $o_j \in MB_i$ **do**
    Add $(L_j, T_j) \in o_j$ to I_i
  if *(number of objects in* I_i*)* $< n_e$ **then**
    Pack empty objects to I_i
  Add I_i to I_T

Remark. (1) Our design is not coupled to a specific R-tree construction. As a result, *TAVLA* can naturally inherit technical advances (such as novel node split/deletion algorithms) for featured spatial databases. (2) Since the size of CRS is fixed in the setup phase, we pack empty MBs or spatial objects to I_R or I_i, to comply with the pre-determined index size n_l and n_e. An alternative strategy is that the ad broker first constructs the spatial database and requires an one-time CRS from TA.

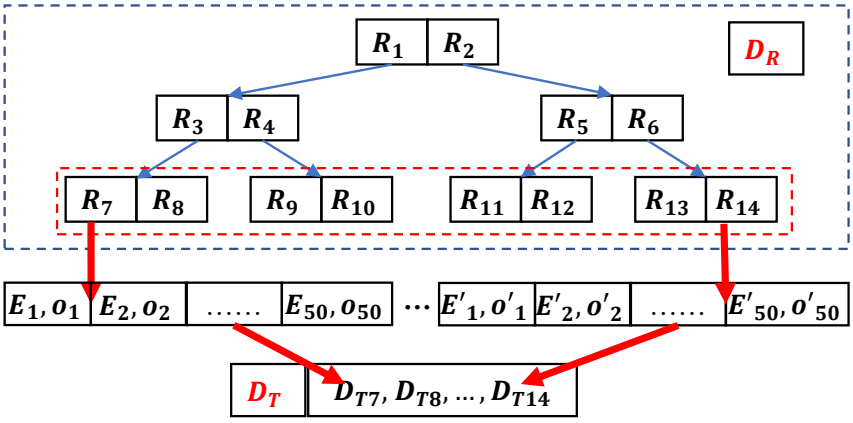


Figure 4.4: An *SKD*-tree Example

We arrange I_R and I_T as arrays. I_R is a $4n_l$-dimension array, since each MB has four coordinates from (L_l, L_r). I_i is an $(m+2)n_e$-dimension array, since each $o_i \in I_i$ has an m-dimension topic description vector and a 2-dimension location L_i. The ad broker computes a spatial authenticator D_R and a topic authenticator $D_{T[i]}$ as follows:

$$\begin{aligned} D_R &= \prod X_i^{I_R[i]}, \ for\ i \in [1, 4n_l], \\ D_{T[i]} &= \prod \hat{X}_j^{I_i[j]}, for\ j \in [1, (m+2)n_e]. \end{aligned} \tag{4.7}$$

$I_R[i]$ represents the i-th element in I_R and $I_i[j]$ represents the j-th element in I_j. The ad broker denotes $D_T = (D_{T[1]}, D_{T[2]}, ..., D_{T[n_l]})$ and uploads the *SKD*-tree authenticator $D = (D_R, D_T)$ to the TAVLA smart contract. A typical example is shown in Fig. 4.4. We can see that the *SKD*-tree is a balanced binary tree with 8 minimum bounding rectangles. I_R consists of MBs from R_7 to R_{14} with an authenticator D_R. Each MB contains 50 spatial objects. For example, R_7 contains spatial objects from o_1 to o_{50} with a topic authenticator $D_{T[7]}$.

Remark We note that the keywords of spatial entities are managed by the ad broker in an off-chain manner and updated onto the blockchain with a succinct digest, which serve the following purposes. (1) Spatial entities can require the ad broker to 'open' the digest anytime and check if their spatial objects are correctly included in the authenticators. The ad broker cannot forge a false opening since he cannot solve the *Decision Diffie-Dellman* problem. (2) The ad broker can prove that each ad dissemination is correctly conducted following the ad dissemination protocol. By doing so, the digest becomes an immutable and auditable evidence of the advertising system.

Spatial Keyword Query Processing

Following **Definition 3**, we present modular designs of the spatial keyword query function, which consists of three algorithms:

$$\{QExe,\ QProv\ , QVeri\}.$$

QExe (Alg. 2) takes an R-tree T_R, a topic index I_T, and a query $Q = (L_q, T_q, k, \lambda)$. The algorithm outputs top k relevant spatial objects as R_T.

Algorithm 2: Query Execution

Input: T_R, I_T, Q
Output: R_T
Run **Seach** with (T_R, L_q), find MB_j that contains L_q
for $o_i \in MB_j$ **do**
  Compute the relevance score $RS(Q, o_i)$
Add k objects with highest scores into R_T

QProv (Alg. 3) takes the public parameters pp, $CRS = (CRS_R, CRS_T)$, the index (I_R, I_T), MB_j, the query Q, the authenticators (D_R, D_T), and the digest keys K_D. It outputs proofs $\pi = (\pi_R, \pi_T, \pi_D, \hat{\pi}_D)$.

Algorithm 3: Query Prove

Input: $pp, CRS, (I_R, I_T), MB_j, Q, (D_R, D_T), K_D$
Output: π
Run $Evaluate(\mathcal{F}_R, CRS_R.ek, I_R, Q)$ to generate π_R
Run $Evaluate(\mathcal{F}_T, CRS_T.ek, I_j, Q)$ to generate π_T
Compute $Z_\pi = \prod_{i\in[1,4n_l]} Z_i^{I_R[i]}$
Compute $Y_\pi = \prod_{i\in[1,4n_l]} Y_i^{I_R[i]}$
Compute $\hat{Z}_\pi = \prod_{i\in[1,(m+2)n_e]} \hat{Z}_i^{I_j[i]}$
Compute $\hat{Y}_\pi = \prod_{i\in[1,(m+2)n_e]} \hat{Y}_i^{I_j[i]}$
Set $\pi_D = (Z_\pi, Y_\pi), \hat{\pi}_D = (\hat{Z}_\pi, \hat{Y}_\pi)$
Set $\pi = (\pi_R, \pi_T, \pi_D, \hat{\pi}_D)$

QVeri (Alg. 4) verifies the correctness of the proofs.

Algorithm 4: Query Verify

Input: $CRS, Q, D, K_V, (j, R_T), \pi$
Output: True or false
Check $Verify(Q, j, \pi_R, CRS_R.vk)$
Check $Verify(Q, R_T, \pi_T, CRS_T.vk)$
Extract $c_x = \prod F_i^{I_R[i]},\ i \in [1, 4n_l]$ from π_R
Extract $\hat{c}_x = \prod \hat{F}_i^{I_j[i]},\ i \in [1, (m+2)n_e]$ from π_T
Check $e(Z_\pi, \tilde{g}) \stackrel{?}{=} e(D_R, A)e(Y_\pi, B)e(c_x, C)$
Check $e(\hat{Z}_\pi, \tilde{g}) \stackrel{?}{=} e(D_{T[j]}, A)e(\hat{Y}_\pi, B)e(\hat{c}_x, C)$
Return *true* if all checks pass. Otherwise, return *false*

Remark. (1) Modular uses of the building blocks is achieved in *TAVLA*, since the same building blocks are sufficiently abstracted in the previous section and are used multiple times. (2) We re-design functions $\mathcal{F}_R$ and $\mathcal{F}_T$ to fit the *digest-and-verify* strategy, which are implemented in C codes and will be discussed in the performance evaluation section.

4.5.3 *TAVLA* Smart Contract

We design a *TAVLA* smart contract that realizes the verifiable spatial keyword query scheme with Solidity [161] of Ethereum. The contract is created by TA, that stores public parameters pp, verification keys ek_R, ek_T, digest verification keys K_V, and digest authenticators $D = (D_R, D_T)$. The details of *TAVLA* contract is shown in Alg. 5. The definitions and algorithms proposed in the previous sections are re-used.

Vehicular users call *SendQuery* function to send spatial keyword queries to the contract, where *addr* is the blockchain address of the message sender. The ad broker retrieves unprocessed queries using the *RetrieveQuery* function. The ad broker executes the queries locally via **QProv** and **QExe** function, and uploads the results with proofs to the smart contract via the *SendResult* function. The correctness of the result and proofs are verified by the smart contract. The smart contract stores valid results to be retrieved by vehicular users via *QueryResult* function. For invalid results, the smart contract generates a verification failure event to notify the public.

Remark. (1) Vehicular user privacy is not a primary concern of *TAVLA*. Vehicular users can apply for one-time blockchain accounts and utilize anonymous payment channels (such as Zerocash). (2) The main goal of the contract is to improve the advertising system transparency and accountability. The accountability in *TAVLA* refers to the *detection*

Algorithm 5: *TAVLA* Smart Contract

Require: (ek_R, ek_T), K_V, (D_R, D_T)
Set Rec_Q to $(addr, Q, flag)$
Set Rec_R to $(addr, MB_j, R_T)$
Function `SendQuery(`*a spatial keyword query* Q`)`
  Set $addr = sender.addr, flag = 0$
  Add $(addr, Q, flag)$ to Rec_Q
Function `RetrieveQuery()`
  Require: msg.sender=ad broker
  Retrieve all $(addr, Q)$ from Rec_Q with $flag = 0$
Function `SendResult(`$addr, (j, R_T), \pi$`)`
  Require: msg.sender=ad broker
  Retrieve $(Q, flag)$ from Rec_Q by $addr$
  Require: $flag = 0$
  if $\boldsymbol{QVeri}(CRS, Q, D, K_V, (j, R_T), \pi) = true$ **then**
    Set $flag$ to 1
    Add $(addr, j, R_T)$ to Rec_R
  else
    Generate a verification failure event
Function `QueryResult()`
  Set $addr$ to msg.sender
  Retrieve $(addr, j, R_T)$ from Rec_R

or *public awareness* of the ad dissemination misbehavior. Potential obligations on the misbehavior can be transfers of the ad broker's pre-deposited crypto currencies or enforcing fines by law enforcement agencies.

4.6 Security Analysis

In this section, we present the security analysis of *TAVLA*. We first review the security properties inherited from the cryptographic building blocks: the blockchain and the verifiable computation framework. Then, we present the detailed analysis on the security properties of *Auditing Security*. Finally, we conclude *Auditing Security*.

4.6.1 Blockchain Security

The consensus protocol of a public blockchain (i.e. Proof-of-Work in *Ethereum*) provides three useful properties: *chain growth*, *chain quality* and *consistency* [81, 175]. Informally, the three properties guarantee: (1) a valid transaction will be accepted by honest blockchain nodes within a certain time (transaction confirmation time); (2) a Byzantine adversary that controls less than 50 percent computation power of the blockchain system cannot control the growth of the chain; and (3) honest blockchain nodes maintain a consistent view of the shared ledger.

4.6.2 Verifiable Computation Framework Security

Completeness. An honest verifier always accepts a result and a proof if they are correctly computed. For Pinocchio VC framework, the QAP-based $SNARG$ system [134] recognizes an NP-complete relation that can be compiled to an arithmetic circuit C. From the QAP theorem, the circuit evaluation of C is equivalent to the divisibility check of the compiled polynomials. The QAP divisibility check is further converted to a linear check over bilinear groups. Based on the correctness of QAP theorem and bilinear groups, the *completeness* is achieved.

Soundness. A computationally-bounded adversary (usually refers to a malicious prover) cannot forge an invalid result with a proof that passes the correctness check (*Verify* function in the VC framework). For a compiled QAP with a degree d and bilinear groups with an order q, we borrow the theorem from [134] that *Soundness* is achieved if (1) **q-PDH,**

2q-SDH, and **d-PKE** assumptions hold for $q \geq 4d + 4$. (2) The trapdoor secret used to generate common reference strings is destroyed.

Succinctness. The VC framework achieves a succinct size of proof that depends on the size of the system security parameter α regardless of the function input size.

4.6.3 *TAVLA* Security

Based on the above security properties from the building blocks, we give a sketch analysis of the security properties in **Definition 4**.

Integrity. Spatial objects are generated by individual spatial entities and sent to the ad broker in a secure channel. Then, the ad broker computes the authenticators D_R and D_T to be uploaded onto the blockchain. To ensure the individual spatial object is digested in the authenticator, the spatial entity can require the ad broker to compute a proof for the correct authenticator generation, which is either a zero-knowledge proof of a linear relation in the discrete logarithm setting or a direct opening of the targeting keywords. The query smart contract receives spatial keyword queries from vehicular users and verifies the correctness of query executions using the on-chain authenticators. If the on-chain storage and advertising contract executions are secure in the Ethereum blockchain, the *Integrity* property is achieved in *TAVLA*.

Correctness. The first property of *Correctness* comes from three folds: (1) *Soundness* of the underlying VC framework. (2) *Integrity* of spatial objects and spatial keyword queries. (3) Unforgeability of the on-chain authenticators. For each spatial keyword query, the ad broker performs the query over the spatial and topic index I_R and I_T. In specific, the ad broker runs the *Evaluate* function of $\mathcal{F}_R$ and $\mathcal{F}_T$ and proves the correctness of query executions. In *TAVLA*, CRS_R and CRS_T are securely generated by TA or a secure multiparty computation protocol. The ad broker can forge a query result with a proof that passes the *Verify* function of $\mathcal{F}_R$ and $\mathcal{F}_T$, iff the ad broker can break the *Soundness* property of the VC framework. The proof π_R and π_T contain multi-exponentiation forms c_x and $\hat{c}_x$ as the representations of I_R and I_T. In specific, c_x and $\hat{c}_x$ follow the form of extended Pedersen commitment for vectors. The ad broker proves that the c_x and $\hat{c}_x$ open to the same value of the on-chain authenticators (D_R, D_T). The ad broker can forge valid proofs $(\pi'_D, \hat{\pi}'_D)$ that pass the check in Alg. 4, iff the ad broker can solve the **SXDH** problem in bilinear groups [141]. For the second property of *Correctness*, honest vehicular users will always accept valid results and proofs due to the *Completeness* property of the VC framework.

Transparency. The Ethereum blockchain is a public and permissionless ledger. Since the succinct digest of the spatial objects are uploaded to the Ethereum, vehicular users can require the ad broker to publish the keywords of a specific spatial entity for further authenticity checking. At the same time, executions of the local advertising contract are also transparent to the public.

Accountability. We emphasize that *Accountability* of *TAVLA* refers to the public *detection* or *awareness* of the ad broker misbehavior. Spatial entities can require the ad broker to publish the proof of correct authenticator generation, while vehicular users directly receive the query results on the smart contract, both of which can be publicly verifiable.

4.7 Performance Evaluation

In this section, we implement *TAVLA* and evaluate its performance with benchmarks in terms of on/off chain overheads.

4.7.1 Off-chain Overheads

Off-chain operations include System Setup, *SKD*-tree Construction, and Spatial Keyword Query Processing. We omit implementations and evaluations of the R-tree and probabilistic topic model, which have been well studied in the non-verifiable setting. In *TAVLA*, our implementation goal is to evaluate additional overheads with the implementation of the VC framework. Thus, we construct testing instances of a balanced R-tree and topic descriptions. With the testing instances, we evaluate performances of $\mathcal{F}_R$ and $\mathcal{F}_T$, in terms of off-chain storage and computation overheads.

We conduct off-chain experiments on a Linux system with Intel Core 2.4 GHz processor and 8 GB memory. The functions $\mathcal{F}_R$ and $\mathcal{F}_T$ are written in C codes. We implement the Python interface of the Pinocchio [122] VC framework that translates the query execution into arithmetic circuits. We note that the Pinocchio interface is compiled with a 32-bit version gcc. We write a circuit parser in C++ to parse the obtained circuits with 'nizk' circuit inputs (both spatial query Q and spatial objects $\mathcal{O}$) and implement the C++ interface of libsanrk [140, 136, 139] for *R1CS* languages. Our design is not specified to particular implementations of *QAP*-based VC framework. Thus, *TAVLA* can inherit any efficiency improvements of future optimizations for *QAP*-based verifiable computations [139].

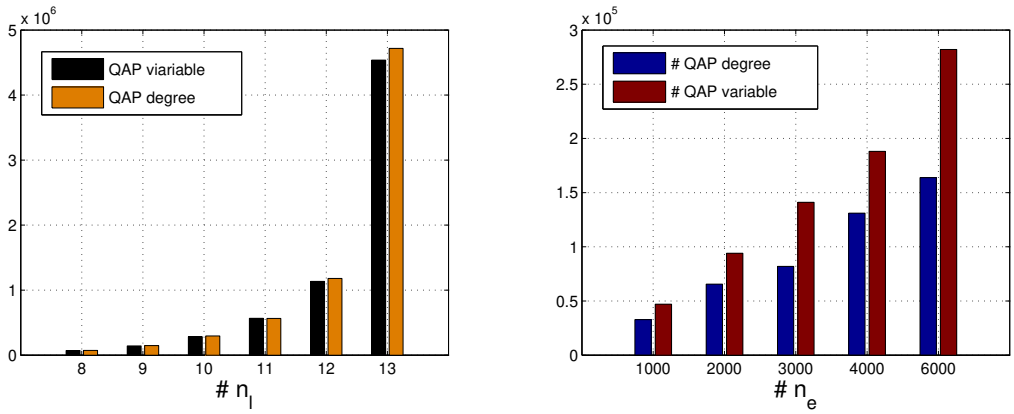
However, we have found several implementation limitations in the circuit-based VC framework: (1) The Pinocchio C program compiler only supports static compilation, which requires fixed-size spatial and topic indexes as inputs [176, 177, 109]. (2) Subscripts of the array access in the C codes must be determined at the program compiling phase. Thus, the logarithmic R-tree search algorithm cannot be implemented. Instead, we implement a linear search algorithm over the leaf nodes in the spatial index. In specific, we must compare L_q of Q with (L_l, L_r) of each MB. (3) The Pinocchio C program compiler only supports integers and simple arithmetic operations. We re-write the relevance score function in Eq. 4.4 as follows:

$$\lambda_1 \times Dist^2(L_q, L_i) + \lambda_2 \times Dist^2(T_q, T_i)$$

The location coordinates L_q, L_i, topic descriptions T_q, T_i and preference factors λ_1, λ_2 are set as integers in the experiments. Square root computations of point and topic distances are eliminated. However, the accuracy of the relevance ranking may be affected compared with the original metric in Eq. 4.4. Thus, we set a larger k in our experiments. Similarly, a local search in Google Map for a specific area will return all the relevant results to users.

We first identify the main off-chain performance metrics. In specific, the complexity of the compiled quadratic arithmetic program is characterized by the QAP variables and degrees. The storage overhead is characterized by the size of CRS. Another important metric is the off-chain processing time of the two functions, in terms of CRS setup and prover computation. We set the number of bounding rectangles n_b as 2, which results in a binary tree structure and is easy to be adjusted to balance when new nodes are inserted. We set the number of returned objects k to be equal to n_e. This is reasonable since a spatial keyword query usually returns all relevant result to users, such as activity spot search in Google Map. The dimension m of the topic description vector is 20, which is sufficient in the probabilistic topic model [167]. The normalization factor η can be adjusted with the change of λ_1. Input sizes of $\mathcal{F}_R$ and $\mathcal{F}_T$, characterized by (n_e, n_l), greatly affects the performance. n_l is represented as the power of 2, since the R-tree is a balanced binary tree in the experiments. From Fig. 4.5(a) and 4.5(b), we can see that the QAP complexity is increasing with n_l and n_e with different rates.

In Fig. 4.6, the same increasing property with n_l and n_e is found for PK size. In Pinocchio vc framework, the VK size is determined by the number of plaintext inputs and outputs. Since we enable 'nizk' for all the input, VK size is solely determined by the number of outputs, which results in a constant-size VK in $\mathcal{F}_R$ and a linearly increasing size of VK in $\mathcal{F}_T$. PK size is much more larger (10^9 magnitude) compared with VK size,



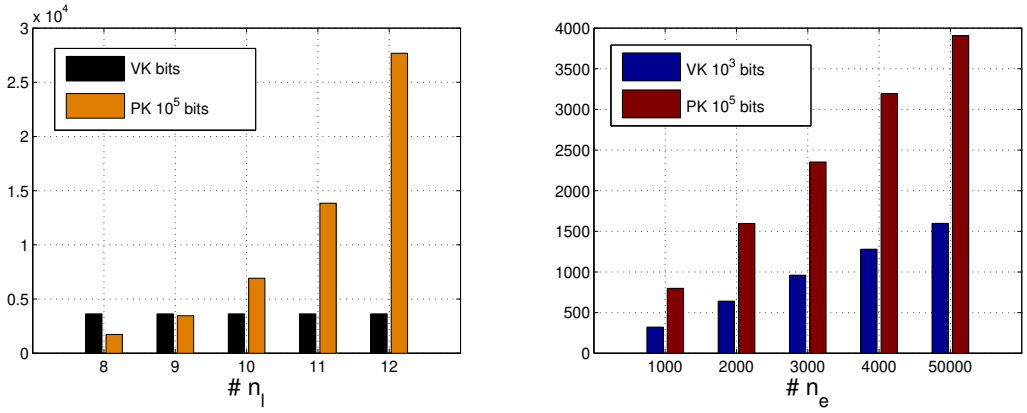
(a) $\mathcal{F}_R$ (b) $\mathcal{F}_T$

Figure 4.5: *QAP* Complexity

since PK must embed information of both input and intermediate gates in the compiled circuits, while VK only embeds information of output gates.

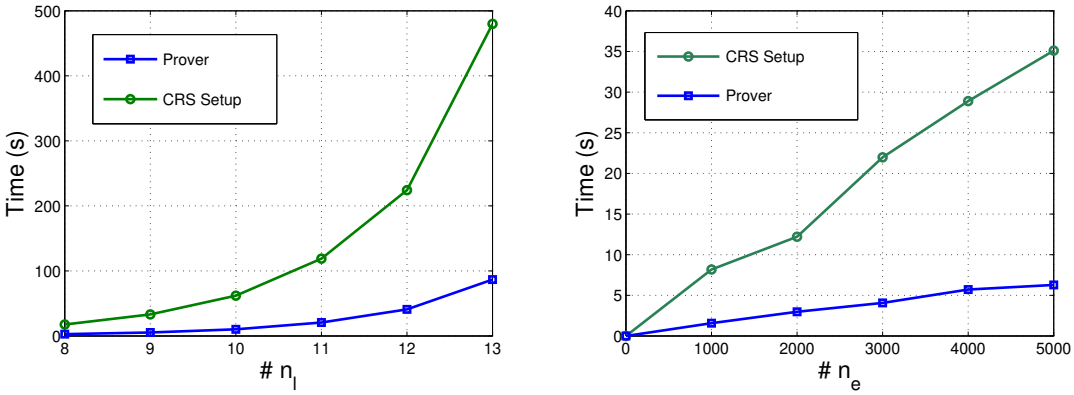
In Fig. 4.7, the prover overhead refers to the *Evaluate* algorithm in the VC framework. The CRS setup is much more costive than the prover overhead, since the setup is an one-time cost. Meanwhile, the processing time is also increasing as n_l and n_e grow. It should be noted that the prover overhead with an input of a few thousand objects is a few seconds on a laptop, which can be improved at the ad broker with powerful computing clusters. Moreover, distributed and parallel optimization techniques for verifiable computations can also be adopted to further enhance the prover performance.

We summarize the storage and computing complexity of cryptographic authenticators in Table 4.4. Although the complexity increases greatly with the input size (n_l, n_e), it is still less significant than the CRS setup and prover computation overheads. The reason is that multi-exponentiation operation is extremely optimized in the *alt-bn128* and *bn-128* curves in the libff library [178] of *libsnark*. For example, a 254-bit multi-exponentiation operation only takes 231.2 μs on a 2.4 GHz Intel Xeon E5-2620 [141]. $|\mathbb{G}_1|$ refers to the size of an element in $\mathbb{G}_1$. E_1 is one exponentiation operation in $\mathbb{G}_1$. The *digest-and-verify* strategy may slightly increases the size of π_R and π_T in our off-chain experiments with the *libsnark* implementation by requiring independent components $c_x, \hat{c}_x$.



(a) $\mathcal{F}_R$ (b) $\mathcal{F}_T$

Figure 4.6: CRS Size



(a) $\mathcal{F}_R$ (b) $\mathcal{F}_T$

Figure 4.7: Off-chain Computation Cost

Table 4.4: Digest Cost vs n_l, n_e

	K_R, K_T Size	D_R, D_T Comp.	$\pi_D, \hat{\pi}_D$ Comp.
$\mathcal{F}_R$	$16n_l\|\mathbb{G}_1\|$	$4n_lE_1$	$8n_lE_1$
$\mathcal{F}_T$	$4(m+2)n_e\|\mathbb{G}_1\|$	$(m+2)n_eE_1$	$2(m+2)n_eE_1$

4.7.2 On-chain Overheads

We implement a Parity Ethereum testing network [151]. The VC framework is instantiated with *alt-bn128* curve, which is compatible for the pre-compiled pairing interface in Ethereum [81]. Since *TAVLA* smart contract is designed for the Ethereum blockchain, we take the main performance metrics from the Ethereum to evaluate the *TAVLA* smart contract: on-chain storage cost, transaction confirmation time and gas cost of function calls. The 'gas' is a unit in the Ethereum to measure the computation complexity of a transaction.

Storage cost. The on-chain storage includes the verification keys for $\mathcal{F}_R$ and $\mathcal{F}_T$, the index authenticators, and the digest verification keys. VK size is determined by the number of non zero-knowledge variables. The authenticator size for the spatial and topic index $I = (I_R, I_T)$ is $(n_l + 1)|\mathbb{G}_1|$. The size of the digest verification key $K_V = (A, B, C)$ is $3|\mathbb{G}_2|$.

Function call. In Ethereum, each function call is instantiated by an Ethereum transaction. The transaction confirmation time can be approximated by ETH status [179]. The most expensive function call is the *SendResult* function, which is dominated by the number of pairing operations in the function. The verifications of $\mathcal{F}_R$ and $\mathcal{F}_T$ include 13×2 pairings. Compared with our off-chain experiments with the *libsnark* implementation, *TAVLA* needs 1 more pairing for the verification, since the verifier needs to check that the appropriate span of c_x or $\hat{c}_x$. The authenticator checks require 4×2 pairings. According to statistics in EIP 1108, total verification cost is approximately $1,201,000$ gas. Note that, the *RetrieveQuery* and *QueryResult* can be conducted without sending transactions to the smart contract. A node with a full copy of the blockchain storage can locally query the contract status.

4.7.3 On/off Chain Tradeoffs

The on-chain cost can be reduced to constant with one authenticator for the whole spatial and topic databases. However, the on-chain strategy will introduce infeasible off-chain cost,

since off-chain cost is linearly increasing with the input size of $\mathcal{F}_R$ and $\mathcal{F}_T$. We identify two split factors that can quantify the on/off chain tradeoffs: f_r for the spatial database and f_t for the topic database. In *TAVLA*, f_r is set to 1 as $\mathcal{F}_R$ takes the whole spatial index as the input. f_t is set to n_l, as $\mathcal{F}_T$ takes into a subindex I_j of topic index I_R. For very large spatial databases, the spatial index can be split into different subindexes. The split factors serve as the tradeoff switches to tune the on/off chain performance: higher f_t and f_r increase the on-chain overheads and reduce off-chain overheads.

Discussion. To further increase the real-time processing capability [180] of the advertising system, we present measures for real-world implementations. (1) The ad broker can adopt a distributed *SNARGs* to implement *Evaluate* function over the powerful computing clusters. (2) The ad broker can tune the on/off-chain tradeoff by dividing the whole spatial database into sub databases. For example, spatial objects can be organized around some hot spots. (3) It has been proven that the scalability of a blockchain system can be significantly improved with novel consensus protocols, such as Proof-of-Stake or Byzantine Fault Tolerant. We emphasize that our design strategies of *TAVLA* do not rely on specific blockchain architectures, and thus can be implemented with efficient consensus protocols.

4.8 Summary

In this chapter, we have proposed a blockchain-based transparent and accountable vehicular local advertising system. With the two design strategies, the proposed system achieves a notably efficient on/off-chain balance for practical implementations. The experimental results provide comprehensive benchmarks with splitting factors that can be used to tune the on/off chain system performance. The practical designs and observed implementation challenges for the vehicular local advertising system may shed light on general constructions of practical blockchain-based vehicular applications.

Chapter 5

Blockchain-based Smart Advertising Network with Privacy-preserving Accountability

5.1 Background

The technical advances of Internet of Things (IoT) [181] and the next generation wireless technology (5G) [182] are reshaping the advertising industry. Specifically, a smart advertising network (SAN) of connected intelligent objects, such as smart vehicles and smart home devices, can help retailers to effectively reach users through multiple advertising channels. For example, a user can receive advertisements of healthy diets from the smart watch, or promotion codes of nearby shopping centers from the mobile devices. At the same time, with the explosive volumes of data generated by SAN, retailers are able to profile behaviors of their users and personalize their advertisements for different users to improve ad recommendation efficiency. As a result, SAN is surpassing traditional advertising strategies, such as TV and billboard in 2018, and will be dominating the advertising industry in the future [183].

In practice, SAN is managed by a third-party broker, such as Google Ads or Facebook Advertising [33]. With its ubiquitous devices and applications, the broker collects massive user behavior data to build user preference profile. For example, Google records user activities from its ecosystem including Android and Google Home devices, and assigns keyword tags to users based on their activities. Retailers can also choose a set of keywords as their targeting policies and rely on the broker to disseminate their ads to users of related

interests. Later, the broker charges the retailers if any user views the ads (per view) or clicks the links in the ads (per click). By doing so, retailers can enjoy the broker's wealth of user data and advertising channels for effective ad disseminations. As a result, *SAN* has achieved great commercial success. According to eMarketer [183], Google and Facebook occupy a quarter of overall ad spending in the US in 2018.

However, there is an emerging challenge for the continuous success of *SAN*: the lack of advertising transparency [33, 43, 42]. First, users feel offended by the broker when they are unknowingly assigned with keyword tags. For example, a simple click on a link of sports news may give users a tag of 'football'. Second, users often find themselves receiving annoying ads that are irrelevant or even biased. For example, ad dissemination based on gender, age, and nationality is considered as ad discrimination by users [184]. Without proper countermeasures, many users prefer ad-free applications or install ad-block extensions [55] to filter out advertisements. This leads to the retailers' decrease in advertising investments, which greatly hinders the developments of *SAN*.

To regain users' confidence on *SAN*, both the industry and the academic are making efforts on increasing the advertising transparency. An initial attempt by the broker is to provide users with personal profile management tools, such as Google's Ads Setting and Facebook's Ads Manager. The brokers also provide users with options of "transparency explanations" regarding why users are receiving specific ads. For example, Facebook explains to users the sources and targeting policies of the ads. However, such explanations are usually insufficient to users [33] and can sometimes be incomplete and misleading [43]. Moreover, prompt actions against the advertising misconduct are insufficient due to profit consideration and slow internal process of the brokers. At the same time, there have been many research activities, that utilize trusted hardware [46] and transparency extensions [42] at user side or introduce an independent organization to enhance transparency of *SAN* [45].

Although existing works have explored a wide range of technologies, they mainly rely on the trustworthiness of a single authority to provide transparency explanations. Due to profit considerations, the single authority may not always act honestly. For example, it is reported that major brokers pay the developers of ad-blocking tool to make their ads on the 'whitelist' [47]. Therefore, a solution that builds upon the blockchain architecture [169] with distributed consensus [185] is more promising for enhancing transparency in *SAN*. Specifically, the blockchain is a public ledger with blocks of peer-to-peer transactions in a fully distributed network. Secured by the cryptography and consensus protocols [186, 187], the blockchain ensures a consistent and transparent view of the shared ledger among mutually distrustful nodes. If we view the blockchain as a state machine, every valid transaction will change the state of the blockchain. As a result, blockchain can be

utilized as a trusted environment to execute computer programs, i.e., smart contract [81]. Specifically, blockchain nodes can call smart contracts on the blockchain by sending and verifying transactions.

A straightforward blockchain-based solution is that the broker stores all retailer policies and user profiles onto the blockchain and designs an advertising smart contract, that implements the ad dissemination. By doing so, a blockchain-based solution achieves two distinctive features: (1) Decentralized transparency. The ad dissemination is publicly verifiable in a distributed network [30]. (2) Automatic accountability. Any advertising misconduct can be automatically detected and publicly held accountable. However, the straightforward solution may not be practical in real-world implementations due to the following challenges. (1) Efficiency. Since on-chain storage and computation resources are limited, directly implementing the ad dissemination on the blockchain is prohibitively costive [93, 109] for *SAN*. (2) Privacy. User profiles contain sensitive personal information, e.g., locations and interests [188], which may be exposed to the public due to the transparency nature of the blockchain.

In this chapter, we propose a blockchain-based Smart Advertising Network with Privacy-preserving Accountability *(SANPA)*. Specifically, the broker commits to the retailer policies and ad dissemination algorithms with succinct cryptographic authenticators. The authenticators are updated to an accountability contract on the blockchain to serve as a public commitment of advertising transparency. Instead of directly implementing *SAN* on the blockchain, *SANPA* enables the broker to manage the ad dissemination in an off-chain manner. Users and retailers can require transparency explanations about advertising activities, e.g., the management of retailer policies and ad dissemination process, by sending challenges to the accountability contract. With the on-chain cryptographic authenticators, the accountability contract can publicly verify the correctness of the challenged advertising activities without sacrificing user profile privacy. At the same time, the accountability contract can hold any advertising misconduct publicly accountable, i.e. confiscating cryptocurrency deposits of misbehaving parties. By doing so, *SANPA* achieves privacy-preserving accountability for *SAN*. Specifically, the contributions are summarized as follows:

- ▷ We propose a composite *SNARG* system from Quadratic Arithmetic Program (QAP)-based relations and multivariate linear relations in the discrete logarithm setting. The composite *SNARG* system is efficient for on-chain verifications of advertising activities and preserves user profile privacy while pursuing public accountability.

- ▷ We design an accountability contract that receives challenges for transparency explanations and enforces accountability on misbehaving parties. The contract implements

the composite *SNARG* system and uses the cryptocurrencies as incentives to boosting honest advertising conducts and promote prompt on-chain responses.

- ▷ Through the security analysis, we formulate and achieve *privacy-preserving accountability* in *SANPA*. Extensive experiments are conducted to demonstrate the feasibility of *SANPA*. The experimental results present comprehensive benchmarks for both the off-chain and on-chain computation and storage overheads.

The chapter is organized as follows. In Section 5.2, we present the smart advertising model, security model, and design goals. In Section 5.3, we introduce the building blocks. In Section 5.4, we propose *SANPA*, and provide the security analysis in Section 5.5. We evaluate the performance of *SANPA* in Section 5.6. We conclude this chapter in Section 5.7.

5.2 Problem Formulation

In this section, we first formulate the smart advertising model in terms of entities and the ad dissemination strategy. Then, we formalize the security model and the design goals.

5.2.1 Smart Advertising Model

We abstract the existing *SAN* model in Fig. 5.1, which consists of three entities: Broker, Retailer, and User.

- ▷ **User:** Users are equipped with multiple devices, e.g., mobile phones or tablets. They run a wide range of applications and can receive advertisements from multiple channels, e.g., web search or application push messages.
- ▷ **Retailer:** Retailers are shops or stores that wish to promote their products by advertisements. Retailers rely on the broker to manage their targeting policies and pay the broker for the ad dissemination services.
- ▷ **Broker:** Broker is a third-party advertising company (e.g., Google Ads). It manages user preference profiles and retailer targeting policies, and charges retailers based on per-view or per-click model.

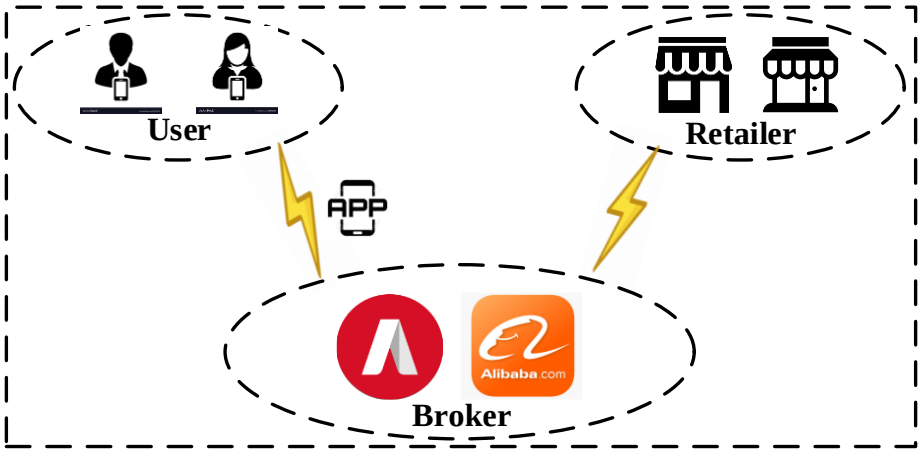


Figure 5.1: Smart Advertising Model

In *SANPA*, we consider the ad dissemination with the popular keyword matching strategy between user profile and retailer targeting policies [10]. Specifically, a keyword dictionary $\mathcal{D} = \{\mathcal{W}_1, \mathcal{W}_2, ..., \mathcal{W}_n\}$ consists of n keywords. n is a few hundreds for a subcategory of the keyword space in Google Ads. The broker can assign each user a set of keywords from the user's interaction with the broker's applications (e.g., Google's Chrome, YouTube and Map), and constructs the user preference profile S_u. At the same time, the user can also access and modify her/his preference profile. Each retailer selects a set of keywords from $\mathcal{D}$ and constructs a targeting policy S_r. The broker measures the similarity between the user profile and retailer policies, and returns the user with advertisements that are most relevant to her preference profile.

5.2.2 Security Model

Users and retailers are both *rational*. That is, either users or retailers will only challenge the advertising system, if there are concerns on the advertising transparency. They will also accept transparency explanations if the explanations are publicly verifiable. The broker is a multi-sector enterprise, that may not always follow the pre-determined ad dissimilation strategy, due to profit considerations, slow internal processes, and the lack of public auditings. Under the security model, we define the security goal as *privacy-preserving accountability* and present its progressive meanings as follows:

Definition 1. *Privacy-preserving Accountability*

- ▷ Public Verifiability: Users and retailers can require the broker to provide advertising transparency explanations about the ad dissemination process and retailer policy management, the correctness of which should be publicly verifiable.
- ▷ Privacy: User preference profiles are concealed from the public view, even in a publicly verifiable transparency explanation.
- ▷ Accountability: Timely and automatic obligations enforcement on the broker should be achieved in case of any advertising misconduct.

5.2.3 Design Goals

SANPA should achieve the following design goals:

- ▷ Compatibility: *SANPA* should support the ad dissemination with the keyword matching strategy.
- ▷ Security: *SANPA* should achieve *privacy-preserving accountability* for the smart advertising network.
- ▷ Efficiency: *SANPA* should incur applicable overhead to the smart advertising network.

5.3 Preliminaries

In this section, we present the preliminaries in *SANPA*, including cryptographic commitment schemes, SNARG systems, and digital signature schemes. (1) The cryptographic commitment is utilized to securely digest targeting policies and user profiles into a succinct authenticator. (2) SNARG systems can achieve verifiable on-chain transparency explanations. (3) Digital signature is used to generate non-repudiable off-chain receipts of advertising activities. Notations are shown in Table 5.1.

Table 5.1: Notations

$\mathbb{G}$	Multiplicative groups
$\mathcal{R}$	Polynomial-time decidable relation
(x, w)	Statement x, witness w
$\mathbf{x}^n$	n-dimension vector
$\mathbf{X}^{m*n}$	$m * n$-dimension matrix
$[n]$	Integers from 1 to n
$\in_R$	Choose a random number
$\mathbf{Com}(\mathbf{x}, CK)$	Cryptographic commitment **Com** Input vector $\mathbf{x}$, commitment key CK
$\mathbb{C} = (EK, VK)$	Common reference string $\mathbb{C}$ Evaluation key EK, verification key VK

5.3.1 Notations

$\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T$ denote three cyclic multiplicative groups [189] with a prime order p and a bilinear pairing $e : \mathbb{G}_1 \times \mathbb{G}_2 \longrightarrow \mathbb{G}_T$. $\mathbb{Z}_p$ denotes a ring of integers modules p. $r \in_R \mathbb{Z}_p$ indicates r is randomly chosen from $\mathbb{Z}_p$. $\mathbb{F}$ denotes a finite field. $[n]$ denotes integers from 1 to n. A bold lower letter $\mathbf{x}^n \in \mathbb{F}^n$ denotes an n-dimension vector from $\mathbb{F}$. A bold capital letter $\mathbf{X}^{m*n} \in \mathbb{F}^{m*n}$ denotes an $m*n$ matrix from $\mathbb{F}$. From the theory of computation [144], we denote $\mathcal{R}$ as a polynomial-time decidable relation with a statement x and a witness w. $(x, w) \in \mathcal{R}$ indicates that $\mathcal{R}$ holds on a pair (x, w), which can be efficiently decided by a non-interactive argument system [143].

5.3.2 Cryptographic Commitment

Cryptographic commitment schemes [125] allow a party to commit to a secret value, such as an integer or a vector of integers. The commitment can either be directly revealed to the public, or combined with the zero-knowledge proof technique [190] to demonstrate that the committed value satisfies a public relation without leaking the value. We define the cryptographic commitment as follows:

Definition 2. A cryptographic commitment of value x using a commitment key CK is denoted as $\mathbf{Com}(x, CK)$.

A typical example of commitment schemes in the cyclic multiplicative groups is Pedersen commitment. Given an $x, r \in_R \mathbb{Z}_p^2$, and $CK = (g_1, g_2) \in \mathbb{G}_1^2$, a Pedersen commitment $\mathbf{Com}(x, CK) = g_1^x g_2^r$. Given a vector $\mathbf{x}_n$, $r \in_R \mathbb{Z}_p$, and $CK' = (g, g_1, g_2, ..., g_n) \in \mathbb{G}_1^{n+1}$, an extended Pedersen commitment $\mathbf{Com}(\mathbf{x}^n, CK') = g^r \prod_{i=1}^n g_i^{x_i}$.

5.3.3 Succinct Non-interactive ARGuments (*SNARG*)

The *SNARG* system allows a prover to demonstrate that a public relation $\mathcal{R}$ holds on a pair (x, w) to a verifier, which can be defined as follows:

Definition 3. A *SNARG* system $\sum$ for an NP-complete relation $\mathcal{R}$ consists of three algorithms:

- ▷ $KeyGen(\mathcal{R}, pp) \rightarrow \mathbb{C} = (EK, VK)$
- ▷ $Prove(EK, x, w) \rightarrow \pi$
- ▷ $Verify(VK, x, w) \rightarrow (0, 1)$

$KeyGen$ takes as inputs the relation $\mathcal{R}$ and public system parameters pp, and outputs the common reference string $\mathbb{C}$ with EK and VK. $Prove$ takes as inputs the EK, a statement x, and a witness w. $Prove$ evaluates $\mathcal{R}$ on (x, w) and generates a proof π. $Verify$ takes VK, the statement x, and the proof π. It outputs 1 if $(x, w) \in \mathcal{R}$; it outputs 0, otherwise. We define four security notions of *SNARG* systems as follows:

- ▷ *Completeness*: A rational verifier will accept (x, π) if $(x, w) \in \mathcal{R}$ and π is correctly computed.
- ▷ *Soundness*: A computationally-bounded adversary cannot forge an invalid tuple (x', w', π'), such that $(x', w') \notin \mathcal{R}$ and $Verify(VK, x', \pi') \rightarrow 1$.
- ▷ *Succinctness*: The proof length is only determined by the system security parameter.
- ▷ *Privacy Preservation*: The verifier only learns whether $(x, w) \in \mathcal{R}$.

For different relations $\mathcal{R}$, *SNARG* systems can be categorized into subsets with different instantiations and security notions. In *SANPA*, we consider two categories: (1) Multivariate linear relations in the discrete logarithm (DLog) setting. (2) Quadratic Arithmetic Program (*QAP*) based *SNARG* in bilinear groups. We do not distinguish the pre-processing *CRS* model for the *QAP*-based *SNARG* and the classical *CRS* model for *SNARG* in the discrete logarithm setting.

Multivariate Linear Relations in the DLog

The *SNARG* system for multivariate linear relations can be efficiently recognized by discrete logarithms in multiplicative groups with a prime order [133]. In *SANPA*, we focus on two specific *SNARG* systems in the DLog: (1) the equality test for two Pedersen vector commitments, and (2) the succinct openness for a subset of a Pedersen vector commitment.

Definition 4. Two extended Pedersen commitments for the same vector $\mathbf{x}^n$ with different commitment keys CK_1, CK_2 are defined as follows:

$$\mathbf{Com}(\mathbf{x}^n, CK_1) = \prod_{i=1}^{n} g_i^{x_i},\ CK_1 = (g_1, ..., g_n) \in \mathbb{G}_1^n$$

$$\mathbf{Com}'(\mathbf{x}^n, CK_2) = \prod_{i=1}^{n} g_i'^{x_i},\ CK_2 = (g_1', ..., g_n') \in \mathbb{G}_1^n$$

A *SNARG* system $\sum_{\mathcal{V}}$ enables a prover with knowledge $\mathbf{x}^n$ to convince a verifier that $\mathbf{Com}$ and $\mathbf{Com}'$ open to the same vector $\mathbf{x}^n$.

Definition 5. Consider $\mathbf{R}^{m*n} \in \mathbb{Z}_P^{m*n}$, $\mathbf{y}^n \in \mathbb{Z}_P^n$ is a subset of $\mathbf{R}^{m*n}$ indexed by $I_S = \{(i_1, j_1), (i_2, j_2), ..., (i_n, j_n)\}$, a Pedersen commitment for $\mathbf{R}$ is defined as follows:

$$\mathbf{Com}(\mathbf{R}, CK) = \prod_{i=1}^{m}\prod_{j=1}^{n} g_{i,j}^{\mathbf{R}_{i,j}},\ CK = \{g_{i,j}\} \in \mathbb{G}_1^{m*n}$$

A *SNARG* system $\sum_{\mathcal{S}}$ enables a prover with knowledge $\mathbf{R}$ to convince a verifier: the subset of $\mathbf{Com}(\mathbf{R}, CK)$ indexed by I_S opens to $\mathbf{y}^n$.

We change the position of randomness r in the original Pedersen commitment to a position in $\mathbf{x}$ or $\mathbf{R}$. Both $\sum_{\mathcal{V}}$ and $\sum_{\mathcal{S}}$ achieve *completeness*, *soundness* and *Succinctness* for secure and efficient verifications. We will present the detailed designs of $\sum_{\mathcal{V}}$ and $\sum_{\mathcal{S}}$ in Section 5.4.

QAP-based *SNARG* in Bilinear Groups

Evaluations of $\mathcal{R}$ on a pair (x, w) is equivalent to the circuit satisfiability evaluations with certain inputs. Gennaro et al. [134] proposed a technique to convert the evaluation of an arithmetic circuit $\mathcal{C}$ to the divisibility check of a Quadratic Arithmetic Program (QAP) $\mathcal{Q}$. In specific, $\mathcal{Q}$ consists of three sets of polynomials $V = \{v_k(x)\}$, $W = \{w_k(x)\}$, $Y = \{y_k(x)\}$, where $k \in [0, z]$ and z denotes the number of input, intermediate and output wires in $\mathcal{C}$. A target polynomial $t(x)$ is defined by picking a random root for each multiplication gate in $\mathcal{C}$. An input $(a_1, a_2, ..., a_o) \in \mathbb{F}^o$ and an output $(a_{z-p+1}, a_{z-p+2}, ..., a_z) \in \mathbb{F}^p$ are valid assignments of $\mathcal{C}$, iff $(a_{o+1}, a_{o+2}, ..., a_{o+q}) \in \mathbb{F}^q$ can be found such that $t(x)$ can divide $p(x)$, where $z = o + p + q$.

$$\begin{aligned} p(x) =& (v_0(x) + \sum_{k=1}^{z} a_k v_k(x)) \times (w_0(x) + \sum_{k=1}^{z} a_k w_k(x)) \\ & - (y_0(x) + \sum_{k=1}^{z} a_k y_k(x)) \end{aligned} \tag{5.1}$$

$(a_{o+1}, a_{o+2}, ..., a_{o+q})$ actually denote the assigned values of intermediate multiplication wires.

We define a relation $\mathcal{R}_{\mathcal{Q}}$ that decides on a pair $(x, \mathbf{w}^z)$ for an arithmetic circuit $\mathcal{C}$. $\mathbf{w}^z \in \mathbb{F}^z$ corresponds to input, output and intermediate multiplication wires of $\mathcal{C}$. Based on Equation 1, $\mathcal{R}_Q$ is represented by a linear combination of $\mathbf{w}^z$, which can be efficiently evaluated in pairing-friendly bilinear groups. Adopting techniques from [122, 138, 141], we obtain a *SNARG* system $\sum_{\mathcal{Q}}$ for $\mathcal{R}_{\mathcal{Q}}$. $\sum_{\mathcal{Q}}$ is *complete*, *sound* and *succinct*.

5.3.4 Digital Signature

A digital signature scheme consists of three algorithms:

- $Gen(\mathbb{G}, 1^\lambda) \rightarrow (pk, sk)$
- $Sig(m)_{sk} \rightarrow \pi_s$

$\triangleright$ $Veri(m, \pi_s)_{pk} \rightarrow (0, 1)$

$KeyGen$ takes into $\mathbb{G}$ and the security parameter λ, and outputs a public/private key pair (pk, sk). Sig takes into a message m and a secret key sk, and outputs a signature π_s. $Veri$ takes into a message and a signature. It outputs 1 if the verification passes; it outputs 0, otherwise. We utilize *ECDSA* signature [79] in *SANPA*, which is compatible in the Ethereum.

5.4 Smart Advertising Network with Privacy-preserving Accountability

In this section, we first give an overview of *SANPA* including *System Model*, *Design Ideas*, and *Workflow*. Then, we present the details of *SANPA*, in terms of *Initialization*, *Off-chain Smart Advertising* and *On-chain Transparency Explanation*.

5.4.1 Overview

System Model

In *SANPA*, we introduce two additional entities to *SAN*: a distributed committee (DC) and a public blockchain in Fig. 5.2. (1) DC can be a set of independent supervising authorities running a Secure Multi-party Computation (SMC) protocol. For example, Zerocash has implemented an SMC protocol [174] to setup the blockchain system. (2) Blockchain is a public ledger, e.g., Ethereum. It is maintained by peer-to-peer blockchain miners and supports secure and automatic executions of smart contracts. We also assume that secure and authenticated off-chain channels are established among all entities.

Design Ideas

SANPA introduces an on/off chain computation model for the blockchain-based architecture: (1) User profile and retailer policy managements, and the ad dissemination are conducted in an off-chain manner by the ad broker. *SANPA* requires each advertising activity is non-repudiable by the broker, which can be achieved by using digital signatures. (2) Broker advertising activities are publicly audited with effective accountability enforcements in an on-chain manner, if required by users or retailers. (3) User profile privacy is guaranteed

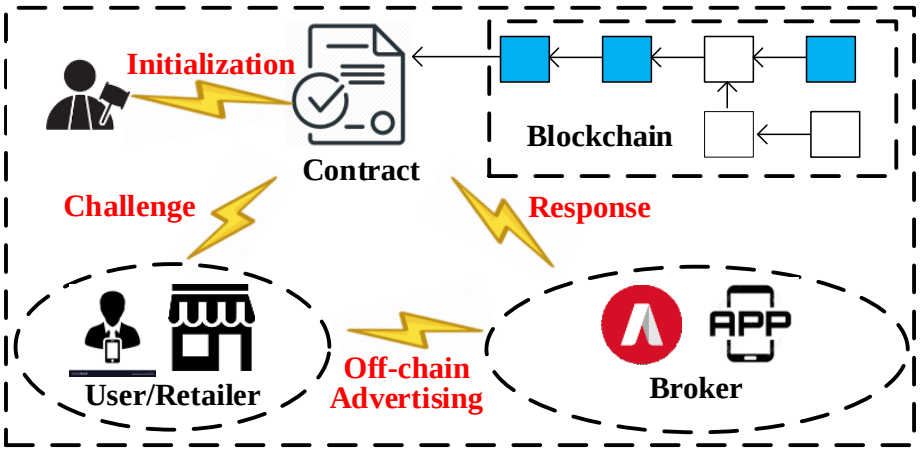


Figure 5.2: SANPA Workflow

in the public transparency explanations with the design of a composite SNARG system. By doing so, *SANPA* achieves distributed, efficient, and privacy-preserving transparency explanations.

Workflow

In Fig. 5.2, *SANPA* consists of the following three phases:

(1) *Initialization.* DC sets up the advertising policies by initializing a set of *SNARG* systems $\sum_{\mathcal{Q}}$, $\sum_{\mathcal{V}}$ and $\sum_{\mathcal{S}}$. $\sum_{\mathcal{S}}$ verifies that retailer policies are correctly managed by the ad broker. $\sum_{\mathcal{V}}$ verifies that the retailer policies are correctly entered into the ad dissemination process. $\sum_{\mathcal{Q}}$ verifies that the ad dissemination process follows the pre-determined keyword matching strategies. DC creates an accountability contract to store evaluation keys of the *SNARG* systems. More details are given in Section 5.4.2.

(2) *Off-chain smart advertising.* Users and retailers register themselves at the broker. Users manage their preference profiles generated by the broker. Retailers set their targeting policies. The broker runs the ad dissemination process with the keyword matching strategy between the user preference profile and retailer targeting policies. The broker finds most relevant advertisements and sends them to the user.

(3) *On-chain transparency explanation.* Users and retailers can make challenges to the accountability contract and require transparency explanations on broker activities, e.g., the ad dissemination process and the retailer policy management. The broker must response

to the challenges promptly, by updating correctness proofs of the advertising activities to the contract. The contract verifies the proofs from the broker and enforces obligations if any advertising misconduct is identified.

5.4.2 Initialization

DC chooses a system security parameter λ and a set of asymmetric multiplicative groups $\mathbb{G} = (\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T)$ with a prime order p and a bilinear pairing e. $g \in \mathbb{G}_1$ and $\tilde{g} \in \mathbb{G}_2$ are two random generators. DC sets the system public parameter $pp = (\mathbb{G}, p, e, g, \tilde{g})$. DC sets a composite relation as follows:

$$\begin{cases} (S_u, S_{\mathcal{R}}, S_o) \in \mathcal{R}_{\mathcal{Q}} \\ \wedge (D_{\mathcal{R}} = \mathbf{Com}(S_{\mathcal{R}}, CK_{\mathcal{R}}), D_{\bar{\mathcal{R}}} = \mathbf{Com}(S_{\mathcal{R}}, CK_{\bar{\mathcal{R}}})) \in \mathcal{R}_{\mathcal{V}} \\ \wedge (D_{\mathcal{R}}, S_r, I_S) \in \mathcal{R}_{\mathcal{S}} \end{cases} \tag{5.2}$$

DC abstracts the ad dissemination process in Algorithm 6 as a relation $\mathcal{R}_{\mathcal{Q}}$ on a pair $(S_u \in \mathbb{Z}_P^n, S_{\mathcal{R}} \in \mathbb{Z}_P^{m*n}, S_o \in \mathbb{Z}_P^k)$. S_u corresponds to an n-dimension user profile. $S_{\mathcal{R}}$ corresponds to m retailer targeting policies, each of which is also n-dimensional. The process outputs S_o, that consists of k retailer identifiers with the most relevant keywords to the user profile. $D_{\mathcal{R}}, D_{\bar{\mathcal{R}}}$ are commitments generated under different commitment keys $CK_{\mathcal{R}}, CK_{\bar{\mathcal{R}}}$. $S_r \in S_{\mathcal{R}}$ is the keyword set for an individual retailer indexed by I_S.

DC initializes the above relations with *SNARG* systems $\sum_{\mathcal{Q}}, \sum_{\mathcal{V}}$, and $\sum_{\mathcal{S}}$. The evaluation of the relation $\mathcal{R}_{\mathcal{Q}}$ is achieved by the design of the $\sum_{\mathcal{Q}}$. To preserve user profile privacy, the $Verify$ algorithm of $\sum_{\mathcal{Q}}$ takes a commitment $D_{\bar{uo}}$ (a commitment of S_u and S_o) and a commitment $D_{\bar{\mathcal{R}}}$ of retailer policies. However, the original commitment $D_{\bar{\mathcal{R}}}$ in the $\sum_{\mathcal{Q}}$ does not support efficient verifications of retailer challenges. Therefore, a *SANRG* system $\sum_{\mathcal{S}}$ is designed to succinctly reveal the retailer policy from a well-structured external commitment $D_{\mathcal{R}}$. At the same time, *SANPA* proves that $D_{\bar{\mathcal{R}}}$ and $D_{\mathcal{R}}$ open to the same vector commitments with the design of the $\sum_{\mathcal{V}}$.

$\sum_{\mathcal{Q}}$ CRS Setup

DC instantiates $\sum_{\mathcal{Q}}$ as a Pinocchio *SNARG* system [122], with the following three algorithms:

$\triangleright$ $KeyGen(\mathcal{R}_{\mathcal{Q}}, pp) \rightarrow \mathbb{C}_{\mathcal{Q}} = (EK_{\mathcal{Q}}, VK_{\mathcal{Q}})$

$\triangleright$ $Prove(EK_{\mathcal{Q}}, (S_u, S_{\mathcal{R}}, S_o)) \rightarrow \pi_{\mathcal{Q}}$

$\triangleright$ $Verify(VK_{\mathcal{Q}}, \pi_{\mathcal{Q}}) \rightarrow (0, 1)$

$KeyGen$ generates CRS $\mathbb{C}_{\mathcal{Q}}$ with an evaluation key $EK_{\mathcal{Q}}$ and a verification key $VK_{\mathcal{Q}}$. $Prove$ evaluates $(S_u, S_{\mathcal{R}}, S_o)$ and outputs a correctness proof $\pi_{\mathcal{Q}}$. $Verify$ outputs 1 if $(S_u, S_{\mathcal{R}}, S_o) \in \mathcal{R}_{\mathcal{Q}}$; Otherwise, it outputs 0. We omit the detailed constructions of the three algorithms, but note that the existence of Pedersen-like commitment keys ($CK_{\bar{u}} \in \mathbb{G}_1^n, CK_{\bar{\mathcal{R}}} \in \mathbb{G}_1^{m*n}, CK_{\bar{o}} \in \mathbb{G}_1^k$) $\in EK_{\mathcal{Q}}$ and Pedersen commitments ($D_{\bar{u}o} \in \mathbb{G}_1, D_{\bar{\mathcal{R}}} \in \mathbb{G}_1$) $\in \pi_{\mathcal{Q}}$.

$\sum_{\mathcal{V}}$ CRS Setup

DC chooses a set of random numbers $\mathbf{Z} \in_R \mathbb{Z}_P^{m*n}$, and computes $CK_{\mathcal{R}} = \{CK_{\mathcal{R}_{i,j}} = g^{\mathbf{Z}_{i,j}}\}_{i\in[m],j\in[n]} \in \mathbb{G}_1^{m*n}$. DC chooses random generators $R = \{R_{i,j}\}_{i\in[m],j\in[n]} \in_R \mathbb{G}_1^{m*n}$ and $\alpha, \beta, \gamma \in_R \mathbb{Z}_P^3$. DC computes $A = \tilde{g}^{\alpha}, B = \tilde{g}^{\beta}, C = \tilde{g}^{\gamma}$. DC computes $T_{i,j} = CK_{\mathcal{R}_{i,j}}^{\alpha} R_{i,j}^{\beta} CK_{\bar{\mathcal{R}}_{i,j}}^{\gamma}$. DC sets $T = \{T_{i,j}\} \in \mathbb{G}_1^{m*n}$. The CRS of $\sum_{\mathcal{V}}$ is as follows:

$$EK_{\mathcal{V}} = (R, T), VK_{\mathcal{V}} = (A, B, C)$$

$\sum_{\mathcal{S}}$ CRS Setup

DC computes $\widetilde{CK}_{\mathcal{R}_{i,j}} = \tilde{g}^{\mathbf{Z}_{i,j}}, \forall i \in [m], j \in [n]$ and sets $\widetilde{CK}_{\mathcal{R}} = \{\widetilde{CK}_{\mathcal{R}_{i,j}}\} \in \mathbb{G}_2^{m*n}$. DC computes $EK_{\mathcal{S}_{(i,j)(i',j')}} = g^{\mathbf{Z}_{i,j}\mathbf{Z}_{i'j'}}, \forall (i, i') \in [m],\ (j, j') \in [n],\ (i, j) \neq (i', j')$. DC sets $EK_{\mathcal{S}} = \{EK_{\mathcal{S}_{(i,j)(i',j')}}\} \in \mathbb{G}_1^{m^2n^2-mn}$.

CRS Distribution

DC sends $\{pp, CK_{\mathcal{R}}, \widetilde{CK}_{\mathcal{R}}, CK_{\bar{u}}, CK_{\bar{o}}\}$ to users and retailers, and $\{pp, EK_{\mathcal{Q}}, EK_{\mathcal{V}}, EK_{\mathcal{S}}, CK_{\mathcal{R}}, \widetilde{CK}_{\mathcal{R}}\}$ to the broker.

For $\sum_{\mathcal{S}}$ and $\sum_{\mathcal{V}}$ from the DLog, the common reference string consists of sets of non-identical generators, whose well-formedness can be easily checked by the public. For $\sum_{\mathcal{Q}}$ from the QAP theorem, a trapdoor secret is used to generate the evaluation and verification keys, which can be securely computed by DC with a SMC protocol or must be securely destroyed if the secret is generated by a single authority.

5.4.3 Off-chain Smart Advertising

The off-chain smart advertising consists of *Registration*, *Retailer Policy Archiving* and *Ad Dissemination*. We assume secure and authenticated communication channels [191] are established among users, retailers and the broker.

Registration

The broker registers itself at DC with a public key $pk_{\mathcal{A}}$ of *ECDSA* signature and a blockchain address $addr_{\mathcal{A}}$. DC validates the identity of the broker and the well-formedness of $pk_{\mathcal{A}}$.

A user registers herself/himself at the broker with a universal ID ID_u and a public key pk_u of *ECDSA* signature. Similar to preference tag management in Google Ads, the user can obtain a set of keywords provided by the broker from the keyword dictionary $\mathcal{D} = \{\mathcal{W}_1, \mathcal{W}_2, ..., \mathcal{W}_n\}$. The user further sets a preference profile S_u as follows:

$$S_u = \{K_{u,i}\}_{i\in[n]} \begin{cases} K_{u,i} = 0, \ if \ \mathcal{W}_i \ is \ not \ selected \\ K_{u,i} = r_i, \ r_i \in_R \mathbb{Z}_P, \ otherwise \end{cases} \tag{5.3}$$

The user sends S_u to the broker. The broker sets $m_u = (ID_u, S_u, T_u)$, where T_u is a valid time stamp. The broker computes a signature as $Sig_{sk_{\mathcal{A}}}(m_u) \rightarrow \pi_u$ and sets the evidence of user preference profile as follows:

$$Evid_u = (m_u, \pi_u) \tag{5.4}$$

The broker sends $Evid_u$ to the user and stores (ID_u, S_u, pk_u) at its storage. The user checks S_u and $Veri(m_u, \pi_u)_{pk_{\mathcal{A}}} \rightarrow 1$ and sends back an acknowledgement to the broker if the checks pass.

A retailer registers herself/himself at the broker, with a universal ID ID_r, a public signature key pk_r and a targeting policy S_r as follows:

$$S_r = \{K_{r,i}\}_{i\in[n]} \begin{cases} K_{r,i} = 0, \ if \ \mathcal{W}_i \ is \ not \ selected \\ K_{r,i} = r_i, r_i \in_R \mathbb{Z}_P, \ otherwise \end{cases} \tag{5.5}$$

For representation simplicity, we assume that there are m retailers in *SANPA* that are sequentially indexed. We denote $r \in [m]$ as the index number of the retailer ID_r, which means that $ID_r = r$. The broker sets $VK_r = \{CK_{\mathcal{R}_{r,j}}\}_{j\in[n]} \in \mathbb{G}_1^n$, where $CK_{\mathcal{R}_{r,j}}$ is the (r, j)-th item in $CK_{\mathcal{R}}$. Similarly, the broker sets $\widetilde{VK}_r = \{\widetilde{CK}_{\mathcal{R}_{r,j}}\}_{j\in[n]} \in \mathbb{G}_2^n$. The broker

sets $m_r = (ID_r, S_r, VK_r, \widetilde{VK}_r, T_r)$, where T_r is a time stamp. The broker computes a signature $\pi_r = Sig_{sk_{\mathcal{A}}}(m_r)$ and sets the evidence of the retailer policy $Evid_r$ as follows:

$$Evid_r = (m_r, \pi_r) \tag{5.6}$$

The broker sends the $Evid_r$ to the retailer and stores (ID_r, S_r, pk_r) at its local storage. The retailer checks that VK_r and $\widetilde{VK}_r$ are correctly chosen from $CK_{\mathcal{R}}$ and $\widetilde{CK}_{\mathcal{R}}$, and $Veri(m_r, \pi_r)_{pk_{\mathcal{A}}} \rightarrow 1$. The retailer sends an acknowledgement to the broker if all checks pass.

Retailer Policy Archiving

The broker collects targeting policies from m retailers as $S_{\mathcal{R}} = (S_1, S_2, ..., S_m)$. We denote $K_{i,j}$ as the j-th item in the keyword set of the i-th retailer S_i. The broker computes a cryptographic commitment $D_{\mathcal{R}}$ as follows:

$$D_{\mathcal{R}} = \prod_{i=1}^{m} \prod_{j=1}^{n} CK_{\mathcal{R}_{i,j}}^{K_{i,j}} \tag{5.7}$$

The broker uploads $D_{\mathcal{R}}$ to the accountability contract.

Ad Dissemination

The broker generates the non-repudiable evidence of ad disseminations for usres. Specifically, the ad dissemination in *SANPA* works with the following steps:

1. The user generates an ad request $m_{sid} = (sid, ID_u, T_{sid})$, where sid is a session id and T_{sid} is a time stamp. The user computes $\pi_{sid} = Sig(m_{sid})_{sk_u}$ and sends (m_{sid}, π_{sid}) to the broker.

2. The broker checks the time stamp, the user ID ID_u and the freshness of sid. If $Veri(m_{sid}, \pi_{sid})_{pk_u} \rightarrow 1$, the broker retrieves user preference profile S_u and conducts the ad dissemination in Algorithm 6.

3. The broker computes:

$$\begin{aligned} &D_{\bar{o}} = \prod_{i=1}^{k} CK_{\bar{o}_i}^{ID_i \in S_o}, D_{\bar{u}} = \prod_{i=1}^{n} CK_{\bar{u}_i}^{K_{u,i}} \\ &D_{\bar{u}o} = D_{\bar{u}} D_{\bar{o}}, \pi'_{sid} = Sig_{sk_{\mathcal{A}}}(m'_{sid}) \end{aligned} \tag{5.8}$$

Algorithm 6: Ad Dissemination with Exact Keyword Matching

Input: User profile S_u, retailer policies $S_{\mathcal{R}}$
Output: Retailer Identifier Set S_o
Set S_o, S_{temp} to be empty
for $S_i \in S_{\mathcal{R}}$ **do**
    Set $flag$ to be 1
    for $K_{u,j} \in S_u$ **do**
        if $K_{u,j} \neq 0$ & $K_{i,j} = 0$ **then**
            Set $flag = 0$
    if $flag = 1$ **then**
        Add the retailer identifier ID_i to S_{temp}
Add k identifiers of S_{temp} to S_o

$m'_{sid} = (D_{\bar{uo}}, m_{sid}, T'_{sid})$. The broker stores (ID_u, sid, S_o) at its storage and sends $(S_o, m'_{sid}, \pi'_{sid})$ to the user.

4. If $Veri(m'_{sid}, \pi'_{sid})_{pk_{\mathcal{A}}} \rightarrow 1$ and $D_{\bar{uo}}$ is correctly computed with $S_u, S_o, CK_{\bar{u}}, CK_{\bar{o}}$, the user sets $Evid_{sid} = (m'_{sid}, \pi_{sid}, \pi'_{sid})$. The user also generates an acknowledgement for the broker to charge the retailer.

5.4.4 On-chain Transparency Explanation

In this section, we first define two types of challenges about the broker advertising activities. Second, we present an overview of the accountability contract. Third, we present the design details of the accountability contract.

Advertising Challenge

Two types of transparency challenges are designed in *SANPA*: (1) *Retailer Challenge*: A retailer can make a challenge with $Evid_r$ to the accountability contract to check whether the targeting policy S_r is correctly included in the on-chain authenticator $D_{\mathcal{R}}$. (2) *User Challenge*: A user can make a challenge with $Evid_{sid}$ to the accountability contract to check whether the ad dissemination process is correctly conducted.

Overview

The accountability contract consists of four phases: *Challenge*, *Prove*, *Resolve* and *Claim*.

- ▷ *Initialization.* DC initializes system public paraders and creates the accountability smart contract.
- ▷ *Challenge.* Retailers or users send their evidence to the accountability contract to require transparency explanations for specific advertising activities.
- ▷ *Resolve.* The broker retrieves the challenges from the accountability contract and proves the correctness of the activities within a pre-determined threshold time. The broker uploads the correctness proof of the challenged activities to the contract and claims the deposits of the challenger if the proof passes the verification.
- ▷ *Claim.* If the broker does not provide the proof within the threshold time, users or retailers can send a request to the accountability contract and claim deposits of the broker.

Accountability Contract

We utilize the public Ethereum blockchain to implement the accountability contract. In *SANPA*, we require that DC is associated with an Ethereum blockchain address that is known to the public. In the following, we present the detailed designs of the above four phases:

Initialization. The broker registers its public key $pk_{\mathcal{A}}$ and blockchain address $addr_{\mathcal{A}}$ at DC. DC initializes the *SNARG* systems for the on-chain transparency explanations. For the retailer challenge, DC instantiates the *SNARG* system $\sum_{\mathcal{S}}$, that is to prove a retailer's targeting policy S_r is correctly included in $D_{\mathcal{R}}$. For the user challenge, *SANPA* designs a composite *SNARG* system with $\sum_{\mathcal{Q}}$ and $\sum_{\mathcal{V}}$. DC creates the accountability contract in Algorithm 7 and 8. The contract stores the broker's public key and blockchain address, the public system parameters, evaluation keys of *SNARG* systems, and the authenticator of retailer policies $D_{\mathcal{R}}$. The contract also stores a threshold time T_H and takes initial deposits $C_{\mathcal{A}}$ from the broker.

Challenge. Users and retailers make challenges of advertising activities to the accountability contract. *UserChallenge* function takes evidence and deposits C_u from users.

RetailerChallenge function takes $Evid_r$ and deposits C_r from a retailer. Both of the functions will verify the signatures of the evidence, set the time stamp for the challenge, and record the valid evidence on the contract.

Resolve. The broker retrieves unprocessed user challenges and conducts off-chain processing via *Prove* algorithms of *SNARG* systems $\sum_{\mathcal{Q}}$ and $\sum_{\mathcal{V}}$ as follows:

- ▷ From Rec_U, the broker obtains ID_u and $Evid_{sid}$. The broker retrieves S_u with ID_u and $S_{\mathcal{R}}$ from its storage. S_u is the user preference profile and $S_{\mathcal{R}}$ is the keyword sets of all retailers.
- ▷ The broker runs $Prove(EK_{\mathcal{Q}}, (S_u, S_{\mathcal{R}}, S_o))$ to obtain a proof $\pi_{\mathcal{Q}}$ for $\mathcal{R}_{\mathcal{Q}}$. Note that, there are commitments $D_{\bar{\mathcal{R}}} = \mathbf{Com}(S_{\mathcal{R}}, CK_{\bar{\mathcal{R}}}), D_{\bar{u}\bar{o}} = \mathbf{Com}(S_u, CK_{\bar{u}}) \cdot \mathbf{Com}(S_o, CK_{\bar{o}}) \in \pi_{\mathcal{Q}}$.
- ▷ With $(T, R) \in EK_{\mathcal{V}}$, the broker runs *Prove* function of $\sum_{\mathcal{V}}$ as follows:

$$X = \prod_{i=1}^{m} \prod_{j=1}^{n} T_{i,j}^{K_{i,j}}, Y = \prod_{i=1}^{m} \prod_{j=1}^{n} R_{i,j}^{K_{i,j}}, K_{i,j} \in S_{\mathcal{R}} \tag{5.9}$$

- ▷ The broker sets $\pi_{\mathcal{V}} = (X, Y)$.

The broker retrieves unprocessed retailer challenges and conducts *Prove* algorithm of $\sum_S$. Specifically, the broker obtains ID_r, S_r from $Evid_r$ and $EK_{\mathcal{S}}$ from its storage. We denote the index I_S for S_r as $\{(r,1),(r,2),...,(r,n)\}$ and $\mathbf{R}^{m*n}$ as $\{(1,1),(1,2),...,(m-1,n),(m,n)\}$. The broker computes a proof π_S as follows:

$$\pi_S = \prod_{(i,j) \in I_S} \prod_{(i',j') \in \mathbf{R}^{m*n} \setminus I_S} EK_{\mathcal{S}_{(i,j)(i',j')}}^{K_{i',j'}} \tag{5.10}$$

The broker uploads $(\pi_V, \pi_{\mathcal{Q}})$ or π_S to the accountability contract by calling *UChallengeResolve* or *RChallengeResolve* functions. Both functions retrieve unprocessed challenges from the blockchain storage, and check the correctness and freshness of the proofs. If the proof is correct, the deposits from users/retailers will be transferred to the broker; Otherwise, users/retailers will take deposits from the broker.

Claim. Users/retailers can *UserClaim* or *RetailerClaim* to claim broker deposits, if they do not receive a timely response. The functions check the threshold time T_H and the processing status $flag$, and transfer the broker deposits to users/retailers if a response delay is identified.

Algorithm 7: Accountability Contract - Part I

Require: $addr_{\mathcal{A}}$, $pk_{\mathcal{A}}$, pp, $VK_{\mathcal{Q}}$,$VK_{\mathcal{V}}$, $D_{\mathcal{R}}$, T_H, $C_{\mathcal{A}}$

Set Rec_U, Rec_R to be *empty*

Function `UserChallenge`($Evid_{sid}$, *deposit* C_u)

- Check $Veri(m_{sid}, \pi_{sid})_{pk_{\mathcal{A}}} \rightarrow 1$
- Check $Veri(m'_{sid}, \pi'_{sid})_{pk_{\mathcal{A}}} \rightarrow 1$
- Check $ID_u||sid \nsubseteq Rec_U$
- Set $addr_u = message\ sender,\ flag = 0$
- Set $T_{recv} = block.time$
- Add the following to Rec_U:
- $(ID_u||sid, addr_u, Evid_{sid}, T_{recv}, flag)$

Function `RetailerChallenge`($Evid_r$, *deposit* C_r)

- Check $Veri(m_r, \pi_r)_{pk_{\mathcal{A}}} \rightarrow 1$
- Check $ID_r \nsubseteq Rec_R$
- Set $addr_r = message\ sender, flag = 0$
- Set $T_{recv} = block.time$
- Add $(ID_r, addr_r, Evid_r, T_{recv}, flag)$ to Rec_R

Function `UserClaim`(ID_u, sid)

- Retrieve tuples from Rec_U by $ID_u||sid$
- Check $addr_u = message\ sender,\ flag = 0$
- **if** $block.time - T_{recv} > T_H$ **then**
  - Transfer C'_u to $addr_u$

Function `RetailerClaim`(ID_r)

- Retrieve tuples from Rec_R by ID_r
- Check $addr_r = message\ sender,\ flag = 0$
- **if** $block.time - T_{recv} > T_H$ **then**
  - Transfer C'_r to $addr_r$

Algorithm 8: Accountability Contract - Part II

Function `UChalResolve`$(ID_u, sid, \pi_{\mathcal{V}}, \pi_{\mathcal{Q}})$

- Check *message sender* $= addr_{\mathcal{A}}$
- Retrieve $Evid_{sid}, flag, T_{recv}$ from Rec_U
- Check $flag = 0$ and $block.time - T_{recv} < T_H$
- Check $(D_{\bar{uo}} \in Evid_{sid}) = (D_{\bar{uo}} \in \pi_{\mathcal{Q}})$
- Check $Verify(VK_{\mathcal{Q}}, \pi_{\mathcal{Q}}) \to 1$
- Check $e(X, \tilde{g}) = e(D_{\mathcal{R}}, A)e(Y, B)e(D_{\bar{\mathcal{R}}}, C)$
- **if** *All checks pass* **then**
  - Transfer C_u to $addr_{\mathcal{A}}$, set $flag = 1$
- **else**
  - Transfer C'_u to $addr_u$, set $flag = 1$

Function `RChalResolve`(ID_r, π_S)

- Check *message sender* $= addr_{\mathcal{A}}$
- Retrieve $Evid_r, flag, T_{recv}$ from Rec_R
- Check $flag = 0$ and $block.time - T_{recv} < T_H$
- Check $e(\frac{D_{\mathcal{R}}}{\prod_{j\in[n]} CK_{\mathcal{R}_{r,j}}^{K_{r,j}}}, \prod_{j\in[n]} \widetilde{CK}_{\mathcal{R}_{r,j}}) = e(\pi_S, \tilde{g})$
- **if** *All checks pass* **then**
  - Transfer C_r to $addr_{\mathcal{A}}$, set $flag = 1$
- **else**
  - Transfer C'_r to $addr_r$, set $flag = 1$

5.5 Security Analysis

First, we summarize the security properties of the *SNARG* systems in terms of *Completeness*, *Soundness*, *Succinctness* and *Privacy Preservation*. Then, we discuss the security and fairness of the accountability contract. Finally, based on the security properties of the *SNARG* systems and the smart contract, we give the security analysis of *privacy-preserving accountability*.

5.5.1 *SNARG* Security

The security properties of $\sum_{\mathcal{Q}}$ inherit from the QAP-based *SNARG* systems [122, 138]. *Completeness* is recognized by the *QAP* theorem and the correctness of the linear combination checks in the bilinear groups. Thus, a rational verifier will accept the proof if it is correctly generated. *Soundness* ensures the unforgeability of the proof $\pi_{\mathcal{Q}}$. If *q-PDH,d-PKE,2q-SDH* assumptions hold in an Elliptic curve-based groups of order q for a *QAP* of degree d where $q \geq 4d + 4$ [138], a computationally-bonded broker without the knowledge of the trapdoor secret cannot forge a valid tuple $(S_u, S_{\mathcal{R}}, S_o) \notin \mathcal{R}_{\mathcal{Q}}$ to pass the $Verify$ function. The proof is *succinct* (a few group elements) due to the high expressiveness of $EK_{\mathcal{Q}}$. Users choose random numbers to construct S_u in the registration phase, which makes the corresponding commitments indistinguishable even for two commitments with the same keyword set. Since only commitments of S_u is utilized in the $Verify$ function, $\sum_{\mathcal{Q}}$ achieves *privacy preservation* for the user profiles.

$\sum_{\mathcal{V}}$ is *succinct* since there are only 2 group elements in the proof $\pi_{\mathcal{V}}$. $\sum_{\mathcal{V}}$ is complete and sound if the *SXDH* assumption holds in $\mathbb{G}$ [141]. $\sum_{\mathcal{S}}$ is *succinct* since there is only one group element in the proof $\pi_{\mathcal{S}}$. $\sum_{\mathcal{S}}$ is *complete* and *sound* if *CDH* assumption holds for computationally-bounded adversaries [133]. $\sum_{\mathcal{S}}$ is not *privacy preserving* as the prover directly opens the S_r to the verifier. This design is reasonable since $\sum_{\mathcal{S}}$ is used for retailer challenge, where the targeting policy S_r is required to be transparent to the public.

5.5.2 Smart Contract Security

Smart contract security relies on the underlying Ehtereum blockchain. For a proof-of-work consensus protocol, the smart contract security is achieved if an adversary cannot control the most (51 percent) of the computing power in the blockchain network [81]. Informally, the smart contract in the Ethereum provides three security properties: (1) *Immutability*. Data stored in the contract cannot be maliciously modified. (2) *Persistence*. In a long

term, honest blockchain nodes will agree on a consistent view of the smart contract states. (3) *Liveness*. A correct contract function call will be verified and executed within a certain period of time, i.e. transaction confirmation time.

5.5.3 Privacy-preserving Accountability

Public Verifiability

The public verifiability [192] in *SANPA* consists of two parts: retailer policy management and the ad dissemination process. *SANPA* designs a trusted setup of the *SNARG* systems and the accountability contract, that stores evaluation keys of the *SNARG* systems and the cryptographic commitment of the $D_{\mathcal{R}}$ to receive challenges from users and retailers. Due to the *immutability* of the smart contract, the on-chain storage are secure and cannot be maliciously modified.

For the retailer challenge, the broker generates the evidence $Evid_r$ for each retailer policy S_r. $Evid_r$ is non-repudiable due to the security of *ECDSA* signature. Any retailer can require the broker to generate a proof by sending a retailer challenge to the accountability contract. The broker proves that S_r is correctly included in $D_{\mathcal{R}}$ by running $Prove$ function of $\sum_{\mathcal{S}}$, which cannot be forged due to the *soundness* of $\sum_{\mathcal{S}}$. For the user challenge, the broker generates the evidence $Evid_{sid}$ for each ad dissemination. The evidence is non-repudiable due to the security of *ECDSA* signature. When receiving a user challenge, the broker proves that the ad dissemination process is correctly conducted with $D_{\bar{uo}}$ in $Evid_{sid}$ and the on-chain authenticator $D_{\mathcal{R}}$. To do so, the broker runs *Prove* function of $\sum_{\mathcal{Q}}$ and $\sum_{\mathcal{V}}$. Due to the *soundness* of $\sum_{\mathcal{Q}}$ and $\sum_{\mathcal{V}}$, the broker cannot forge an invalid proof that passes $Verify$ functions in the accountability contract.

With the *completeness* of the *SNARG* systems, a rational user or retailer will accept the proof if it is correctly computed. Since the accountability contract is implemented over the Ethereum blockchian, anyone in the public can verify correctness of the transparency challenges. In summary, *public verifiability* is achieved in *SANPA*.

Privacy Preservation

For the user challenge, users upload the evidence $Evid_{sid}$ to the accountability contract. $Evid_{sid}$ contains the commitment $D_{\bar{uo}}$ of S_u, S_o. The accountability contract uses $Verify$ functions of $\sum_{\mathcal{Q}}$ and $\sum_{\mathcal{V}}$ for public verifications of the proof $\pi_{\mathcal{Q}}$ and $\pi_{\mathcal{V}}$. Due to *privacy preservation* property of $\sum_{\mathcal{Q}}$, the verifications on the contract will not leak the plaintext

of S_u and S_o, other than if $(S_u, S_{\mathcal{R}}, S_o) \in \mathcal{R}_{\mathcal{Q}}$. That is, *privacy preservation* for the user profile is achieved in *SANPA*.

Obligation Enforcement

SANPA utilizes the smart contract to enforce timely obligations against advertising misconducts. Specifically, the accountability contract takes initial deposits from the broker and will transfer the deposits if any advertising misconduct is identified. With the *liveness* of the smart contract, an advertising challenge cannot be maliciously delayed or ignored. Moreover, to promote timely responses from the broker, the accountability contract utilizes the Ethereum block time and sets a threshold T_H for each advertising challenge. That is, obligation enforcement is achieved in *SANPA*.

5.6 Performance Evaluation

Since *SANPA* is an add-on component of the existing *SAN*, we mainly evaluate the additional overhead of *SANPA* in terms of *SNARG* systems and the accountability contract. First, we present the theoretical and experimental analysis of *SANPA* systems $\sum_{\mathcal{Q}}$, $\sum_{\mathcal{V}_1}$, $\sum_{\mathcal{V}_2}$ and $\sum_{\mathcal{S}}$. Second, we analyze the accountability contract in terms of on-chain gas cost and the off-chain cost for the broker to response to the advertising challenges. All experiments are conducted on a Linux System with 2.4 GHz processor and 8 GB memory.

5.6.1 *SNARG* Systems

Since *SANPA* is an add-on component of the existing *SAN*, we mainly evaluate the additional overhead of *SANPA* in terms of the *SNARG* systems and the accountability contract. We utilize a single entity to implement DC in the experiments for illustrative purposes. First, we present the theoretical and experimental analysis of *SANPA* systems $\sum_{\mathcal{Q}}$, $\sum_{\mathcal{V}}$ and $\sum_{\mathcal{S}}$. Second, we analyze the accountability contract in terms of the on-chain gas cost. All experiments are conducted on a Linux System with 2.3GHz processor and 8 GB memory.

$\sum_{\mathcal{V}}$ and $\sum_{\mathcal{S}}$ Complexity

$\sum_{\mathcal{V}}$ and $\sum_{\mathcal{S}}$ are instantiated based on *alt-bn128* curve in the *libff* library [178]. In Table 5.2 and 5.3, we summarize the storage and computation complexity of $\sum_{\mathcal{V}}$ and $\sum_{\mathcal{S}}$ with

the following observations:

- ▷ Storage cost of proof and computation cost of $Verify$ remain *succinct* regardless of the input size m. This is critical since the verification is conducted on the accountability contract with expensive on-chain computation and storage costs.
- ▷ $Prove$ and $KeyGen$ consists of multi exponentiations E_1 and E_2, which is optimized in the *libff* with a table of intermediate powers.

Therefore, $\sum_{\mathcal{V}}$ and $\sum_{\mathcal{S}}$ are practical for both the on-chain implementations of the $Verify$ function and off-chain implementations of $Prove$ and $Keygen$ at the broker with powerful computing resources.

Table 5.2: *SNARG* Complexity - I

	EK	VK	Proof
$\sum_{\mathcal{V}}$	$2mn\|\mathbb{G}_1\|$	$3\|\mathbb{G}_2\|$	$2\|\mathbb{G}_1\|$
$\sum_{\mathcal{S}}$	$(m^2n^2 - mn)\mathbb{G}_1$	$n(\mathbb{G}_1 + \mathbb{G}_2)$	$\mathbb{G}_1$

* $|\mathbb{G}_1|, |\mathbb{G}_2|, |\mathbb{Z}_P|$, size of a group element in $\mathbb{G}_1, \mathbb{G}_2, \mathbb{Z}_P$; E_1, E_2, exponentiation operations in $\mathbb{G}_1, \mathbb{G}_2$; P, paring operation in $\mathbb{G}$; m, number of retailers; n, dimension of the keyword dictionary.

Table 5.3: *SNARG* Complexity - II

	KeyGen	Prove	Verify
$\sum_{\mathcal{V}}$	$4mnE_1 + 3E_2$	$2mnE_1$	$4P$
$\sum_{\mathcal{S}}$	$(m^2n^2 - mn)E_1 + mnE_2$	$n(mn - n)E_1$	$nE_1 + 2P$

* $|\mathbb{G}_1|, |\mathbb{G}_2|, |\mathbb{Z}_P|$, size of a group element in $\mathbb{G}_1, \mathbb{G}_2, \mathbb{Z}_P$; E_1, E_2, exponentiation operations in $\mathbb{G}_1, \mathbb{G}_2$; P, paring operation in $\mathbb{G}$; m, number of retailers; n, dimension of the keyword dictionary.

$\sum_{\mathcal{Q}}$ Complexity

We write the ad dissemination process of Algorithm 6 in C. The python interface of Pinocchio is adopted to translate the C codes to circuits. We re-compile the circuit-to-SNARG interface [140, 193, 139] in *libsnark*, by instantiating the *R1CS ppzkSNARK* with *alt-bn128* curve, which is supported by pairing operations in the Ethereum [81]. We choose three tunable system parameters: the dimension of the keyword dictionary n, the number of

retailers m, the number of returned results k. We also identify three sets of performance indicators. First, we measure the time costs for the three functions: $KeyGen$, $Prove$ and $Verify$. Second, we measure the QAP complexity by the number of variables and degrees in the compiled QAP program. Third, we measure the size of PK and VK in bits.

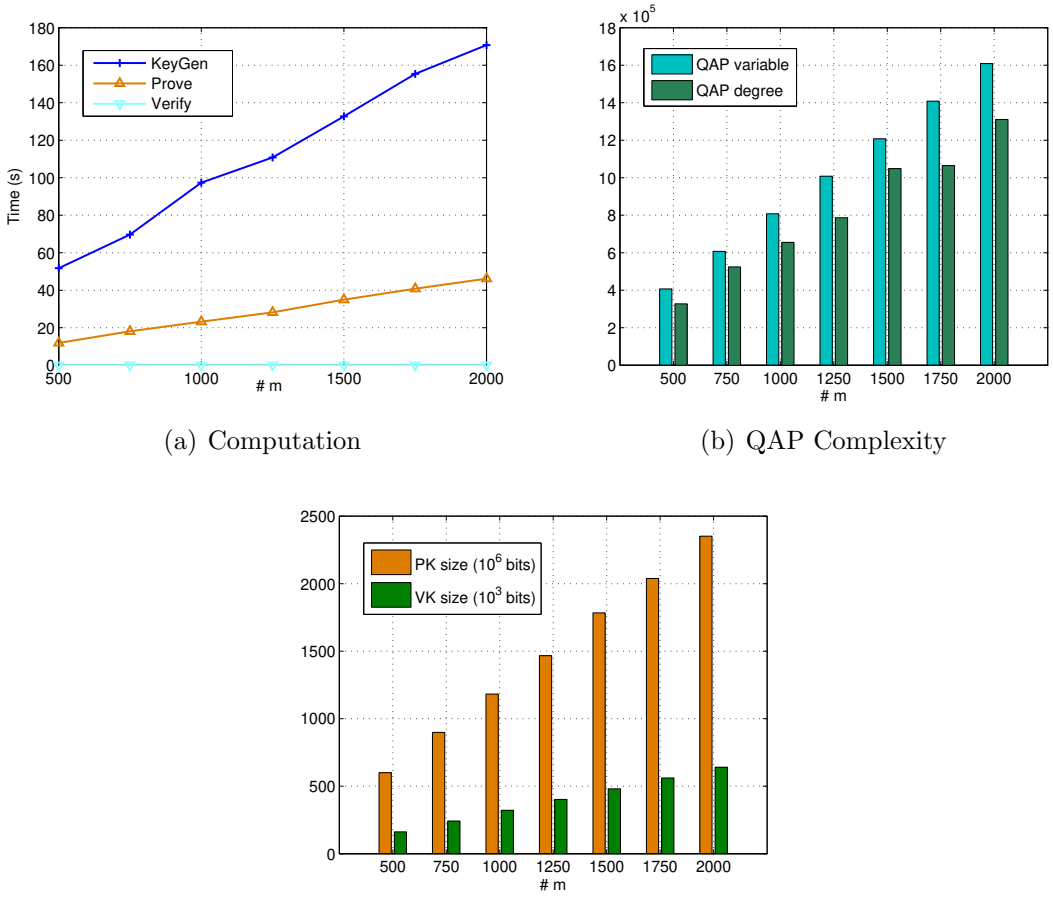
In Fig. 5.3(a), 5.4(a) and 5.5(a), the time cost for $Prove$ and $KeyGen$ increases with m, n, k. At the same time, the time cost for $Verify$ remains the same: around 0.27s in all experiments. This is because $\sum_{\mathcal{Q}}$ achieves *succinct* verification cost with only 12 pairings. Although $Keygen$ operation is much more expensive compared with $Prove$ and $Verify$, it is acceptable since it is an one-time setup and the entity can remain offline after the setup. Similar properties for the QAP complexity are found in Fig. 5.3(b), 5.4(b) and 5.5(b). In Fig. 5.3(c), 5.4(c) and 5.5(c), the PK size is much larger than the VK size. Since the original S_o instead of its commitment is used in the experiment, the VK size linearly grows with the number of outputs S_o. The VK size remains the same in Fig. 5.4(c) as the number of outputs is fixed at 1000. In Fig. 5.5(c), the VK size is reduced to a few thousand bits if $\sum_{\mathcal{Q}}$ only outputs a few results.

5.6.2 Accountability Contract

We mainly estimate the gas cost of storing data and conducting cryptographic operations single algebraic operations and read/store operations are negligible [194] in the contract. For example, it costs 2,000 gas to store a 256-bit word in the contract storage and 3,000 gas to verify an $ECDSA$ signature. For *alt-bn128* group operations, we take the estimations of latest optimized precompiled contracts [195].

In Table 5.4, we summarize the storage, computation, and gas cost for the four complex functions of the accountability contract. For the estimation, we take the theoretical results of *alt-bn128* curve, where $|\mathbb{Z}_p|$ is 256 bit, $|\mathbb{G}_1|$ is 512 bit, and $|\mathbb{G}_2|$ is 1024 bit. We regard the user/retailer ID and the time stamp as a 32-bit word. The combination of $\sum_{\mathcal{Q}}$ and $\sum_{\mathcal{V}}$ increases the size and verification cost of $\pi_{\mathcal{Q}}$ compared with our *libsnark* implementations. Theoretically, two additional elements $D_{\bar{\mathcal{R}}}, D_{u\bar{o}} \in \mathbb{G}_1$ are introduced. Two more elements in $\mathbb{G}_1$ for appropriate span checks of $D_{\bar{\mathcal{R}}}, D_{u\bar{o}}$ are also added. Two more elements in $\mathbb{G}_1$ and two more pairings are also introduced to check $D_{\bar{\mathcal{R}}}, D_{u\bar{o}}$ are well formed. In Table 5.4, the retailer challenge is more expensive than the user challenge. As a result, we can increase the amount of C_r', such that the broker loses more cryptocurrencies if the misconduct of retailer policy management is identified.

We summarize some insights into the experimental results. (1) Linear off-chain complexity at the prover is mainly caused by the arithmetic computations. Specifically, dy-

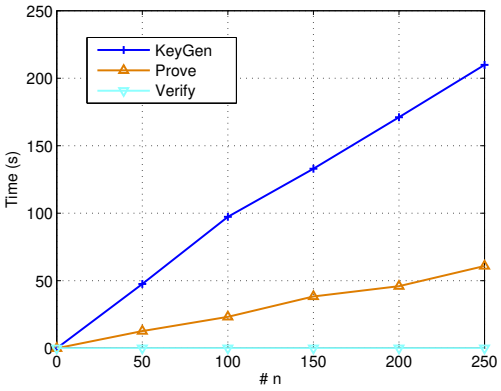


(a) Computation

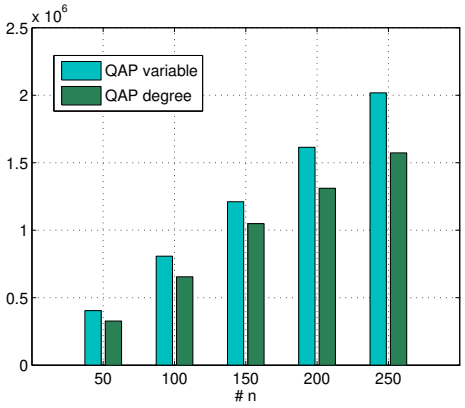
(b) QAP Complexity

(c) Key Size

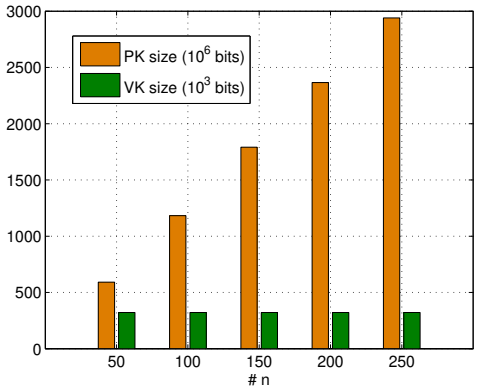
Figure 5.3: Overhead vs m, $n = 100, k = m$



(a) Computation

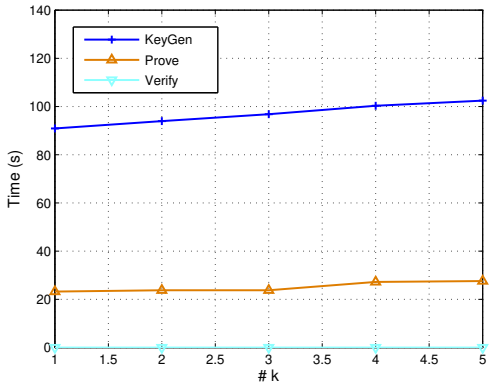


(b) QAP Complexity

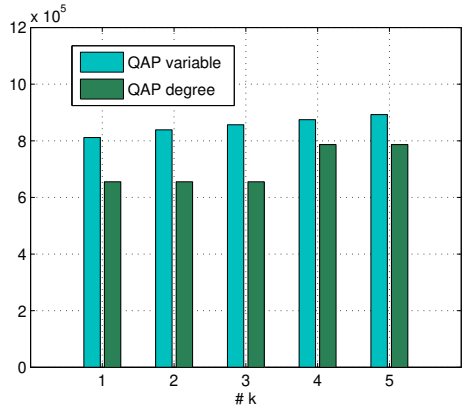


(c) Key Size

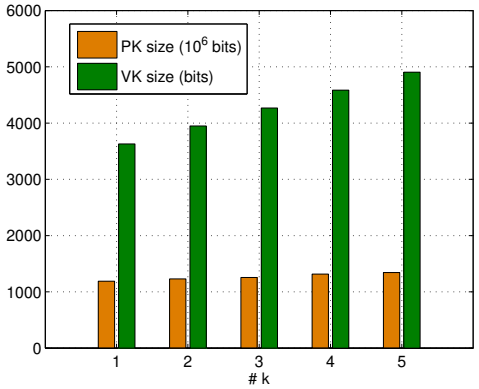
Figure 5.4: Overhead vs n, $m = k = 1000$



(a) Computation



(b) QAP Complexity



(c) Key Size

Figure 5.5: Overhead vs k, $m = 1000, n = 100$

namic subscript assignment for arrays and loop breaks are not supported in the implementation. (2) The efficiency at the on-chain verifier is achieved due to the efficient verifications of the QAP theorem, which significantly reduces the implementation cost as on-chain storage and computation are expensive.

Table 5.4: Function Complexity & Gas Cost

Function	Storage	Computation	Approximate Gas Cost
UserChallenge	$2\pi_E + 4 * W_{32} + \lvert\mathbb{G}_1\rvert$	$2ER$	20,000
RetailerChallenge	$\pi_E + 2 * W_{32} + n(\lvert\mathbb{G}_1\rvert + \lvert\mathbb{G}_2\rvert + \lvert\mathbb{Z}_P\rvert)$	ER	14,000 n +9,000
UChalResolve	$2 * W_{32} + 13\lvert\mathbb{G}_1\rvert + \lvert\mathbb{G}_2\rvert$	18P	719,000
RChalResolve	$W_{32} + \lvert\mathbb{G}_1\rvert$	$2P + nE_1$	6,000n +119,000

* $\lvert\mathbb{G}_1\rvert, \lvert\mathbb{G}_2\rvert, \lvert\mathbb{Z}_P\rvert$, size of a group element in $\mathbb{G}_1, \mathbb{G}_2, \mathbb{Z}_P$; E_1, exponentiation operation in $\mathbb{G}_1$; P, paring operation in $\mathbb{G}$; π_E, $ECDSA$ signature; ER, $ECDSA$ verification operation; W_i, an i-bit word; n, dimension of the keyword dictionary.

5.7 Summary

In this chapter, we have proposed a blockchain-based Smart Advertising Network with Privacy-preserving Accountability (*SANPA*). *SANPA* can increase the public awareness of the advertising transparency with effective enforcements on the advertising misconduct. We have conducted extensive experiments to demonstrate that *SANPA* is compatible with the existing SAN and is feasible for real-world implementations. This research work should shed light on the future research and practice of a more trustworthy advertising network.

Chapter 6

Conclusions and Future Works

In this chapter, the conclusions of this thesis are drawn and the future works are summarized.

6.1 Conclusions

The thesis addresses the security and privacy threats in the mobile advertising by designing, implementing, and evaluating a blockchain-based architecture. The thesis takes advantage of the blockchain technology to enhance the transparency and accountability of the mobile advertising. At the same time, the thesis presents a set of design strategies to significantly reduce the on-chain storage and computation overhead while preserving user profile and review privacy. Specifically, the conclusions of this thesis are summarized as follows.

- ▷ *Anonymous Reputation System.* The thesis first studies the reliable and transparent review system for the marketplace in the mobile advertising system. To enable mobile users to leave reviews without the fear of privacy leakage, the research explores an anonymous review token generation and private review aggregation technique based on group signature, zero-knowledge proof and homomorphic encryption. The research also builds a blockchain architecture based on Proof-of-Stake consensus protocol with a public rating board, where the rating accumulation of each merchant is transparent and verifiable to the public. Moreover, the research divides the review accumulation into different phases: token generation, review generation and review aggregation. Individual review is encrypted and aggregated on the blockchain, and only aggregated

reviews are revealed to the public. Double-review attacks are prevented while preserving the anonymity and confidentiality of the reviewer. The research also explores the implementation challenges with a blockchain testing network. The insights and experiences obtained from the first research pave the way for the following research works.

- ▷ *Transparent and Accountable Vehicular Local Advertising.* This research investigates a use case of the mobile advertising, i.e., vehicular local advertising. Specifically, mobile users traveling in a smart vehicle can receive advertisements based on their locations and interests. The research explores the characteristics of vehicular local advertising models as a spatial keyword query function. The research identifies the lack of transparency issues, where inaccurate or fake advertisements could be sent out to vehicular users. The research takes into considerations the two dimensions (spatial and textual) of the vehicular advertising based on the R-tree technique and user interest representation techniques, to design a verifiable spatial keyword query scheme based on the smart contract. Considering the efficiency issue of the blockchain, the research proposes two design strategies: *digest-and-verify* and *divide-then-assemble* to reduce the on-chain overheads. The implementation and observations of the vehicular advertising system also provide comprehensive benchmarks for future research on the mobile advertising system.

- ▷ *Smart Advertising Network with Privacy-preserving Accountability.* The research further investigates the privacy issues, especially the user profile privacy, in the mobile advertising system. The research identifies the public accountability enforcement on the advertising misbehavior is important to tackle the security threats in the mobile advertising system. Specifically, the research designs an accountability smart contract for the smart advertising network. Both users and merchants can enjoy the efficient ad dissemination services with the design of an on/off-chain advertising model. At the same time, mobile users and merchants can require transparency explanations of specific advertising activities by sending challenges to the accountability contract. With the design of a composite *SNARG* system for the pre-determined advertising policy, the accountability contract can verify the correctness of the challenged advertising activities without exposing user profile privacy. Since the accountability contract is only executed when transparency explanations are required, the research achieves a notably effective balance between user privacy and advertising efficiency at the same time. The research provides comprehensive experimental results for the off-chain overheads and theoretical analysis of the accountability contract based on the latest Ethereal status.

6.2 Future Works

6.2.1 Ad-fraud Attack Mitigation with Public Accountability

A major concern for the continuous success of the mobile advertising model is the *ad-fraud* attack [2]. It is highly motivated for the application developers to gain more revenues with fraud displaying of ad contents. For example, the developers can hide the display of ads (ad hidden fraud) or trigger multiple ad impression claims for one single ad clicks (impression number fraud) [55]. The *ad-fraud* attacks are reported to cause millions US dollars every year. Therefore, the *ad-fraud* attacks reduce the retailer confidence with a lower ad conversion rate, and more and more mobile users are choosing ad-free applications or install ad-block software [55]. To build an ecosystem that resolves the *ad-fraud* issue, extensive research efforts have been directed to build in-app countermeasures, that monitor the application runtime environment based on UI-state transition graph [57], machine learning [32] and trusted hardware [46]. Meanwhile, the ad broker has also proposed system-level methods that track the abnormal impression traffic or gain reviews from the mobile users in the app store.

Although the existing countermeasures have great impacts on resolving the *ad-fraud* attacks, there still remain some issues that are not fully addressed. First, the lack of system transparency for the ad broker still exists [169]. The success of anti-fraud methods relies highly on the trustworthiness of the ad broker. The huge gap between retailers and mobile users put forward the requirements of increasing the advertising transparency when dealing with the *ad-fraud* attacks. Second, the prompt response and actions of the broker over the fraudulent applications are not well achieved due to benefit and regulation considerations. According to statistics [40], 21.3/26.9 percent of mobile apps are considered fraudulent in Apple Store or Google Play. As a result, a framework that increases the advertising system transparency and engages prompt actions is an urgent task. In the future, further research will be directed to build a blockchain-based accountability architecture for ad-fraud attack mitigation [196], where joint accountability will be pursued against misbehaving developers and ineffective ad brokers.

6.2.2 Blockchain-based Mobile Advertising with GDPR Compliance

In the current mobile advertising system, ad brokers rely on user tracking techniques to collect user activities and build user profiles for the targeted ad dissemination [10]. Euro-

pean General Data Protection Regulation (GDPR) [41] has taken effect since 2018, which has a great impact on the mobile advertising industry [31]. GDPR grants mobile users with the a set of rights on their personal data. Mobile users have the "right to be informed" about how their data are processed at ad brokers and "right to be forgotten" if they would like to turn off the ad personalization services. More importantly, GDPR enforces strict controls over sharing of the personal data with third parties, which significantly restricts the mobile application's ability to share user data with ad brokers. As a result, it becomes an urgent need to design an infrastructure that can transparently and efficiently manage the data flows in the mobile advertising system.

Blockchain with distributed consensus is a promising intermediate for data management in the mobile advertising system that enforces GDPR compliance. However, there are still some issues remaining unsolved. First, it may need a redesign of the fundamental blockchain infrastructure taking into consideration the roles and motivations of the stakeholders in the mobile advertising system. Therefore, a hybrid blockchain architecture is envisioned. The architecture will integrate the features of both public blockchain [81] and consortium blockchain [89] to define the views of the public and the involved stakeholders in the advertising system. Second, the fine-grained access control policy for the data sharing over a blockchain architecture requires complex designs of on-chain data storage and processing model, which not only requires developments of cryptographic primitives, but also collaborative inter-discipline innovations. In the future, further research efforts will be directed to design a versatile and efficient blockchain-based data management architecture for the mobile advertising system.

6.3 Final Remarks

In this thesis, we have designed, implemented, and evaluated a blockchain-based architecture that resolves the security and privacy issues in the mobile advertising system. The observations in designing the architecture and the experimental results may shed light on future research on the blockchain-based transparent architecture for the mobile advertising. Future research directions have been discussed to facilitate collaborations and implementations from both the academy and the industry.

References

[1] M. M. Tsang, S.-C. Ho, and T.-P. Liang, "Consumer attitudes toward mobile advertising: An empirical study," *International journal of electronic commerce*, vol. 8, no. 3, pp. 65–78, 2004.

[2] D. Grewal, Y. Bart, M. Spann, and P. P. Zubcsek, "Mobile advertising: a framework and research agenda," *Journal of Interactive Marketing*, vol. 34, pp. 3–14, 2016.

[3] S. Dhar and U. Varshney, "Challenges and business models for mobile location-based services and advertising," *Communications of the ACM*, vol. 54, no. 5, pp. 121–128, 2011.

[4] K. Ullah, L. Jaimes, R. S. Yokoyama, and E. dos Santos Moreira, "Advertising roadside services using vehicular ad hoc network (vanet) opportunistic capabilities," in *Proc. of International Conference on Advances in Vehicular Systems, Technologies and Applications*, 2015, pp. 7–13.

[5] H. Aksu, L. Babun, M. Conti, G. Tolomei, and A. S. Uluagac, "Advertising in the iot era: Vision and challenges," *arXiv preprint arXiv:1802.04102*, 2018.

[6] 75+ MOBILE MARKETING STATISTICS FOR 2020 AND BEYOND. https://www.bluecorona.com/blog/mobile-marketing-statistics/. Accessed June 2020.

[7] D. Zhang, L. Guo, L. Nie, J. Shao, S. Wu, and H. T. Shen, "Targeted advertising in public transportation systems with quantitative evaluation," *ACM Transactions on Information Systems*, vol. 35, no. 3, pp. 1–29, 2017.

[8] P.-T. Chen and H.-P. Hsieh, "Personalized mobile advertising: Its key attributes, trends, and social impact," *Technological Forecasting and Social Change*, vol. 79, no. 3, pp. 543–557, 2012.

[9] V. Toubiana, A. Narayanan, D. Boneh, H. Nissenbaum, and S. Barocas, "Adnostic: Privacy preserving targeted advertising," 2010.

[10] J. Jiang, Y. Zheng, Z. Shi, X. Yuan, X. Gui, and C. Wang, "A practical system for privacy-aware targeted mobile advertising services," *IEEE Transactions on Services Computing*, 2017, DOI: 10.1109/TSC.2017.2697385.

[11] E. McCallister, *Guide to protecting the confidentiality of personally identifiable information*. Diane Publishing, 2010, vol. 800, no. 122.

[12] L. Aalto, N. Göthlin, J. Korhonen, and T. Ojala, "Bluetooth and wap push based location-aware mobile advertising system," in *Proc. of International Conference on Mobile Systems, Applications, and Services*, 2004, pp. 49–58.

[13] S. S. Banerjee and R. R. Dholakia, "Mobile advertising: Does location based advertising work?" *International Journal of Mobile Marketing*, 2008.

[14] C.-H. Wong, G. W.-H. Tan, B.-I. Tan, and K.-B. Ooi, "Mobile advertising: the changing landscape of the advertising industry," *Telematics and Informatics*, vol. 32, no. 4, pp. 720–734, 2015.

[15] B. Stone-Gross, R. Stevens, A. Zarras, R. Kemmerer, C. Kruegel, and G. Vigna, "Understanding fraudulent activities in online ad exchanges," in *Proc. of ACM SIGCOMM Conference on Internet Measurement Conference*, 2011, pp. 279–294.

[16] H. Haddadi, P. Hui, and I. Brown, "Mobiad: private and scalable mobile advertising," in *Proc. of ACM International Workshop on Mobility in the Evolving Internet Architecture*, 2010, pp. 33–38.

[17] Google Ads. https://ads.google.com/. Accessed June 2020.

[18] Y. Jin, K. Seipp, E. Duval, and K. Verbert, "Go with the flow: effects of transparency and user control on targeted advertising using flow charts," in *Proc. of the International Working Conference on Advanced Visual Interfaces*, 2016, pp. 68–75.

[19] Y. Pang, B. Wang, F. Wu, G. Chen, and B. Sheng, "Prota: A privacy-preserving protocol for real-time targeted advertising," in *Proc. of IEEE IPCCC*, 2015, pp. 1–8.

[20] M. Green, W. Ladd, and I. Miers, "A protocol for privately reporting ad impressions at scale," in *Proc. of ACM CCS*, 2016, pp. 1591–1601.

[21] H. Corrigan-Gibbs and D. Boneh, "Prio: Private, robust, and scalable computation of aggregate statistics." in *Proc. NSDI*, 2017, pp. 259–282.

[22] J. Ni, X. Lin, and X. S. Shen, "Efficient and secure service-oriented authentication supporting network slicing for 5g-enabled iot," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 3, pp. 644–657, 2018.

[23] G. Araniti, M. Condoluci, P. Scopelliti, A. Molinaro, and A. Iera, "Multicasting over emerging 5g networks: challenges and perspectives," *IEEE Network*, vol. 31, no. 2, pp. 80–89, 2017.

[24] W. Ji, Y. Chen, M. Chen, B.-W. Chen, Y. Chen, and S.-Y. Kung, "Profit maximization through online advertising scheduling for a wireless video broadcast network," *IEEE Transactions on Mobile Computing*, vol. 15, no. 8, pp. 2064–2079, 2015.

[25] A. Asadi and V. Mancuso, "Network-assisted outband d2d-clustering in 5g cellular networks: Theory and practice," *IEEE Transactions on Mobile Computing*, vol. 16, no. 8, pp. 2246–2259, 2016.

[26] A. Zhang, J. Chen, R. Q. Hu, and Y. Qian, "Seds: Secure data sharing strategy for d2d communication in lte-advanced networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2659–2672, 2015.

[27] F. Lyu, H. Zhu, N. Cheng, H. Zhou, W. Xu, M. Li, and X. Shen, "Characterizing urban vehicle-to-vehicle communications for reliable safety applications," *IEEE Trans. Intell. Transp. Syst. DOI:10.1109/TITS.2019.2920813*, to appear.

[28] D. Liu, J. Ni, H. Li, and X. S. Shen, "Achieving adaptive linkability for cellular v2x group communications in 5g," in *Proc. of IEEE GLOBECOM*, 2018.

[29] I. De Felipe, V. Hristidis, and N. Rishe, "Keyword search on spatial databases," in *IEEE International Conference on Data Engineering*, 2008, pp. 656–665.

[30] M. Li, J. Weng, A. Yang, J.-N. Liu, and X. Lin, "Toward blockchain-based fair and anonymous ad dissemination in vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 11 248–11 259, 2019.

[31] T. Urban, D. Tatang, M. Degeling, T. Holz, and N. Pohlmann, "Measuring the impact of the gdpr on data sharing in ad networks," in *Prof. of ASIA CCS*, 2020.

[32] J. Crussell, R. Stevens, and H. Chen, "Madfraud: Investigating ad fraud in android applications," in *Proc. ACM MobiSys*, 2014, pp. 123–134.

[33] A. Andreou, M. Silva, F. Benevenuto, O. Goga, P. Loiseau, and A. Mislove, "Measuring the facebook advertising ecosystem," in *Proc. of NDSS*, 2019.

[34] G. Chen, W. Meng, and J. Copeland, "Revisiting mobile advertising threats with madlife," in *Proc. of WWW*, pp. 207–217.

[35] V. Rastogi, R. Shao, Y. Chen, X. Pan, S. Zou, and R. Riley, "Are these ads safe: Detecting hidden attacks through the mobile app-web interfaces," in *Proc. of NDSS*, 2016.

[36] K. Soska, A. Kwon, N. Christin, and S. Devadas, "Beaver: A decentralized anonymous marketplace with secure reputation," *IACR Cryptology ePrint Archive*, p. 464, 2016.

[37] J. Blömer, F. Eidens, and J. Juhnke, "Practical, anonymous, and publicly linkable universally-composable reputation systems," in *Proc. of CTRSA*. Springer, 2018, pp. 470–490.

[38] M. Hardt and S. Nath, "Privacy-aware personalization for mobile advertising," in *Proc. of ACM CCS*, 2012, pp. 662–673.

[39] S. Son, D. Kim, and V. Shmatikov, "What mobile ads know about mobile users," in *Proc. of NDSS*, 2016.

[40] Mobile Ad Fraund. https://www.businessofapps.com/ads/ad-fraud/research/ad-fraud-statistics/. Accessed May 2020.

[41] General Data Protection Regulation (GDPR). https://gdpr-info.eu. Accessed April 2020.

[42] J. Parra-Arnau, J. P. Achara, and C. Castelluccia, "Myadchoices: Bringing transparency and control to online advertising," *ACM Transactions on the Web (TWEB)*, vol. 11, no. 1, pp. 1–47, 2017.

[43] A. Andreou, G. Venkatadri, O. Goga, K. Gummadi, P. Loiseau, and A. Mislove, "Investigating ad transparency mechanisms in social media: A case study of facebook's explanations," in *Proc. of NDSS*, 2018.

[44] Ad blocking rate. https://www.statista.com/topics/3201/ad-blocking/. Accessed June 2020.

[45] G. Venkatadri, A. Mislove, and K. P. Gummadi, "Treads: Transparency-enhancing ads," in *Proc. of the ACM Workshop on Hot Topics in Networks*, 2018, pp. 169–175.

[46] W. Li, H. Li, H. Chen, and Y. Xia, "Adattester: Secure online mobile advertisement attestation using trustzone," in *Prof. of ACM MobiSys*, 2015, pp. 75–88.

[47] Google, Microsoft and Amazon pay to get around ad blocking tool. https://www.ft.com/content/80a8ce54-a61d-11e4-9bd3-00144feab7de. Accessed March 2020.

[48] A. Boukis, "Exploring the implications of blockchain technology for brand–consumer relationships: a future research agenda," *Journal of Product and Brand Management*, 2019.

[49] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Manubot, Tech. Rep., 2019.

[50] Y. Ding, D. Luo, H. Xiang, C. Tang, L. Liu, X. Zou, S. Li, and Y. Wang, "A blockchain-based digital advertising media promotion system," in *International Conference on Security and Privacy in New Computing Environments*. Springer, 2019, pp. 472–484.

[51] W. Chen, Z. Xu, S. Shi, Y. Zhao, and J. Zhao, "A survey of blockchain applications in different domains," in *Pro. of International Conference on Blockchain Technology and Application*, 2018, pp. 17–21.

[52] The AdChain Registry. https://www.altoros.com/blog/adchain-registry-blockchain-to-prevent-fraud-in-digital-advertising/. Accessed May 2020.

[53] M. Pärssinen, M. Kotila, R. C. Rumin, A. Phansalkar, and J. Manner, "Is blockchain ready to revolutionize online advertising?" *IEEE Access*, vol. 6, pp. 54 884–54 899, 2018.

[54] A. Zhang and X. Lin, "Towards secure and privacy-preserving data sharing in e-health systems via consortium blockchain," *Journal of Medical Systems*, vol. 42, no. 8, p. 140, 2018.

[55] J. Gui, M. Nagappan, and W. G. Halfond, "What aspects of mobile ads do users care about? an empirical study of mobile in-app ad reviews," *arXiv preprint arXiv:1702.07681*, 2017.

[56] R. Shao, V. Rastogi, Y. Chen, X. Pan, G. Guo, S. Zou, and R. Riley, "Understanding in-app ads and detecting hidden attacks through the mobile app-web interface," *IEEE Transactions on Mobile Computing*, vol. 17, no. 11, pp. 2675–2688, 2018.

[57] F. Dong, H. Wang, L. Li, Y. Guo, T. F. Bissyandé, T. Liu, G. Xu, and J. Klein, "Frauddroid: Automated ad fraud detection for android apps," in *Proc. of ESEC/FSE*, 2018, pp. 257–268.

[58] L. Jin, B. He, G. Weng, H. Xu, Y. Chen, and G. Guo, "Madlens: Investigating into android in-app ad practice at api granularity," *IEEE Transactions on Mobile Computing*, 2019, to appear.

[59] S. Guha, B. Cheng, and P. Francis, "Privad: Practical privacy in online advertising," in *Proc. of NSDI*, 2011, pp. 169–182.

[60] M. Backes, A. Kate, M. Maffei, and K. Pecina, "Obliviad: Provably secure and practical online behavioral advertising," in *Proc. of IEEE S&P*, 2012, pp. 257–271.

[61] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," in *Proc. of IEEE 36th Annual Foundations of Computer Science*, 1995, pp. 41–50.

[62] C. Dwork, "Differential privacy: A survey of results," in *International conference on theory and applications of models of computation*. Springer, 2008, pp. 1–19.

[63] D. Davidson, M. Fredrikson, and B. Livshits, "Morepriv: Mobile os support for application personalization and privacy," in *Proc. of the 30th Annual Computer Security Applications Conference*, 2014, pp. 236–245.

[64] R. Ostrovsky and W. E. Skeith, "Private searching on streaming data," in *Proc. of CRYPTO*. Springer, 2005, pp. 223–240.

[65] J. Groth and M. Kohlweiss, "One-out-of-many proofs: Or how to leak a secret and spend a coin," in *Proc. of EUROCRYPT*. Springer, 2015, pp. 253–280.

[66] J. Qian, F. Qiu, F. Wu, N. Ruan, G. Chen, and S. Tang, "Privacy-preserving selective aggregation of online user behavior data," *IEEE Transactions on Computers*, vol. 66, no. 2, pp. 326–338, 2016.

[67] R. Dennis and G. H. Owenson, "Rep on the block: A next generation reputation system based on the blockchain," in *Proc. of International Conference for Internet Technology and Secured Transactions (ICITST)*, 2016.

[68] R. Di Pietro, X. Salleras, M. Signorini, and E. Waisbard, "A blockchain-based trust system for the internet of things," in *23nd ACM on Symposium on Access Control Models and Technologies*, 2018, pp. 77–83.

[69] E. Zhai, D. I. Wolinsky, R. Chen, E. Syta, C. Teng, and B. Ford, "Anonrep: Towards tracking-resistant anonymous reputation," in *Proc. of NSDI*, 2016, pp. 583–596.

[70] J. Blömer, J. Juhnke, and C. Kolb, "Anonymous and publicly linkable reputation systems," in *International Conference on Financial Cryptography and Data Security*. Springer, 2015, pp. 478–488.

[71] M. A. Azad, S. Bag, and F. Hao, "Privbox: Verifiable decentralized reputation system for the on-line marketplaces," *Future Generation Computer Systems*, 2018.

[72] R. Bazin, A. Schaub, O. Hasan, and L. Brunie, "A decentralized anonymity-preserving reputation system with constant-time score retrieval," *IACR Cryptology ePrint Archive*, p. 416, 2016.

[73] S. Bag, M. A. Azad, and F. Hao, "A privacy-aware decentralized and personalized reputation system," *Computers & Security*, vol. 77, pp. 514–530, 2018.

[74] J. Camenisch and M. Stadler, "Efficient group signature schemes for large groups," in *Proc. of CRYPTO*. Springer, 1997, pp. 410–424.

[75] P. Rogaway and T. Shrimpton, "Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance," in *Proc. of International Workshop on Fast Software Encryption*. Springer, 2004, pp. 371–388.

[76] A. Gervais, G. O. Karame, K. Wüst, V. Glykantzis, H. Ritzdorf, and S. Capkun, "On the security and performance of proof of work blockchains," in *Proc. of ACM CCS*, 2016, pp. 3–16.

[77] A. Kiayias, A. Russell, B. David, and R. Oliynykov, "Ouroboros: A provably secure proof-of-stake blockchain protocol," in *Proc. of CRYPTO*. Springer, 2017, pp. 357–388.

[78] R. Grinberg, "Bitcoin: An innovative alternative digital currency," *Hastings Sci. & Tech. LJ*, vol. 4, p. 159, 2012.

[79] D. Johnson, A. Menezes, and S. Vanstone, "The elliptic curve digital signature algorithm (ecdsa)," *International journal of information security*, vol. 1, no. 1, pp. 36–63, 2001.

[80] R. C. Merkle, "A digital signature based on a conventional encryption function," in *Proc. of CRYPTO*. Springer, 1987, pp. 369–378.

[81] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger byzantium version," *Ethereum project yellow paper*, pp. 1–39, 2018-06-05.

[82] V. Buterin *et al.*, "A next-generation smart contract and decentralized application platform," *white paper*, vol. 3, no. 37, 2014.

[83] C. Dannen, *Introducing Ethereum and Solidity*. Springer, 2017, vol. 1.

[84] J. Mendling, I. Weber, W. V. D. Aalst, J. V. Brocke, C. Cabanillas, F. Daniel, S. Debois, C. D. Ciccio, M. Dumas, S. Dustdar *et al.*, "Blockchains for business process management-challenges and opportunities," *ACM Transactions on Management Information Systems (TMIS)*, vol. 9, no. 1, p. 4, 2018.

[85] K. Korpela, J. Hallikas, and T. Dahlberg, "Digital supply chain transformation toward blockchain integration," in *50th Hawaii International Conference on System Sciences*, 2017.

[86] S. Dziembowski, L. Eckey, and S. Faust, "Fairswap: How to fairly exchange digital goods," in *Proc. of ACM CCS*, 2018, pp. 967–984.

[87] R. Böhme, N. Christin, B. Edelman, and T. Moore, "Bitcoin: Economics, technology, and governance," *Journal of economic Perspectives*, vol. 29, no. 2, pp. 213–38, 2015.

[88] Bitcoin (BTC) price stats and information. https://bitinfocharts.com/bitcoin/. Accessed May 2020.

[89] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. De Caro, D. Enyeart, C. Ferris, G. Laventman, Y. Manevich *et al.*, "Hyperledger fabric: a distributed operating system for permissioned blockchains," in *Proc. of EuroSys Conference*, 2018, pp. 1–15.

[90] S. Underwood, "Blockchain beyond bitcoin," 2016.

[91] S.-F. Sun, M. H. Au, J. K. Liu, and T. H. Yuen, "Ringct 2.0: a compact accumulator-based (linkable ring signature) protocol for blockchain cryptocurrency monero," in *European Symposium on Research in Computer Security*. Springer, 2017, pp. 456–474.

[92] M. Zamani, M. Movahedi, and M. Raykova, "Rapidchain: Scaling blockchain via full sharding," in *Proc. of ACM CCS*, 2018, pp. 931–948.

[93] J. Eberhardt and S. Tai, "Zokrates-scalable privacy-preserving off-chain computations," in *IEEE International Conference on Blockchain*, 2018.

[94] S. Bowe, A. Chiesa, M. Green, I. Miers, P. Mishra, and H. Wu, "Zexe: Enabling decentralized private computation," *IACR Cryptology ePrint Archive*, 2018.

[95] M. Green and I. Miers, "Bolt: Anonymous payment channels for decentralized currencies," in *Proc. of ACM CCS*, 2017, pp. 473–489.

[96] E. B. Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza, "Zerocash: Decentralized anonymous payments from bitcoin," in *Proc. IEEE S&P*, 2014, pp. 459–474.

[97] A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papamanthou, "Hawk: The blockchain model of cryptography and privacy-preserving smart contracts," in *Proc. of IEEE S&P*, 2016, pp. 839–858.

[98] R. Neisse, G. Steri, and I. Nai-Fovino, "A blockchain-based approach for data accountability and provenance tracking," in *Proc. of ACM International Conference on Availability, Reliability and Security*, 2017, p. 14.

[99] Z. Wu, A. B. Williams, and D. Perouli, "Dependable public ledger for policy compliance, a blockchain based approach," in *Proc of IEEE ICDCS*, 2019, pp. 1891–1900.

[100] J. Frankle, S. Park, D. Shaar, S. Goldwasser, and D. Weitzner, "Practical accountability of secret processes," in *Proc. of USENIX Security*, 2018, pp. 657–674.

[101] M. Li, J. Weng, J.-N. Liu, X. Lin, and C. Obimbo, "Bb-vdf: Enabling accountability and fine-grained access control for vehicular digital forensics through blockchain," Cryptology ePrint Archive, Report 2020/011, 2020, https://eprint.iacr.org/2020/011.

[102] A. Boudguiga, N. Bouzerna, L. Granboulan, A. Olivereau, F. Quesnel, A. Roger, and R. Sirdey, "Towards better availability and accountability for iot updates by means of a blockchain," in *Proc. of EuroS&PW*, 2017, pp. 50–58.

[103] R. Künnemann, I. Esiyok, and M. Backes, "Automated verification of accountability in security protocols," in *Proc. of IEEE CSF*, 2019, pp. 397–39 716.

[104] M. Li, L. Zhu, and X. Lin, "Efficient and privacy-preserving carpooling using blockchain-assisted vehicular fog computing," *IEEE Internet of Things Journal*, 2018.

[105] Y. Zhang, C. Xu, J. Ni, H. Li, and X. Shen, "Blockchain-assisted public-key encryption with keyword search against keyword guessing attacks for cloud storage," *IEEE Transactions on Cloud Computing*, DOI: 10.1109/TCC.2019.2923222, 2019.

[106] M. Li, J. Weng, A. Yang, W. Lu, Y. Zhang, L. Hou, J.-N. Liu, Y. Xiang, and R. H. Deng, "Crowdbc: A blockchain-based decentralized framework for crowdsourcing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 6, pp. 1251–1266, 2018.

[107] Y. Lu, Q. Tang, and G. Wang, "Zebralancer: Private and anonymous crowdsourcing system atop open blockchain," *arXiv preprint arXiv:1803.01256*, 2018.

[108] A. Dorri, M. Steger, S. S. Kanhere, and R. Jurdak, "Blockchain: A distributed solution to automotive security and privacy," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 119–125, 2017.

[109] D. Liu, J. Ni, C. Huang, X. Lin, and X. Shen, "Secure and efficient distributed network provenance for iot: A blockchain-based approach," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7564–7574, 2020.

[110] R. Frey, D. Wörner, and A. Ilic, "Collaborative filtering on the blockchain: a secure recommender system for e-commerce," in *Proc. of INFORMATION SYSTEMS SECURITY AND PRIVACY (SIGSEC)*, 2016.

[111] K. Nguyen, G. Ghinita, M. Naveed, and C. Shahabi, "A privacy-preserving, accountable and spam-resilient geo-marketplace," in *Proceedings of ACM SIGSPATIAL*, 2019, pp. 299–308.

[112] M. Klems, J. Eberhardt, S. Tai, S. Härtlein, S. Buchholz, and A. Tidjani, "Trustless intermediation in blockchain-based decentralized service marketplaces," in *International Conference on Service-Oriented Computing*. Springer, 2017, pp. 731–739.

[113] Y. Gu, X. Gui, P. Xu, R. Gui, Y. Zhao, and W. Liu, "A secure and targeted mobile coupon delivery scheme using blockchain," in *International Conference on Algorithms and Architectures for Parallel Processing*. Springer, 2018, pp. 538–548.

[114] A. Schaub, R. Bazin, O. Hasan, and L. Brunie, "A trustless privacy-preserving reputation system," in *IFIP International Information Security and Privacy Conference*. Springer, 2016, pp. 398–411.

[115] S. Hu, C. Cai, Q. Wang, C. Wang, X. Luo, and K. Ren, "Searching an encrypted cloud meets blockchain: A decentralized, reliable and fair realization," in *Proc. of IEEE INFOCOM*, 2018, pp. 792–800.

[116] C. Xu, C. Zhang, and J. Xu, "vchain: Enabling verifiable boolean range queries over blockchain databases," in *Proc. of SIGMOD*, 2019, pp. 141–158.

[117] S. Goldwasser, S. Micali, and C. Rackoff, "The knowledge complexity of interactive proof systems," *SIAM Journal on computing*, vol. 18, no. 1, pp. 186–208, 1989.

[118] M. Bellare and O. Goldreich, "On defining proofs of knowledge," in *Proc. of CRYPTO*. Springer, 1992, pp. 390–420.

[119] D. Pointcheval and O. Sanders, "Reassessing security of randomizable signatures," in *Proc. of CTRSA*. Springer, 2018, pp. 319–338.

[120] M. H. Au, W. Susilo, and Y. Mu, "Constant-size dynamic k-taa," in *International Conference on Security and Cryptography for Networks*. Springer, 2006, pp. 111–125.

[121] D. Boneh, X. Boyen, and H. Shacham, "Short group signatures," in *Proc. of CRYPTO*. Springer, 2004, pp. 41–55.

[122] B. Parno, J. Howell, C. Gentry, and M. Raykova, "Pinocchio: Nearly practical verifiable computation," in *Proc. of IEEE S&P*, 2013, pp. 238–252.

[123] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in *Proc. of CRYPTO*. Springer, 2001, pp. 213–229.

[124] J. Katz and Y. Lindell, *Introduction to modern cryptography*. CRC press, 2014.

[125] T. P. Pedersen, "Non-interactive and information-theoretic secure verifiable secret sharing," in *Proc. of CRYPTO*. Springer, 1991, pp. 129–140.

[126] J. M. Pollard, "Kangaroos, monopoly and discrete logarithms," *Journal of cryptology*, vol. 13, no. 4, pp. 437–447, 2000.

[127] A. Fiat and A. Shamir, "How to prove yourself: Practical solutions to identification and signature problems," in *Proc. of EUROCRYPT*. Springer, 1986, pp. 186–194.

[128] D. Pointcheval and O. Sanders, "Short randomizable signatures," in *Proc. of CTRSA*. Springer, 2016, pp. 111–126.

[129] J. Bootle, A. Cerulli, P. Chaidos, J. Groth, and C. Petit, "Efficient zero-knowledge arguments for arithmetic circuits in the discrete log setting," in *Proc. of CRYPTO*. Springer, 2016, pp. 327–357.

[130] B. Bünz, J. Bootle, D. Boneh, A. Poelstra, P. Wuille, and G. Maxwell, "Bulletproofs: Short proofs for confidential transactions and more," in *Proc. of IEEE S&P*, 2018.

[131] J. Bootle and J. Groth, "Efficient batch zero-knowledge arguments for low degree polynomials," in *Proc. of PKC*. Springer, 2018, pp. 561–588.

[132] I. Damgård, J. Luo, S. Oechsner, P. Scholl, and M. Simkin, "Compact zero-knowledge proofs of small hamming weight," in *Proc. of PKC*. Springer, 2018, pp. 530–560.

[133] R. W. Lai and G. Malavolta, "Subvector commitments with application to succinct arguments," in *Proc. of CRYPTO*, 2019, pp. 530–560.

[134] R. Gennaro, C. Gentry, B. Parno, and M. Raykova, "Quadratic span programs and succinct nizks without pcps," in *Proc. of EUROCRYPT*. Springer, 2013, pp. 626–645.

[135] E. Ben-Sasson, A. Chiesa, D. Genkin, E. Tromer, and M. Virza, "Snarks for c: Verifying program executions succinctly and in zero knowledge," in *Proc. of CRYPTO*. Springer, 2013, pp. 90–108.

[136] E. Ben-Sasson, A. Chiesa, E. Tromer, and M. Virza, "Succinct non-interactive zero knowledge for a von neumann architecture," in *Prof. of USENIX Security*, 2014, pp. 781–796.

[137] C. Costello, C. Fournet, J. Howell, M. Kohlweiss, B. Kreuter, M. Naehrig, B. Parno, and S. Zahur, "Geppetto: Versatile verifiable computation," in *Proc. of IEEE S&P*, 2015, pp. 253–270.

[138] J. Groth, "On the size of pairing-based non-interactive arguments," in *Prof. of EUROCRYPT*, 2016, pp. 305–326.

[139] A. Kosba, C. Papamanthou, and E. Shi, "xjsnark: a framework for efficient verifiable computation," in *Proc. of IEEE S & P*, 2018, pp. 944–961.

[140] libsnark: a C++ library for zkSNARK proofs. https://github.com/scipr-lab/libsnark. Accessed January 2020.

[141] D. Fiore, C. Fournet, E. Ghosh, M. Kohlweiss, O. Ohrimenko, and B. Parno, "Hash first, argue later: Adaptive verifiable computations on outsourced data," in *Proc. of ACM CCS*, 2016, pp. 1304–1316.

[142] M. Chase, C. Ganesh, and P. Mohassel, "Efficient zero-knowledge proof of algebraic and non-algebraic statements with applications to privacy preserving credentials," in *Proc. of CRYPTO*. Springer, 2016, pp. 499–530.

[143] S. Agrawal, C. Ganesh, and P. Mohassel, "Non-interactive zero-knowledge proofs for composite statements," in *Proc. of CRYPTO*, 2018, pp. 643–673.

[144] M. Campanelli, D. Fiore, and A. Querol, "Legosnark: Modular design and composition of succinct zero-knowledge proofs," in *Proc. of ACM CCS*, 2019.

[145] Z. Li, J. Kang, R. Yu, D. Ye, Q. Deng, and Y. Zhang, "Consortium blockchain for secure energy trading in industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3690–3700, 2018.

[146] J. Gregory, "The internet of things: revolutionizing the retail industry," *Accenture Strategy*, 2015.

[147] L. Ardito, A. M. Petruzzelli, U. Panniello, and A. C. Garavelli, "Towards industry 4.0: Mapping digital technologies for supply chain management-marketing integration," *Business Process Management Journal*, 2018.

[148] B. Nguyen and L. Simkin, "The internet of things (iot) and marketing: the state of play, future trends and the implications for marketing," *Journal of Marketing Management*, 2017.

[149] T. Minkus and K. W. Ross, "I know what youre buying: Privacy breaches on ebay," in *International Symposium on Privacy Enhancing Technologies Symposium*. Springer, 2014, pp. 164–183.

[150] K. Zhang, X. Liang, R. Lu, K. Yang, and X. Shen, "Exploiting mobile social behaviors for sybil detection," in *Proc. of IEEE INFOCOM*, 2015, pp. 271–279.

[151] Parity Ethereum. https://github.com/paritytech/parity-ethereum. Accessed October 2019.

[152] M. P. Andersen, J. Kolb, K. Chen, G. Fierro, D. E. Culler, and R. A. Popa, "Wave: A decentralized authorization system for iot via blockchain smart contracts," 2017.

[153] B. David, P. Gaži, A. Kiayias, and A. Russell, "Ouroboros praos: An adaptively-secure, semi-synchronous proof-of-stake blockchain," in *Proc. of EUROCRYPT*. Springer, 2018, pp. 66–98.

[154] J. Camenisch, A. Kiayias, and M. Yung, "On the portability of generalized schnorr proofs," in *Proc. of EUROCRYPT*. Springer, 2009, pp. 425–442.

[155] H. Liu, Y. Zhang, and T. Yang, "Blockchain-enabled security in electric vehicles cloud and edge computing," *IEEE Network*, vol. 32, no. 3, pp. 78–83, 2018.

[156] T. ElGamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," *IEEE transactions on information theory*, vol. 31, no. 4, pp. 469–472, 1985.

[157] T. P. Pedersen, "A threshold cryptosystem without a trusted party," in *Workshop on the Theory and Application of of Cryptographic Techniques*. Springer, 1991, pp. 522–526.

[158] Java Pairing based Cryptography. https://github.com/emilianobonassi/jpbc. Accessed July 2018.

[159] Proof-of-Authority (PoA) Chains. https://wiki.parity.io/Proof-of-Authority-Chains. Accessed January 2020.

[160] Web3j - Lightweight Java library for integration with Ethereum clients. https://docs.web3j.io/. Accessed January 2020.

[161] Solidity. https://solidity.readthedocs.io/en/v0.4.25/. Accessed January 2020.

[162] K. Suto, H. Nishiyama, and N. Kato, "Postdisaster user location maneuvering method for improving the qoe guaranteed service time in energy harvesting small cell networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 9410–9420, 2017.

[163] L. T. Tan and R. Q. Hu, "Mobility-aware edge caching and computing in vehicle networks: A deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10 190–10 203, 2018.

[164] F. Lyu, H. Zhu, H. Zhou, W. Xu, N. Zhang, M. Li, and X. Shen, "Ss-mac: A novel time slot-sharing mac for safety messages broadcasting in vanets," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3586–3597, 2017.

[165] Local Search Statistics. https://www.webfx.com/blog/marketing/google-ads-statistics/. Accessed March 2020.

[166] Z. Li, L. Chen, and Y. Wang, "G*-tree: An efficient spatial index on road networks," in *Proc. of IEEE ICDE*, 2019, pp. 268–279.

[167] Z. Qian, J. Xu, K. Zheng, P. Zhao, and X. Zhou, "Semantic-aware top-k spatial keyword queries," *World Wide Web*, vol. 21, no. 3, pp. 573–594, 2018.

[168] D. Zhang, Y. Li, X. Cao, J. Shao, and H. T. Shen, "Augmented keyword search on spatial entity databases," *The VLDB Journal*, vol. 27, no. 2, pp. 225–244, 2018.

[169] D. Liu, A. Alahmadi, J. Ni, X. Lin, and X. Shen, "Anonymous reputation system for iiot-enabled retail marketing atop pos blockchain," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3527–3537, 2019.

[170] W. Hua, Z. Wang, H. Wang, K. Zheng, and X. Zhou, "Short text understanding through lexical-semantic analysis," in *Proc. of IEEE ICDE*, 2015, pp. 495–506.

[171] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[172] N. Kato *et al.*, "Location awareness system for drones flying beyond visual line of sight exploiting the 400 mhz frequency band," *IEEE Wireless Communications*, DOI:10.1109/MWC.2019.1800570, 2019.

[173] A. Guttman, "R-trees: A dynamic index structure for spatial searching," pp. 47–57, 1984.

[174] S. Bowe, A. Gabizon, and M. D. Green, "A multi-party protocol for constructing the public parameters of the pinocchio zk-snark," in *Proc.of FC*. Springer, 2018, pp. 64–77.

[175] J. Garay, A. Kiayias, and N. Leonardos, "The bitcoin backbone protocol: Analysis and applications," in *Proc. of EUROCRYPT*, 2015, pp. 281–310.

[176] Y. Zhang, C. Papamanthou, and J. Katz, "Alitheia: Towards practical verifiable graph processing," in *Proc. of ACM CCS*, 2014, pp. 856–867.

[177] Y. Zhang, D. Genkin, J. Katz, D. Papadopoulos, and C. Papamanthou, "vsql: Verifying arbitrary sql queries over dynamic outsourced databases," in *Prof. of IEEE Symposium on S&P*, 2017, pp. 863–880.

[178] C++ library for Finite Fields and Elliptic Curves. https://github.com/scipr-lab/libff. Accessed January 2020.

[179] Ethereum Status. https://etherscan.io. Accessed October 2019.

[180] F. Ye, Y. Qian, and R. Q. Hu, "A real-time information based demand-side management system in smart grid," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 2, pp. 329–339, 2015.

[181] J. Chen, K. Hu, Q. Wang, Y. Sun, Z. Shi, and S. He, "Narrowband internet of things: Implementations and applications," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 2309–2314, 2017.

[182] L. Zhu, J. Zhang, Z. Xiao, X. Cao, D. O. Wu, and X.-G. Xia, "Joint tx-rx beamforming and power allocation for 5g millimeter-wave non-orthogonal multiple access networks," *IEEE Transactions on Communications*, vol. 67, no. 7, pp. 5114–5125, 2019.

[183] Advertising Industry Statistics. https://www.emarketer.com/content/mobile-advertising-is-expected-to-surpass-tv-ad-spending Accessed March 2020.

[184] M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, and A. Rieke, "Discrimination through optimization: How facebook's ad delivery can lead to skewed outcomes," *arXiv preprint arXiv:1904.02095*, 2019.

[185] H. Liu and J. Chen, "Distributed privacy-aware fast selection algorithm for large-scale data," *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, no. 2, pp. 365–376, 2017.

[186] J. Misic, V. B. Misic, X. Chang, S. G. Motlagh, and Z. M. Ali, "Modeling of bitcoin's blockchain delivery network," *IEEE Transactions on Network Science and Engineering*, 2019, DOI: 10.1109/TNSE.2019.2928716.

[187] Y. Xu, C. Zhang, Q. Zeng, G. Wang, J. Ren, and Y. Zhang, "Blockchain-enabled accountability mechanism against information leakage in vertical industry services," *IEEE Transactions on Network Science and Engineering*, 2020, DOI: 10.1109/TNSE.2020.2976697.

[188] W. Tang, J. Ren, and Y. Zhang, "Enabling trusted and privacy-preserving healthcare services in social media health networks," *IEEE Transactions on Multimedia*, vol. 21, no. 3, pp. 579–590, 2018.

[189] J. Ni, K. Zhang, Q. Xia, X. Lin, and X. Shen, "Enabling strong privacy preservation and accurate task allocation for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, 2019, DOI: 10.1109/TMC.2019.2908638.

[190] J. Camenisch, M. Drijvers, and M. Dubovitskaya, "Practical uc-secure delegatable credentials with attributes and their application to blockchain," in *Proc. of ACM CCS*, 2017, pp. 683–699.

[191] D. Chen, N. Zhang, R. Lu, N. Cheng, K. Zhang, and Z. Qin, "Channel precoding based message authentication in wireless networks: Challenges and solutions," *IEEE Network*, vol. 33, no. 1, pp. 99–105, 2018.

[192] J. Shao, R. Lu, Y. Guan, and G. Wei, "Achieve efficient and verifiable conjunctive and fuzzy queries over encrypted data in cloud," *IEEE Transactions on Services Computing*, 2019, DOI: 10.1109/TSC.2019.2924372.

[193] JSnark. https://github.com/akosba/jsnark. Accessed January 2020.

[194] Ethereum Gas Table. https://ethgastable.info. Accessed March 2020.

[195] EIP 1108. Reduce alt-bn128 precompile gas costs. https://eips.ethereum.org/EIPS/eip-1108. Accessed Febauary 2020.

[196] A. Anjum, M. Sporny, and A. Sill, "Blockchain standards for compliance and trust," *IEEE Cloud Computing*, vol. 4, no. 4, pp. 84–90, 2017.

Author's Publications

1. D. Liu, A. Alahmadi, J. Ni, X. Lin, and X. Shen, "Anonymous Reputation System for IIoT-enabled Retail Marketing atop PoS Blockchain," IEEE Transactions on Industrial Informatics, vol. 15, no. 6, pp. 3527–3537, 2019.

2. D. Liu, J. Ni, X. Lin, and X. Shen, "Transparent and Accountable Vehicular Local Advertising with Practical Blockchain Designs," IEEE Transactions on Vehicular Technology, Accepted with minor.

3. D. Liu, C. Huang, J. Ni, X. Lin, and X. Shen, "Exploiting Blockchain for Transparent Mobile Advertising with Privacy-preserving Accountability," IEEE Transactions on Network Science and Engineering, under revision.