Privacy-Enhanced Payment Systems*

Agostino Capponi† Michael Junho Lee‡ Brian Zhu§

Abstract

Technological innovations enable digital privacy, but pose fundamental conflicts between freedom and control. We study the design of privacy-enhanced payment systems, valued for legitimate transactions but vulnerable to illicit financial activities. We distinguish two dimensions of privacy: *identity privacy*, the masking of beneficiary owners of accounts, and *transaction privacy*, confidentiality in activity. Identity-private systems are suboptimal, enabling illicit operations to scale by distributing activity across multiple accounts. The optimal provision of transaction privacy is complex and tailored to payoff frontiers of illicit actors but adopts a simple design when considering robustness to innovations in techniques of illicit actors. Our framework generates a ranking of payment systems; public blockchains, despite abuse of identity-privacy, can dominate legacy systems under strong enforcement capabilities. However, when enforcement capabilities are too strong, the dominance switches in favor of legacy systems as public blockchains fail to provide meaningful privacy to users.

Keywords: Privacy, payment systems, regulation, innovation, blockchain

JEL Classification Numbers: E41, E42, G20, G23, G28

*The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. Any and all errors are ours.

†Columbia University. Email: ac3827@columbia.edu.

‡Federal Reserve Bank of New York. Email: michael.j.lee@ny.frb.org.

§Columbia University. Email: bzz2101@columbia.edu.

1 Introduction

Privacy has been argued to have intrinsic social and economic value (Acquisti, Taylor and Wagman 2016). Yet, the pervasive digitalization of economic activity enables surveillance and monetization of the vast amounts of personal data. Payments lie at an important juncture of this value chain due to the highly sensitive and valuable nature of the information contained in payments (Garratt and Van Oordt 2021). With the growing obsolescence of physical cash, both private and public sectors today explore technologies to imbue payment systems with privacy features that are reminiscent of cash (Kahn, McAndrews and Roberds 2005).

Recent innovations offer technical solutions to enhance privacy.[1] Blockchains, distributed ledgers with lists of records securely linked together via cryptographic hashes, enable anonymous users to transfer value among each other (Nakamoto 2008). Cryptographic protocols, including those that use zero-knowledge proofs (Sasson, Chiesa, Garman, Green, Miers, Tromer and Virza 2014, Van Saberhagen 2013), whether applied at the system or program layer, enhance privacy by obfuscating details of transactions, e.g. information about the sender and/or receiver. These advancements expand on the realm of possibilities for *technological* provision of privacy in digital value-transfer systems.

At the same time, financial technologies have been heavily implicated in a wide span of illegal financial and economic activities (Department of Homeland Security 2022), including money laundering (Coelho, Fishman and Ocampo 2021), market manipulation (Cong, Li, Tang and Yang 2023c), financial fraud (Phua, Sang, Wei and Yu 2022), narcotics trafficking (TRM Labs 2024), and cyber crime (Cong, Campbell, Rabetti and Wu 2023a). Black markets, which historically relied on cash or trusted intermediaries to settle transactions (Rogoff (2016)), have now shifted to and have even become empowered by cryptocurrencies.[2] Integration with traditional financial markets further strengthens the effectiveness with which blockchain-based systems facilitate illicit activities. Private payments pose material challenges to governments' powers to implement policy and enforce the law (Brownworth, Durfee, Lee and Martin 2024), and have become a top risk to national and economic security for developed and developing countries alike (White House 2022, Europol 2023, Treasury 2024).[3]

This dichotomous nature of privacy presents a fundamental conflict between *freedom* and

[1] Or more specifically, *hard* privacy, that is, technological-assured privacy as described in Auer, Böhme, Clark and Demirag (2025).

[2] Cong et al. (2023a) show that ransomware gangs have transformed into complex, corporate-style organizations featuring physical offices, franchise models, and affiliate networks through the use of cryptocurrencies.

[3] Mounting evidence indicates that crypto-asset systems may facilitate capital outflows (Hu, Lee and Putniņš 2021, Chen and Sarkar 2022), reduce government tax revenues (Meling, Mogstad and Vestre 2024), and enable corruption (Alnasaa, Gueorguiev, Honda, Imamoglu, Mauro, Primus and Rozhkov 2022).

control. For governments and regulatory agencies, what form should policy take with regard to privacy-enhanced payment systems and services? For central banks interested in designing payment systems and central bank digital currencies (CBDCs), how can privacy be integrated to achieve key objectives, such as fostering consumer protection and ensuring safe, efficient payment systems (Garratt and Lee 2021)? For private sector payment providers, what does it mean to provide privacy in a balanced and compliant way? We formally explore this conflict in the design of privacy-enhanced payment systems.

We formulate a model in which an operator considers the design of a privacy-enhanced payment system that allows users to create accounts and use it to settle transactions. Users intrinsically value privacy for different reasons: *good* users value privacy for personal, sensitive transactions; *bad* users seek privacy in order to make illicit transactions whilst evading ex-post enforcement. Our model of privacy is multi-dimensional: a system can offer *identity-privacy*, i.e. the privacy of a user's identity in relation to accounts, and offer (state-contingent) *transaction-privacy*, i.e. privacy regarding an account's transactions. The operator can design the system to guarantee privacy without recourse, but the provision of privacy directly hampers the ability to detect illicit activity. Hence, privacy enhancements to the payment system act as an irrevocable ex-ante commitment to the provision of privacy.

In the model, bad users possess a set of technologies, with varying production characteristics, that mimic payment patterns of good users, making it impossible to discern from payment characteristics alone.[4] These technologies represent technical and operational apparatus developed by criminal enterprises to mix illicit transactions with legal businesses or activity that make it difficult to discern without comprehensive forensic investigations (Europol 2023). Hence, although the operator can choose to condition transaction-privacy based on observable actions or states, volume serves as a proxy for intent and when bad users utilize technologies, account activity is distinguished primarily by total transaction volume.[5]

Our first key result is that in identity-private systems, any provision of transaction-privacy is suboptimal from a welfare perspective. When identities are private, transaction-privacy can only be imposed at the account (rather than the user) level. For any fixed transaction-privacy rule, we show that bad users strategically select a subset of technologies to meet the conditions for transaction-privacy, and scale their operations across multiple accounts. This renders any

[4]In the background, we assume that the operator is able to deny privacy to any known transaction patterns strongly associated with illicit activity.

[5]For example, Chainalysis' 2023 Crypto Crime Report revealed that illicit cryptocurrency transaction volumes reached an all-time high of $20.1 billion in 2022. Despite this record figure, these transactions accounted for only 0.24% of the total cryptocurrency transaction volume for the year. The report highlights that certain categories, such as transactions involving sanctioned entities, experienced an extraordinary increase of over 10,000%, comprising 44% of all illicit activity in 2022.

provision of transaction-privacy prohibitively costly from a social standpoint.

Scalability under identity-privacy extends to identity-private systems without transaction-privacy, such as public blockchains like Bitcoin and Ethereum. Under no transaction-privacy, all monitor transaction activity, with varying effectiveness dependent on transaction volume. In equilibrium, the bad user employs a strategy that resembles the use of two common forms of money laundering: "structuring," whereby payments are made in smaller values to evade detection, and "smurfing," whereby payments are processed via multiple accounts to obfuscate links between transactions. These techniques are found to be utilized in legacy and token systems.[6] Importantly, this strategy lowers the effectiveness of enforcement and enables bad users to achieve high transaction throughput. Ergo, identity-privacy systems are shown to be generally suboptimal.

Without identity-privacy, users still have opportunities to make private payments via transaction-privacy, which hides the identities of senders and recipients. Since identities are not private, transaction-privacy can now be specified at the *user* level, limiting the scalability of bad user activity through the use of multiple accounts. In the baseline model, the equilibrium payment system design is an identity-required system with a "complex" transaction-privacy rule. The transaction-privacy rule is complex in the sense that it provides privacy to disjoint ranges of transaction volume. These ranges are selected to maximize the social benefits of privacy to good users while limiting the scope of activity for bad users.

Intuitively, the regions of transaction-privacy typically consist of low-value payments and large-value payments that depend on the *payoff frontier* of bad user technologies. In the low-value region, the payment demand of good users is large while the low transaction limit imposes a ceiling on bad users' activity. In the high-volume region, bad users find it unprofitable to conduct activities based on their technologies, meaning that the operator can provide transaction-privacy knowing that the possibility of bad user activity is low.

In the baseline model, implicit requirements are that the operator is sufficiently informed about bad users' technologies, and that the set of technologies and their properties are stable. However, in practice, criminal enterprises are known to constantly innovate around legal and enforcement boundaries in order to avoid suspicion (Erol and Lee 2019). This ability to evade authorities has been boosted through early adoption of emerging technologies, including encrypted messaging[7], privacy-centric devices[8], and more recently, cryptocurrency mixers

[6] Hong Kong authorities uncovered a multi-year, sophisticated money-laundering operation amounting to $14 billion, applying various techniques to hide its illicit nature. Similar techniques are shown to be utilized by Lazarus Group, a sanctioned North Korean hacking group, known to have stolen roughly $1.7 billion in cryptoassets.

[7] Examples include Telegram and the Daren Li case.

[8] Examples include the utilization of devices provided by Sky Global and Phantom Secure

(Brownworth et al. 2024).

We extend our model to incorporate the possibility of bad users engaging in innovation. Two forms of innovation are considered. First is *incremental* innovation, whereby existing technologies' productivity are enhanced through investment. These could represent efforts and investments to improve operational efficiency, or make adjustments to increase the scale of a particular type of arrangement. The second form is *radical* innovation, whereby an entirely new technology is developed to expand on the existing technology set. A prime example is acquiring talent and building technological infrastructure to run operations through encrypted systems, or developing a new business model that offers differentiated economies of scale.

Our key result is that the equilibrium payment system design arising in the presence of innovation is characterized by two simple, implementable features: identity-required with a simple cutoff over transactions are auditable. This design exhibits robustness in the sense that the system is resilient to any future changes in the capabilities of bad users. Both forms of innovation push the operator to adopt a robust design, but for different reasons. With incremental innovation, bad users make investments to shift their production technology upward, which make transaction-privacy too costly for higher payment volumes for the same utility profile of good users. With radical innovation, bad users target transaction-privacy regions with customized technologies, essentially exploiting regions where transaction-privacy exists as "loopholes." The operator chooses the transaction-privacy rule accordingly to mitigate any possibility for future technologies to exploit privacy regions beyond tolerable levels.

Our model provides normative implications for privacy-enhanced payment systems being considered by central banks. Di Iorio, Kosse and Mattei (2024) find that the vast majority of central banks across developed and developing countries are exploring or have already begun the issuance of CBDCs, with privacy as an important concern for retail CBDCs.[9] Our results reveal potential shortcomings of current approaches to privacy in CBDCs.

One common design consideration is to relax standards for creating accounts with limited functionality with the intention of enhancing privacy or financial inclusion.[10] In the context of our framework, this corresponds to some degree of identity-privacy. We demonstrate how identity-private systems, even with rigid transaction limits, are generally undesirable due to

[9]For example, the European Central Bank's Digital Euro project recognizes privacy as a priority, and states:

> A digital euro would be designed to provide a higher level of privacy than other digital payment methods typically offer. The Eurosystem would not identify people based on their payments. Moreover, personal transaction details from offline digital euro payments would be known only to the payer and the payee.

See explainer on the Digital Euro and ECB's memo on privacy for more detail.

[10]For example, the Sand Dollar and Digital Yuan allow for the creation of wallets with minimal or no identity requirements, though balance and transaction limits are still imposed.

the potential for scalable abuse by illicit actors.[11] In addition, several CBDC designs also impose limits on balances in wallets. Our model suggests that transaction throughput is the main determinant of the degree of illicit activity on payment systems. Given this, caps on balances, rather than flows, may be inadequate for curtailing illegal activity while presenting another constraint on the adoption of privacy-enhanced payment systems by legitimate users. Additionally, our framework also captures the ideas of "anonymity vouchers" (European Central Bank (2019)) and "controlled anonymity" ((Pocher and Veneris 2022)) that have been explored by various central banks, which correspond to some provisions of transaction-privacy.[12] Our results thus provide implications for the implementation of these ideas in CBDCs.

Finally, we use our framework to classify the broad spectrum of legacy and token payment systems based on their privacy features. In addition to the optimal privacy-enhanced payment system characterized in the paper, this covers three other broad classes of existing payment systems. Public blockchain systems, such as Bitcoin and Ethereum, provide identity-privacy but no transaction-privacy for basic transactions. Privacy-centric systems, such as ZCash and Monero, offer both identity- and transaction-privacy. Legacy payment systems, including those on private payment rails (such as Venmo, Visa), and interbank payment rails (banking system, ACH), provide neither forms of privacy.

Our model implies a ranking of payment systems, from a social and regulatory perspective, based on the types and degree of privacy provided. Predictably, a identity-required system with a prudent transaction-privacy policy dominates all other systems while identity-private systems with transaction-privacy are inferior to all other systems due to uninhibited bad user activity. The relative ranking between identity-private and identity-required systems with no transaction-privacy depends on *enforcement capability*, or the general effectiveness with which the enforcer is able to detect bad user activity. When enforcement capabilities are poor, identity-private systems offer too much scalability in bad user activity. However, with strong enforcement, identity-private systems provide limited privacy to good users, compared to identity-required systems, which offer no privacy to any users. The pecking order based on enforcement capability offers clear policy prescriptions based on a jurisdiction's resources to monitor and enforce activity. In jurisdictions where resources or technical expertise is limited, identity-private systems, even with no transaction-privacy, creates significant economies

[11] Similar principles are applied to cash-based transactions to curtail money laundering. For example, the EU is set to impose restrictions on cash payments exceeding EUR 10,000, setting uniform standards on top of restrictions enforced by individual countries. Some jurisdictions, such as the UK, impose cash restrictions on high-value goods dealers recognizing AML considerations.

[12] The ECB introduced anonymity vouchers that when used, allow CBDC transfers to be anonymous. The People's Bank of China proposed the concept of controlled or managed anonymity that shields all eCNY transactions from third-party viewing, but only small-value transactions from all parties including the central bank itself.

of scale for illicit actors that make the provision of privacy costly. In contrast, in jurisdictions where enforcement capabilities are sufficiently strong, privacy enhancements may be desirable relative to legacy systems.

Differences in enforcement capabilities could be pivotal. Whereas U.S. enforcement agencies have leveraged blockchain traceability to track transactional drug trafficking networks and seize assets, the general effectiveness in monitoring and enforcement varies significantly based on expertises and resources (Chainalysis 2025). These capabilities may further diverge for privacy-centric cryptocurrencies, such as Monero, which is implicated in illicit activity.[13]

Related Literature. Our study contributes to the literature on privacy in payment systems[14], whose importance has long been recognized (Kahn et al. (2005)), but has re-emerged in interest with the advent of new technologies that allow for transformative methods of providing privacy. Privacy exhibits features of a public good due to information externalities (Garratt and Van Oordt 2021) and is a key consideration in the design of central bank digital currencies (Lee and Garratt (2021), Ahnert, Hoffmann and Monnet (2022), Wang (2023), and Uhligh, Alonso and Frost (2023)). We refer to Monnet, Frost, Gambacorta, Auer and Rice (2022) for a review of CBDCs focusing on economic considerations related to operational architecture, technology, and privacy.[15] Relative to these studies that treat privacy as a monolithic good and an intrinsically valued commodity, our paper is the first to unpack privacy on both dimensions: we consider a "gradient" of privacy based on identity, transaction sizes, and contingencies, and based on socially valued and socially costly motives for valuing privacy. This allows us to provide sharp, normative implications on the design principles for privacy-enhanced payment systems, whether oriented to government initiatives or regulatory guardrails put in place for private-sector initiatives.

We highlight a new dimension to consider in the regulation of payments. Existing studies have focused on policies targeting systemic risk, for instance by detecting insolvency probabilities in payment networks (Rochet and Tirole (1996)) or providing capital to absorb the systemic implications of these defaults (Acharya, Engle and Richardson (2012)). Other studies have highlighted the importance of policies which increase competitiveness, such reducing the costs of card-based payments (Rochet and Tirole (2006)). Our unique contribution lies in using

[13] For example, see this.

[14] A subset of papers consider the informational role of payment data on lending, competition, and welfare (e.g. Parlour, Rajan and Zhu (2022), He, Huang and Zhou (2023), and Ghosh, Vallee and Zeng (2022)).

[15] Some studies abstract away from privacy, analyzing competition between CDBCs and private digital money (Cong and Mayer (2024)), and interactions between traditional banking sector and CDBCs (Villaverde, Sanches, Schilling and Uhlig (2021), Villaverde, Schilling and Uhlig (2020), and Whited, Wu and Xiao (2022)).

our framework to evaluate the social impact of both legacy and token-based payment systems based on their privacy features, and leveraging these rankings to identify where enforcement efforts are most effective.

Privacy is an ever-growing concern and a key consideration in the design of decentralized finance (DeFi) applications, as financial transactions comprise sensitive information yet these applications are built on public blockchains. The risk of frontrunning from the transparency of blockchains has led to the introduction of private submission pools for transactions (Daian, Goldfeder, Kell, Li, Zhao, Bentov, Breidenbach and Juels 2020). Capponi, Jia and Wang (2024) study the economics of these private pools, demonstrating that they increase social welfare but are not fully adopted due to misaligned incentives. Cryptographic protocols called zero-knowledge proofs have also been integrated into DeFi, allowing for transactions to be verified without revealing any underlying details. Mechanisms have been proposed to balance the privacy enhancements of ZKPs with regulatory standards, such as the *Privacy Pools* protocol ((Buterin, Illum, Nadler, Schär and Soleimani 2024)) and *involuntary selective de-anonymization* ((Sahu, Gajera, Chaudhary and Ivey-Law 2023a)). We contribute to these aspects of financial privacy by (i) showing the suboptimality of public blockchain-based platforms in balancing privacy benefits with illicit activity and (ii) showing that zero-knowledge protocols offering state-contingent privacy can be used in payment system design to optimize social good. Our findings provide guidance on identity and privacy considerations for tokenized real-world assets, for which privacy may facilitate adoption.

Our paper contributes to the broader literature on the design of tokenized platforms. The emergence of distributed ledger technologies and their transformative potential for financial services has raised important questions about their optimal design (Townsend 2020). In these systems, tokens can have their supply dynamically adjusted to enable optimal investment in platform quality over time (Cong, Li and Wang 2021), be used as utility tokens to foster competition (Goldstein, Gupta and Sverchkov 2023) or enable platform owners to relinquish control through decentralized governance mechanisms (Sockin and Xiong 2023), and be used to eliminate settlement risk by enforcing immediate transaction implementation (Lee, Martin and Townsend 2024b) or enable asset programmability to reduce settlement uncertainty (Lee, Martin and Townsend 2024a). Other studies have analyzed the design of tokenized systems for asset trading (Chiu and Koeppl (2019)), their potential for achieving settlement efficiency (e.g. Lee et al. (2024b), Lee et al. (2024a)), and joint design of trading and settlement systems (Lee et al. 2024a). We contribute to this body of literature by focusing on the value and cost of privacy in the design of financial platforms.

Our paper is also related to the literature on the economics of crime and illicit financing

(Becker (1968)). Calvo-Armengol and Zenou (2004) argue that social ties amplify delinquent behavior by enabling malicious actors to learn and adapt within their networks. These dynamics underscore the importance of robust designing systems to limit the value bad users can extract. In the context of DeFi, Foley, Karlsen and Putnins (2019) estimate that 44% of Bitcoin transactions are associated with illicit activity, and Cong, Harvey, Rabetti and Wu (2023b) give strong evidence of sophistication in crypto-enabled cybercrimes. Mounting evidence indicates that crypto-asset systems could facilitate capital outflows (Hu et al. 2021, Chen and Sarkar 2022), reduce government tax revenues (Meling et al. 2024), evade capital control in emerging markets (Graf von Luckner, Reinhart and Rogoff 2023, von Luckner, Reinhart and Rogoff 2023, Biais, Rochet and Villeneuve 2024), and enable corruption (Alnasaa et al. 2022). Our framework captures these perverse incentives by incorporating both incremental and radical innovation in criminal enterprises.

2 Environment

We begin by outlining the model environment. In the model, an operator designs a payment system, which allows users to create accounts that can be used to settle transactions. Users value privacy in payments for various reasons: some are "good" (i.e. civil liberties), and some are "bad" (i.e. money laundering). Bad payments impose negative social externalities on society, and an outside enforcer seeks to punish users engaging in such transactions. The operator does not know *a priori* whether a payment system user is of good or bad type (only the probability β), and must rely on the enforcer. On the other hand, the operator does not have an *ex-post* tool to restrict the enforcer from accessing transaction data, but can incorporate privacy-enhanced features in the design of the payment system, thereby achieving *ex-ante commitment* to the provision of privacy. We proceed to characterize strategies, payoffs, and other basic mechanics of our model.

2.1 Benchmark Model

There are three periods that are indexed by $t = 0, 1, 2$ and three agents: an operator, a user, and an enforcer.

Users and Payments. At $t = 0$, the user privately learns their type θ, which can be good ($\theta = g$) with probability β, or bad ($\theta = b$) with probability $1 - \beta$.[16] "Good" users realize a

[16]For exposition, we consider a single user of hidden type; results generalize to a setting with a mass of users.

stochastic demand for payments that are sensitive in nature, and incur disutility from non-privacy. In contrast, "bad" users seek to settle transactions that facilitate illicit activity.

If a user is good, they are assumed to use a single account to submit a random payment demand with total volume of q_g, for some continuous random variable q_g with distribution F_g supported in the interval $[\underline{q}, \overline{q}]$ where $0 \leq \underline{q} < \overline{q}$. Settling payments yields a normalized payoff of q_g. If the user's transactions can be traced back to their identity, then the payoff is $(1 - \delta_g)q_g$, for some $\delta_g \in (0, 1)$, reflecting a disutility from non-privacy for sensitive transactions.

If a user is bad, they seek to settle illicit transactions without recourse from the enforcer. The bad user has access to a set $[T]$ of *technologies* that can mimic the transaction patterns of a good user. These technologies represent various technical and operational measures, such as mixing illicit transactions with legal businesses, that make it difficult to separate without comprehensive forensic investigations. Hence, we assume that bad payments processed through technologies are ex-ante indistinguishable from good payments, and the only observable characteristic is total payment volume.

Technologies vary in scalability, representing different efficiencies in mimicking legal transactions. Technology $i \in [T]$ is specified by (y_i, c_i, ϕ_i), where $y_i(q) : \mathbb{R}_+ \to \mathbb{R}_+$ represents how much payment volume can be mimicked given an input volume q (i.e. the *production* function), $c_i(Q_i) : \mathbb{R}_+ \to \mathbb{R}_+$ represents the cost of inputting a volume of Q_i through the technology (i.e. the *variable cost* function), and $\phi_i \in \mathbb{R}_{++}$ represents a start-up cost that is incurred if technology i is used (i.e. the *fixed cost*). For tractability, we set production to be *linear* in q, i.e. $y_i(q) = \theta_i q$ for $\theta_i > 0$, and cost is *quadratic* in q, i.e. $c_i(q) = \eta_i q^2$ for $\eta_i > 0$. Thus, the overall payoff f_i of sending a volume of Q_i thorugh technology i, before any applicable disutilities of non-privacy, is

$$f_i(q) = \theta_i q - \eta_i q^2 - \phi_i.$$

The linear-quadratic characterization of technologies excludes the possibility that any single technology is infinitely scalable. We make the additional assumption regarding the technology set:

Assumption 1. *The bad user's set of technologies are sufficiently effective, in the sense that*

- *There exists a technology with sufficiently low fixed cost, i.e.* $\exists\, i \in [T]$ *such that* $\phi_i < \frac{\theta_i}{\zeta e}$.
- *There exists a technology with sufficiently large scale, i.e.* $\exists\, i \in [T]$ *such that* $\overline{q} < \theta_i / \eta_i$.

The bad user chooses the subset $S \subseteq [T]$ of technologies to use, the number N of accounts to create (if multiple accounts are permissible), and transactions volumes $Q_{a,i}$ to sub-

mit through the each technology per account, receiving a payoff of the total output minus the total costs. A multiplicative disutility of non-privacy $\delta_b = 1$ is incurred on the output if transactions can be traced back to their identity. Hence, disutility associated with non-privacy is strictly greater for the bad user relative to good users.[17]

Payment System Design. The operator designs a payment system in which users can create accounts and submit transactions through the accounts. The operator decides on a set of rules governing the payment system, namely on *identity-privacy*, i.e. whether users must reveal their identity in order to create an account, and *transaction-privacy*, i.e. whether and when certain transactions are granted privacy. A private transaction has critical data, including the sender and receiver of the transaction, obfuscated from the operator and enforcement, even when there is no identity-privacy. Formally, the operator's the action and strategy spaces are given by $(\mathcal{I}, p) \in \{0,1\} \times (\mathbb{R}_+ \to \{0,1\})$, where:

- Identity-privacy (IP) rule: $\mathcal{I} \in \{0,1\}$. If $\mathcal{I} = 1$, then users can create arbitrarily many accounts on the payment system without providing identity. If $\mathcal{I} = 0$, or *identity-required* (IR), then account creation requires a mapping between a user's real-world identity and that user's address on the payment system. In this case, the user's identity is known only to the payment system and can only be made known to enforcement depending on the transaction-privacy rule and the user's transaction volume. Account activity is assumed consolidated at the user level, such that there is a single representative account per user.

- Transaction-privacy (TP) rule: $p : \mathbb{R}_+ \to \{0,1\}$. The mapping p takes as input the total volume q of submitted transactions of an account, and outputs $\{0,1\}$ indicating whether transactions originating from that account are transaction-private (i.e. 1) or transaction-auditable (TA) (i.e. 0).

Enforcement. An enforcer monitors all public transactions on the payment system, seeking to detect illicit transactions submitted by bad users ex-post.[18] In systems with IR ($\mathcal{I} = 0$), the enforcer is able to detect bad user activity among public payments with probability 1. In IP systems ($\mathcal{I} = 1$), the enforcer possess a technology that imperfectly traces transactions to identities at the account level. This technology is specified by a detection function $\pi : \mathbb{R}_+ \to$

[17] Since payoff conditional on non-privacy is 0, this could be interpreted as a penalty corresponding to a complete seizure of all wealth associated with account(s) controlled by the bad user.

[18] In other words, although bad transactions are ex-ante indistinguishable from good transactions, the enforcer is able to investigate and sometimes unravel their true nature.

$[0,1]$ which specifies the probability of successfully tracing a transaction to its originator's identity that depends on total transaction volume q of an account. We assume detection probability is given by

$$\pi(q) = \min\{\zeta q, 1\}$$

for $\zeta > 0$, where larger ζ represents stronger *enforcement capabilities*.[19] The likelihood of detection probability is assumed to be independent across accounts, including those owned by the same user.

Strategy Space and Payoffs. The bad users' action set is characterized as follows. If $\mathcal{I} = 1$ (i.e. the system is identity-priate), then the bad user chooses the number $N \in \mathbb{Z}_+$ of accounts to create, the subset $S \subseteq [T]$ of technologies to use, and a *transaction volume matrix* $Q \in \mathbb{R}_+^{N \times |S|}$ where the (a,i)-th entry $Q_{a,i}$ denotes the volume sent on account a through technology i. Given S, we define the bad user's *aggregate fixed cost* as

$$\Phi_S = \sum_{i \in S} \phi_i$$

and the *per-account transaction volume* as

$$q_a = \sum_{i \in S} Q_{a,i}.$$

Denoting summations over $a \in [N]$ and $i \in S$ as simply "a" and "i," respectively, the bad user's payoff expected payoff is

$$u_1(N,S,Q) = \Big((1 - \psi(N,S,Q)) + (1 - \delta_b) \cdot \psi(N,S,Q)\Big) \sum_{a,i} \theta_i Q_{a,i} - \sum_{a,i} \eta_i Q_{a,i}^2 - \Phi_S \qquad (1)$$

where $\psi(N,S,Q)$ is the overall probability of detection given by

$$\psi(N,S,Q) = 1 - \prod_a \Big(1 - \pi(q_a) \cdot \mathbb{1}_{p(q_a)=0}\Big). \qquad (2)$$

[19]The enforcer may represent law enforcement and other agencies on various value-transfer systems, including legacy payment systems to permissionless blockchains, or monitoring by the operator for compliance purposes. In payment systems where identity is required, non-private transactions contain identity-revealing metadata for the enforcer. In identity-private payment systems, unmasking identities through activity and other sources can be more challenging, and may become increasingly challenging for accounts with limited activity and depend on the sophistication and resources of the enforcer. We capture this feature through the explicit form of π.

Note that this is the probability that *at least* one account is detected by the enforcer to be fraudulent, as the detection of just one account is sufficient for the enforcer to identify the bad user.

If $\mathcal{I} = 0$ (i.e. identity is required to use the payment system), then the bad user effectively is limited to a single account ($N = 1$), but still decides on technology selection S and transaction volumes $Q = (Q_i)_{i \in S}$ sent through each selected technology.[20] Letting

$$q_b = \sum_{i \in S} Q_i$$

denote the total volume transacted by the bad user, the bad user's expected payoff is

$$u_0(S, Q) = \left(1 - \delta_b \cdot \pi(q_b) \cdot \mathbb{1}_{p(q_b)=0}\right) \sum_i \theta_i Q_i - \sum_i \eta_i Q_i^2 - \Phi_S. \tag{3}$$

Operator's Objective. The goal of the operator is to design a payment system that maximize an objective that balances *welfare benefit*, or the value of privacy accruing to good users, and a *welfare cost*, or the social loss associated with bad user activity. The welfare benefit denoted $WB(\mathcal{I}, p)$, this is the payoff of the good user, given by:

$$WB(\mathcal{I}, p) = \begin{cases} \mathbb{E}[(1 - \delta_g \cdot \pi(q_g) \cdot \mathbb{1}_{p(q_g)=0})\, q_g] & \mathcal{I} = 1 \\ \mathbb{E}[(1 - \delta_g \cdot \mathbb{1}_{p(q_g)=0})\, q_g] & \mathcal{I} = 0. \end{cases} \tag{4}$$

The welfare cost, denoted $WC(\mathcal{I}, p, q)$, is the bad user's total transaction volume submitted to each technology via the payment system:

$$WC(\mathcal{I}, Q) = \begin{cases} \sum_{a,i} Q_{a,i} & \mathcal{I} = 1 \\ \sum_i Q_i & \mathcal{I} = 0. \end{cases} \tag{5}$$

For a payment system with rules $(\mathcal{I}, p)$ and bad user's transactions, the operator's objective, called the *net welfare*, is the welfare benefit less the welfare cost, both probability-weighted. We allow for a flexible representation of the objective based on the preferences of the operator

[20] We remark that identity-privacy with only "one account" is not a meaningful constraint to be imposed. The default action space is that the bad user can create multiple accounts, but these only matter for identity-private systems since they are not associated with each other.

using weights ω_g and ω_b for the welfare benefit and cost components, respectively:

$$W(\mathcal{I}, p, Q) = \omega_g(1-\beta) \cdot WB(\mathcal{I}, p) - \omega_b \beta \cdot WC(\mathcal{I}, Q).$$

Remark 1. *In this paper, we focus on a setting in which ex-ante discrimination between good and bad users is not possible. This is driven by the assumptions regarding the preferences of the users – bad users strictly value untraced transactions more than good users.*

In the spirit of mechanism design, the operator chooses the privacy design of the system to maximize a welfare criterion, taking into account the ex-post interaction between users and the enforcer. In particular, the model can be recast as a principal-agent problem, in which the operator acts as a principal to design a system with the aim of curtailing (unobservable) bad activity of agents as while enabling good users to benefit from transaction privacy. In doing so, we study the provision of "hard" privacy, taking as given practical challenges of fully preventing bad users from exploiting privacy features of a system.

Timeline. The timeline of events is as follows.

- $t = 0$: Operator sets identity-privacy rule $\mathcal{I}$ and transaction-privacy rule p.
- $t = 1$: User type is realized; the user submits transactions.
- $t = 2$: Privacy rules are applied; the enforcer looks for illicit transactions; payoffs and welfare are realized.

Solution Concept. We solve for subgame-perfect Nash equilibria $(\mathcal{I}^\star, p^\star, (N^\star, S^\star, Q^\star))$ in the operator–user game, where (i) $(N^\star, S^\star, Q^\star)$ is the bad user's best-response action to $(\mathcal{I}^\star, p^\star)$ and (ii) $(\mathcal{I}^\star, p^\star)$ maximizes $W(\mathcal{I}, p, (N^{BR}, S^{BR}, Q^{BR})(\mathcal{I}, p))$ where $(N^{BR}, S^{BR}, Q^{BR})(\mathcal{I}, p)$ is the bad user's best-response as a function of $(\mathcal{I}, p)$.

3 Equilibrium Payment System Design

We solve the model via backward induction. In the second stage, the user submits transactions within the payment system. If the user is good, then the transaction volume is randomly determined. If the user is bad, then they choose their action taking as given the payment system design and ex-post enforcement. In the first stage, the operator decides on the payment system design, taking into account the expected actions of users.

3.1 Bad Users' payoff frontier

Bad users have access to a set of technologies $[T]$, allowing bad actors to strategically adapt to the underlying privacy environment set by the operator. For any given payment system design chosen by the operator, a bad user strategically selects a subset of technologies, total transaction volume, and an allocation of transaction across technologies to maximize its payoff. We start by characterizing the *payoff frontier* of the bad user, defined by the maximum achievable payoff per total payment volume $q \in \mathbb{R}_+$.

First, observe that for a fixed total transaction volume q, the bad user must choose a subset $S \subseteq [T]$ of technologies to use and an allocation of transaction volumes Q_i across the set of technologies $i \in S$. For a fixed $S \subseteq [T]$, let $M_S(q)$ be the bad user's maximum achievable payoff conditional on the bad user's transactions being private at q with probability one. We can express $M_S(q)$ as the solution to the following optimization problem:

$$\begin{aligned} M_S(q) = \max_{(Q_i)_{i \in S}} & \sum_{i \in S} (\theta_i Q_i - \eta_i Q_i^2) \\ \text{s.t.} & \sum_{i \in S} Q_i = q \\ & Q_i \geq 0 \; \forall \; i \in S \end{aligned}$$

Intuitively, taking as given a set of technologies, the bad user makes an allocation Q_i across technologies that takes into account the marginal production and costs of each technology. We summarize useful properties of M_S in the following lemma:

If we use payoff frontier, should we still refer to production of a technology?

Lemma 1. *For any subset S of technologies, the payoff frontier M_S achievable by S is single-peaked in the total transaction volume q.*

A key implication of Lemma 1 is that M_S is single-peaked. Consequently, a bad user faces certain limits to the scale of their activity for any given subset of technologies. Next, consider the bad user's choice of technologies to use. For any target transaction volume $q \in \mathbb{R}_+$, the bad user selects a subset $S \subseteq [T]$ that yields the highest payoff at q over all possible subsets of T. We refer to the maximal payoff resulting from this choice as the *payoff frontier* $M(q)$, formally defined below.

Definition. *The bad user's payoff frontier $M : \mathbb{R}_+ \to \mathbb{R}$ is*

$$M(q) = \max_{S \subseteq [T]} \left(M_S(q) - \sum_{i \in S} \phi_i \right).$$

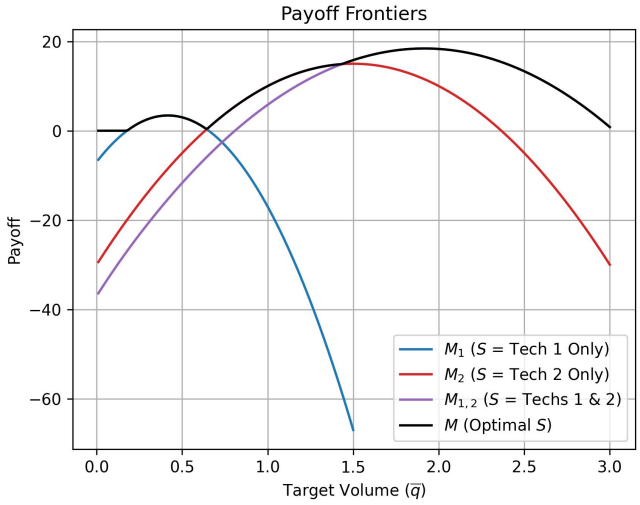

Figure 1: Payoff frontier (PF) for a bad user with two technologies: $\theta_1 = 50$, $\eta_1 = 60$, $\phi_1 = 7$, $\theta_2 = 60$, $\eta_2 = 20$, $\phi_2 = 30$. The red and blue lines are the PFs of technologies 1 and 2 only, respectively; the purple line is the PF from using both technologies in tandem; the black line is the overall PF given the technology set.

This payoff frontier succinctly captures the bad user's payoff attained by their best-responses given a candidate total transaction volume q. The key consideration for the choice of technologies is the fixed cost ϕ_i associated with each technology. For a high target q, the bad user may choose a set of technologies involving high fixed costs but scale better, i.e. peak at large transaction volumes; for a low target q, the bad user may choose a subset of technologies that economize on fixed costs.

We approach the operator's optimal design problem as follows. We first characterize the optimal TP rule taking as given the IP rule, stating properties of $p_1^\star$ and $p_0^\star$, the optimal TP rules when fixing $\mathcal{I} = 1$ and $\mathcal{I} = 0$, respectively. We then show that the payment system given by $(1, p_0^\star)$ yields a higher net welfare than that given by $(0, p_1^\star)$, establishing that the optimal design of the payment system is given by IR an TP rule $p_0^\star$, i.e. $(1, p_0^\star)$.

3.2 Identity-Private Payment Systems

In order to understand how IP interacts with TP, we start by restricting attention to payment systems with IP ($\mathcal{I} = 1$), in which the system does not require users to reveal identity when creating an account. This has two operational implications: it allows bad users to create multiple accounts on which it can allocate transactions without the system discerning com-

monality of ownership, and restricts the application of TP rules to the account level as the underlying mapping between users and accounts is unobservable to the system. We show for any nonzero TP rule, the bad user is able to arbitrarily scale operations within the restricted range of TP across accounts, resulting in prohibitively high welfare costs.

Lemma 2. *In an identity-private payment system ($\mathcal{I} = 1$), if the good user has some chance of benefitting from transaction-privacy (i.e. p such that $p(q) = 1$ for some $q \in [\underline{q}, \overline{q}]$), then the welfare cost can be arbitrarily large.*

Lemma 2 states that the bad user can exploit any provision of TP to scale their operations without limit, and only requires a minimal condition that the bad user's payoff frontier $M(q)$ is positive over the relevant range of $q \in [\underline{Q}, \overline{Q}]$.[21] As long as the bad users are able to achieve some positive payoff within the payment volume specified by the TP rule, it is able to linearly scale the payment value across multiple accounts. Furthermore, because all activity is conducted within the TP region, bad users face no risk of enforcement.

Before characterizing the optimal TP under IP, we demonstrate that the potential for IP to be exploited by bad users is a generic feature. Consider an IP system with no TP, in which case all transactions are subject to monitoring by the enforcer. The bad user's best response is given by the following lemma.

Lemma 3. *In an identity-private payment system ($\mathcal{I} = 1$) that is fully transaction-auditable ($p(q) = 0$ for all $q \in \mathbb{R}_+$), the bad user's optimal action is to equally distribution payments across arbitrarily many accounts (i.e. $N \to \infty$), which results in a limited welfare cost.*

Like the case when there is some TP, the bad user utilizes multiple accounts in their best response, but the underlying motives differ. In the presence of TP, the bad user uses multiple accounts to scale its operations after matching the transaction volume to meet the TP region per account. With no TP, the bad user targets an optimal payment volume and distributes activity across multiple accounts in order to minimize the total likelihood of detection by the enforcer, taking advantage of the fact that the enforcer's detection probability diminishes with lower per-account activity. Hence, although the bad user's operations are bounded (in contrast to the case with TP), the welfare cost associated with the bad user's activity can be disproportionately large relative to privacy value accruing to good users.

Notably, the bad user's optimal action in an IP system with no TP resembles two common forms of money laundering: "structuring" and "smurfing." Structuring refers to the practice by which smaller transactions are made in order to lower the likelihood of raising suspicion

[21] Note that this result does not depend on the number of technologies in T.

from monitoring. Smurfing refers to the practice of processing payments through multiple accounts to obfuscate the association between transactions. In the model, bad users combine both by processing very small payments through numerous accounts to minimize the probability of detection by the enforcer. These techniques are known to be used by criminals in public blockchains, such as Bitcoin and Ethereum (both IP systems with no TP).

Proposition 1. *For an identity-private payment system ($\mathcal{I} = 1$), any optimal transaction-privacy rule, denoted $p_1^\star$, does not provide transaction-privacy on $[\underline{q}, \bar{q}]$. Furthermore, the zero function (i.e. $p(q) = 0$ for all $q \in \mathbb{R}_+$) is an optimal transaction-privacy rule.*

Proposition 1 characterizes the equilibrium TP rule when the operator is constrained to designing and IP system, and show that a zero transaction-privacy rule is welfare-optimal. At a high level, these results indicate that identity privacy *and* transaction privacy are incompatible in practice. The interaction between IP and TP enable the bad user to effectively scale its operations within the specified privacy environment, through the strategic use of multiple accounts. Consequently, for an IP system, the operator choose not to provide any TP, ultimately prevents good users from benefiting from direct provision of TP.

3.3 Identity-Required Payment Systems

We now turn our attention to identity-required systems (i.e. $\mathcal{I} = 0$). In the absence of IP, only one account can be made by each user. As such, the TP rule can now be imposed at the user level. Let the transaction-privacy rule be p. Since the bad user's payoff from no privacy is zero (as $\beta_b = 1$), the bad user always chooses a $\mathbf{q}$ such that $p(\|\mathbf{q}\|_1) = 1$. For a total volume q, let $\mathbf{q}^\star(q) = \arg\max_{\mathbf{q}} M_{S^\star(q)}(q)$ where $S^\star(q) = \arg\max_{S \subseteq [T]} M_S(q)$. It follows that the optimal total transaction volume $q^\star$ for the bad user is

$$q^\star = \arg\max_q \{M(q) : M(q) > 0, p(q) = 1\}, \tag{6}$$

so the best response of the bad user is $\mathbf{q}^\star(q^\star)$. Given the best-response of the bad user, we focus on the operator's problem for setting the transaction-privacy rule.

Lemma 4. *In an identity-required payment system ($\mathcal{I} = 0$), out of all transaction-privacy rules in which the bad user's optimal payoff is u (where $u \in [0, \sup_q M(q)]$), there exists a unique transaction-privacy rule that maximizes net welfare.*

The operator aims to a set the TP rule that balances between the bad user's payoff against the privacy gains accruing to good users over the privacy region specified by the TP rule.

Lemma 4 establishes the existence and uniqueness of an optimal TP rule for any set of TP rules from which bad users attain some fixed payoff u. This transforms the infinite-dimensional optimization problem of finding the optimal $p : \mathbb{R}_+ \to [0, 1]$ to a single-dimensional optimization problem over bad user payoff levels u. We build on Lemma 4 to fully characterize the optimal TP rule.

Proposition 2. *In an identity-required payment system ($\mathcal{I} = 0$), an optimal transaction-privacy rule, denoted $p_0^\star$, can consist of disjoint intervals of payment volumes in which transaction-privacy is provided.*

For a payment system where identity is required, the operator maximizes net welfare by selectively offering TP across regions so as to maximize the provision of privacy of good users whilst targeting a fixed payoff level of the bad user. Proposition 2 shows that the dominant TP rule can be "complex" in the sense that the privacy regions consists of disjoint intervals over the transaction volume space. The possible non-monotonicity of M as established in Lemma 1 suggests that M can be potentially multi-peaked. Thus, for certain payoff levels, the level set of M at those payoff levels can be a union of disjoint intervals, which gives rise to a complex rule. The operator proceeds to search across payoff levels for the welfare-optimal level and selects the optimal transaction-privacy rule corresponding to this level. Figure 2 illustrates the connection between the bad user's payoff frontier and the optimal TP rule, also offering an example of how a "complex" rule can arise.

Intuitively, a complex rule arises when the operator targets "gaps" in the bad user's payoff frontier where technologies are collectively less effective and thus are not desirable for the bad user. In general, the TP rule targets a region of low-value payments, a region of large-value payments, and others depending on the bad user's payoff frontier. In the low payment volume region, good users' demand is high whilst the low transaction limit poses a ceiling on bad users' activity and fixed technological costs limit the utilization of some technologies. In the high payment volume region, the operator opts to provide transaction privacy when the the scalability of the bad users' technologies is expected to drop.

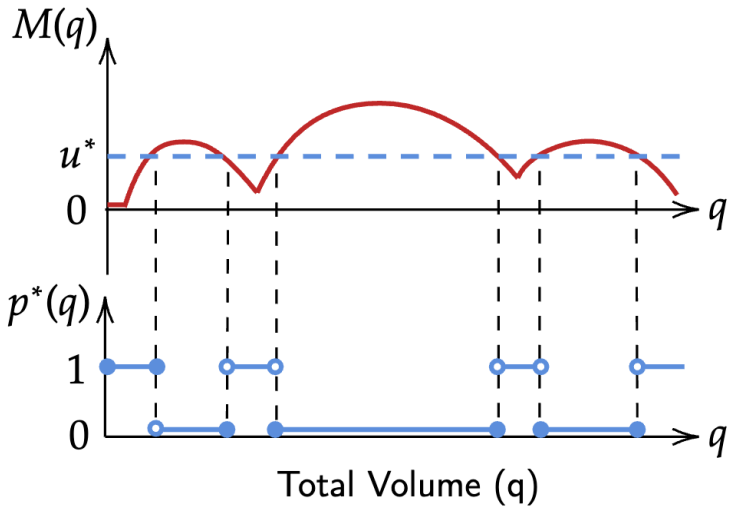

Figure 2: Illustration of a bad user's payoff frontier and the corresponding optimal TP rule. Once an optimal payoff level $u^\star$ is found, TP is provided to volumes yielding lower payoffs than $u^\star$. To ensure that the bad user selects the volume yielding $u^\star$ with the lowest welfare cost, only the first interval of privacy is closed.

3.4 Equilibrium

Having characterized the optimal TP rules $p_1^\star$ and $p_0^\star$ conditional on each type of IP rule ($\mathcal{I} = 1$ and $\mathcal{I} = 0$, respectively), it remains to compare net welfare under each combination (($\mathcal{I} = 0, p = p_1^\star$) and ($\mathcal{I} = 1, p = p_0^\star$)) to determine the payment system design chosen by the operator in equilibrium. We show that the equilibrium design entails a identity-required system with the selective TP rule as characterized by Proposition 2 below.

Theorem 3. *Let $p_1^\star$ be an optimal transaction-privacy rule for a identity-required payment system and $p_0^\star$ be the zero function. Then net welfare under privacy rules $\mathcal{I} = 0$ with $p = p_1^\star$ exceeds that under privacy rules $\mathcal{I} = 1$ with $p = p_0^\star$. The equilibrium design is thus an identity-required payment system with transaction-privacy rule $p_1^\star$.*

Theorem 3 states that the welfare-optimal IP system always yields a lower net social value than the welfare-optimal identity-required system. Thus, the equilibrium system requires identity but can offer TP at certain transaction volumes. As explained previously, the equilibrium TP rule is highly tailored to the bad user's technology set, assumed to be static, which can result in potentially "complex" forms as shown in Figure 2. In the next section, we consider the possibility for set of technologies to change.

4 Payment System Design Under Innovation

A major concern regarding the baseline model equilibrium is whether the regions where TP is provided can be implemented in a precise manner that sufficiently controls for current and future bad activity. In reality, illicit actors are known to constantly innovate around legal and enforcement boundaries to avoid suspicion. Furthermore, criminals are often found to be early adopters of new technologies; known examples include the use of dark or anonymous communication systems, such as encrypted messaging and privacy-centric devices, and more recently, blockchains.

We extend our baseline model to account for the possibility for bad users to engage in innovation by introducing two variations in which the bad user can enhance their technology set. The first extension captures *incremental innovation*, whereby the bad user can improve the production functions of each technology at some investment cost. This represents effort and capital investment to improve the efficiency of existing operations to maximize output. The second extension captures *radical innovation*, whereby the bad user can develop entirely new technologies to complement the existing set of technologies.

4.1 Incremental Innovation

Improvements in Productivity. We now suppose that the bad user can choose an *innovation level* $\alpha \geq 1$. At an innovation level of α, all production is scaled by α, at an extra cost of $\gamma(\alpha)$ incurred, where $\gamma : [1, \infty) \to \mathbb{R}_+$ is the *innovation cost function.*

Assumption 2. *Innovation cost is quadratic in α, i.e. $c(\alpha) = \gamma(\alpha - 1)^2$.*

Building on the baseline model, we modify the bad user's payoff frontier under incremental innovation. For a fixed $q \in \mathbb{R}_+$ and $S \subseteq [T]$, let $M_S^{inc}(q)$ to be the bad user's maximum achievable payoff by submitting transactions with a total volume of q given S and choosing an innovation level of α, conditional on the bad user's transactions being private with probability one. We can express $M_S^{inc}(q)$ as the solution of the following optimization problem:

$$\begin{aligned} M_S^{inc}(q) = \max_{(Q_i)_{i\in S}, \alpha} & \sum_{i\in S} (\alpha\theta_i Q_i - \eta_i Q_i^2) - \gamma(\alpha - 1)^2 \\ \text{s.t.} & \sum_{i\in S} Q_i = q \\ & Q_i \geq 0 \ \forall \ i \in S \\ & \alpha \geq 1. \end{aligned}$$

They key difference relative to the benchmark model is that the bad user now has an additional decision variable α which captures the level of incremental innovation. Incremental innovation shifts outputs by a factor of α at an innovation cost of $\gamma(\alpha - 1)^2$. Under Assumption 2, we provide the following characterization of M_S^{inc}:

Lemma 5. *If $\theta_i/\eta_i > 4\gamma$ for all $i \in [T]$, then M_S^{inc} is continuous and increasing in q for any $S \subseteq [T]$.*

With incremental innovation, we show a marked difference in the payoff frontier of the bad user. In contrast to the baseline model, M_S^{inc} is increasing if baseline technologies' productivity is sufficiently large relative to the cost of innovation. Through incremental innovation, the bad user is able to "fine-tune" the production characteristics of its technologies so as to achieve payoffs that scale with transaction volume. Similarly to the baseline model, for a given q, the bad user can choose between all subsets $S \subseteq [T]$ to determine which yields the optimal payoff, yielding the payoff frontier, defined below. The next theorem fully characterizes the equilibrium with incremental innovation.

Definition. *The bad user's payoff frontier under incremental innovation $M : \mathbb{R}_+ \to \mathbb{R}$ is*

$$M^{inc}(q) = \max_{S \subseteq [T]} \left(M_S^{inc}(q) - \sum_{i \in S} \phi_i \right).$$

Theorem 4. *Suppose that $\theta_i/\eta_i > 4\gamma$ for all $i \in [T]$. Under incremental innovation:*

- *in an identity-private system ($\mathcal{I} = 1$), the zero function is an optimal transaction-privacy rule;*
- *in a identity-required system ($\mathcal{I} = 0$), the optimal transaction-privacy rule, denoted $p_0^\star$, provides privacy when the total payment volume is less than a certain volume threshold;*
- *the equilibrium design under incremental innovation is an idnetity-required payment system with transaction-privacy rule $p_1^\star$.*

While the operator follows the same principles as before in determining the equilibrium TP rule, a striking result arises: when incremental innovation is sufficiently effective, then the optimal TP rule consists of a "simple" rule consisting of a single threshold. To see this, consider the case where the operator selects a complex rule, along the lines of that described in previous section. With incremental innovation, the bad user observes the TP rule, and can choose make investments into technologies to selectively high-transaction volume intervals with TP. This intuition is formalized in Lemma 5, which shows that the bad user's indirect utility when targeting a total volume of q now increases in q, which stands in stark contrast to the single-peakedness in the baseline case.

4.2 Radical Innovation

Creating New Technologies. We now suppose that the bad user can develop an entirely new technology to add to it technology set. To discipline the potential set of new technologies, we assume the following. Suppose that the bad user can select one technology τ from the chosen subset S to serve as the basis of creating new technologies. Specifically, upon choosing τ, the bad user selects a technology from the *technological cone* of τ, parameterized by $\alpha \geq 1$, along with the other technologies $S \backslash \{\tau\}$. We model the technological cone in such a way that imposes some bounds on innovation, with maximum productivity and fixed costs scaling linearly with respect to elements in the original technology set.

Technological Cones. The *cone of technology* τ is a set of technologies parameterized by $\alpha \geq 1$ such that technology $\tau(\alpha)$ has production and cost functions given by

$$y_{\tau(\alpha)}(q_\tau) = \theta_\tau q_\tau,$$
$$c_{\tau(\alpha)}(q_\tau) = \alpha^{-1}\eta_\tau \cdot q_\tau^2 + \alpha\phi_\tau,$$

respectively. Under this specification, the payoff function (production minus costs) of technology $\tau(\alpha)$ has a maximum value of α times that of $\tau(1)$, achieves the maximum at α times the arg max of the utility function of $\tau(1)$, and has a fixed cost of α times ϕ_τ.[22] An example of a technological cone is shown in Figure 3.

For a fixed $q \in \mathbb{R}_+$, $S \subseteq [T]$ and $\tau \in S$ (the technology that the bad user chooses to innovate on), let $M^{rad}_{S,\tau}(q)$ be the bad user's largest achievable payoff by submitting transactions with a total volume q using technologies from $S \backslash \{\tau\}$ and a technology from the technological cone of τ. We can then write:

$$M^{rad}_{S,\tau}(q) = \max_{(Q_i)_{i\in S},\alpha} \sum_{i \in S\backslash\{\tau\}} (\theta_i Q_i - \eta_i Q_i^2) + (\theta(\tau,\alpha)\cdot q_\tau - \eta(\tau,\alpha)\cdot q_\tau^2 - \alpha\phi_\tau) \tag{7}$$

$$\text{s.t.} \sum_{i\in S} Q_i = q \tag{8}$$

$$Q_i \geq 0 \;\forall\; i \in S \tag{9}$$

$$\alpha \geq 1. \tag{10}$$

Similar to incremental innovation, the bad user has an additional choice α capturing the production-cost features of the new technology. A larger α corresponds to a larger fixed cost

[22] Note that the utility function of $\tau(1)$ achieves a maximum value of $\theta_\tau^2/4\eta_\tau - \phi_\tau$ at $\theta_\tau/2\eta_\tau$.

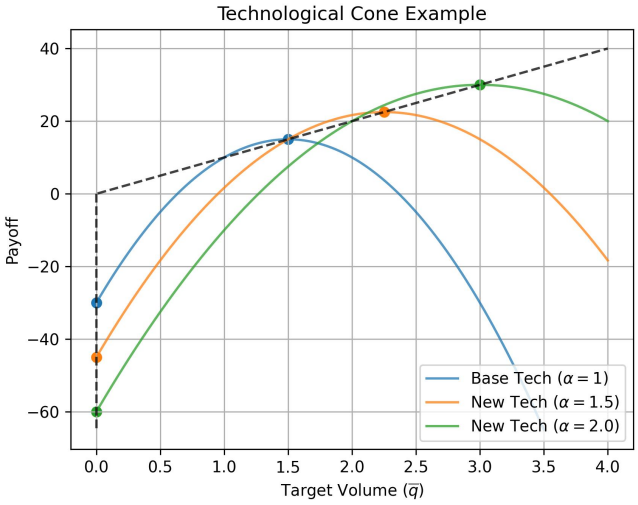

Figure 3: Elements of the technological cone of a technology with $\theta_\tau = 60$, $\eta_\tau = 20$, $\phi_\tau = 30$. The payoff frontier (output minus variable and fixed costs) of the baseline technology is in blue, with those of the elements of the technological cone corresponding to $\alpha = 1.5$ and $\alpha = 2$ in orange and green, respectively.

but better scalability for larger payments regarding the bad user's newly created technology. However, the bad user now has an additional choice of the technology τ from which to create a new technology. In the next lemma, we show that $M^{rad}_{S,\tau}$ is generally increasing in q.

Lemma 6. *$M^{rad}_{S,\tau}$ is continuous and increasing in q for any $S \subseteq [T]$ and $\tau \in S$.*

Radical innovation, like incremental innovation, substantially changes the profile of the payoff frontier of the bad user. One notable difference is that under radical innovation, the payoff frontier is increasing in general (without further assumptions on innovation costs). This difference reflects the core nature of radical innovation, which allows the bad user to commit regulatory arbitrage by surgically developing a new technology. We proceed to define the payoff frontier and characterize the equilibrium payment system design under radical innovation.

Definition. *The bad user's payoff frontier under radical innovation $M : \mathbb{R}_+ \to \mathbb{R}$ is*

$$M^{rad}(q) = \max_{S \subseteq [T], \tau \in S} \left(M^{rad}_{S,\tau}(q) - \sum_{i \in S \setminus \{\tau\}} \phi_i \right).$$

Theorem 5. *If the bad user can create a new technology from a technological cone, then*

- *in an identity-private system ($\mathcal{I} = 1$), the zero function is an optimal TP rule;*
- *in an identity-required system ($\mathcal{I} = 0$), the optimal transaction-privacy rule, denoted $p_1^\star$ provides privacy when the total payment volume is less than a certain volume threshold.*
- *the equilibrium design under radical innovation is an idnetity-required payment system with transaction-privacy rule $p_1^\star$.*

The transaction-privacy rules that arise in equilibria when the operator is constrained to different types of identity privacy have similar properties, notably, that the optimal TP rule is also a simple threshold policy when $\mathcal{I}$ is fixed to one. While the bad user is able to target larger volumes at a profit by appropriately balancing productivity improvements with investment costs in the model of incremental innovation, under radical innovation, the bad user targets larger volumes by creating a technology with the appropriate scalability by moving along a technological cone. Whether bad users improve upon existing technologies or create new ones entirely, simple threshold policies achieve robustness to innovation while allowing for good users to capture some benefits of transaction privacy.

5 Classification and Implementation of Payment Systems

So far, we used our model to characterize the optimal set of rules governing privacy in a payment system. In order to do so, we study an environment that considers a broad array of privacy configurations. As such, our framework nests a variety of existing payment systems that exist today, including permissionless blockchains, legacy payment systems, and privacy-centric cryptocurrencies. In this section, we use our model to analyze the relative welfare ranking between different existing systems, and study key determinants that affect the relative value of each type of system.

5.1 A Privacy-Based Classification of Payment Systems

	Identity-Private	**Identity-Required**
TX-Auditable	Public blockchain	Legacy payment systems, credit cards
(Partial) TX-Private	Monero, ZCash	Zero-knowledge protocols (e.g. SeDe),[23] white papers (e.g. digital Euro & Yuan)[24]

Table 1: Classification and examples of payment systems based on types of privacy (identity privacy and transaction privacy) provided to users.

The two dimensions of privacy considered in our model provides an intuitive quadrant-based classification of payment systems, as shown in Table 1. The horizontal axis is the identity-privacy axis and the vertical axis is the transaction-privacy axis. Starting from the lower left quadrant and going clockwise, we abbreviate the types of payment systems as *(IP, TP)*, *(IP, TA)*, *(IR, TA)*, and *(IR, TP)*. In this section, we use our model to rank these four types of systems based on net welfare.

First, we can build on the results in Section 3 and 4 to make several observations:

- net welfare in an (IP, TA) system is always lower-bounded by that in all other systems, since allowing some transaction-privacy can lead to an infinite welfare cost while the welfare benefit remains bounded;
- net welfare in an (IR, TA) system is always upper-bounded by that in an (IR, TP) system, since the zero transaction-privacy rule is a feasible solution to the operator's problem in designing the optimal transaction-privacy rule;
- net welfare in an (IP, TA) system is always upper-bounded than that in a (IR, TP) system, since the latter is the type of system that arises in equilibrium.

These comparisons highlight the different role that transaction privacy plays in payment systems based on their privacy treatment of identity. In an IP system, TP is harmful because the ability to create multiple accounts without providing identity can lead to bad actors taking advantage of it on a large scale. On the other hand, in a IR payment system, TP can be potentially beneficial to net welfare, since processing account activity at the *user* level allows the operator to control bad user activity while letting a good user reap the benefits of privacy.

It remains to compare the (IP, TA) and (IR, TA) systems, which we alias as "Public Blockchain" and "Legacy," respectively. In legacy systems, a good user receives no benefits

from transaction privacy and a bad user does not submit any transactions since the systems lacks identity and transaction privacy (cf. Section 3.3). In a public blockchain, the good user receives some benefits from transaction privacy and a bad user submits transactions with volumes totaling ζ^{-1}. The changes in welfare benefit and cost of an public blockchain relative to an legacy system are then given by

$$\Delta WB = \delta_g \int_{\min\{\underline{Q},\zeta^{-1}\}}^{\min\{\overline{Q},\zeta^{-1}\}} (1-\zeta q) q \, dF_g(q),$$

$$\Delta WC = \zeta^{-1}.$$

It follows that a legacy system is better in terms of net welfare than a public blockchain if and only if $\omega_g(1-\beta)\cdot\Delta WB > \omega_b\beta\cdot\Delta WC$. This may depend on several parameters, such as β, δ_g, and ζ. We first characterize the comparison with respect to the enforcement strength ζ.

Proposition 6. *Let "Public Blockchain" denote a (IP, no TP) system and "Legacy" denote a (IR, no TP) system. Suppose that smaller volumes of good user payment demand are more likely to be realized than larger volumes, i.e. F_g has density proportional to to $q^{-\kappa}$ where $\kappa > 1$. Then either:*

- *"Legacy" is socially preferable to "Public Blockchain" for all $\zeta \in \mathbb{R}_+$.*
- *There exists $0 < \zeta_L^\star < \zeta_H^\star$ such that "Public Blockchain" is socially preferable to "Legacy" if and only if $\zeta \in (\zeta_L^\star, \zeta_H^\star)$.*

When evaluating the net welfare of public blockchain and legacy systems in terms of enforcement strength, holding other parameters fixed, two possibilities arise. Either the legacy system is always socially preferred to the public blockchain, or there exists an interval $(\zeta_L^\star, \zeta_H^\star)$ of enforcement strength values in which the public blockchain is socially preferred to the legacy system. In the latter case, we find an interesting tradeoff as ζ varies. For $\zeta < \zeta_L^\star$, the enforcer is too weak, allowing illicit activity on the blockchain to go unnoticed with high probability. However, for $\zeta > \zeta_H^\star$, the enforcer is very strong, meaning that good users cannot reap the benefits of privacy on the blockchain as the enforcer can trace their transactions with high probability. This leaves a "Goldilocks zone" where enforcement is neither too weak nor strong in which public blockchains can result in higher net welfare to the operator than legacy systems. We illustrate this in Figure 4.

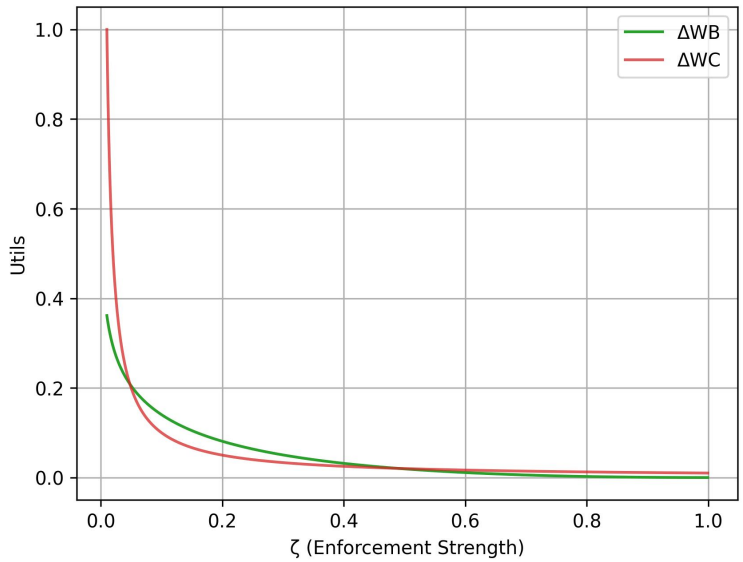

Figure 4: A sketch of the increases in welfare benefit and welfare cost, ΔWB and ΔWC, respectively, when viewed as a function of enforcement ζ.

We proceed with the comparison along two more parameters, δ_G and β, that have tangible real-world analogs. As enforcement strength varies between jurisdiction, so may the societal value on privacy (captured by δ_G) and prevalence of financial crime (captured by β) within a jurisdiction. Figure 5 shows the region of the δ_G-ζ space (left) and β-ζ space (right) where public blockchains result in more net welfare than legacy systems. As the value of privacy increases, the $(\zeta_L^\star, \zeta_H^\star)$ interval where public blockchains are preferred widens. Thus, in jurisdictions that place high value on privacy, there is more room for public blockchains to be adopted as payment systems compared to legacy systems. On the other hand, as the prevalence of financial crime in a jurisdiction increases, the $(\zeta_L^\star, \zeta_H^\star)$ narrows, suggesting that there is less room for public blockchain-based payment systems due to the pervasiveness of illicit activity.

5.2 Towards Implementation of Privacy-Enhanced Payment Systems

Our framework outlines key principles of the *economic* design of privacy-enhanced payment systems. The key properties, outlined in Theorems 4 and 5, offer tangible guidelines on the nature of privacy provision that could be implemented in various technical approaches. We now provide an overview of the tools and components that could facilitate implementation of privacy-enhancements, with a particular focus on state-contingent transaction privacy.

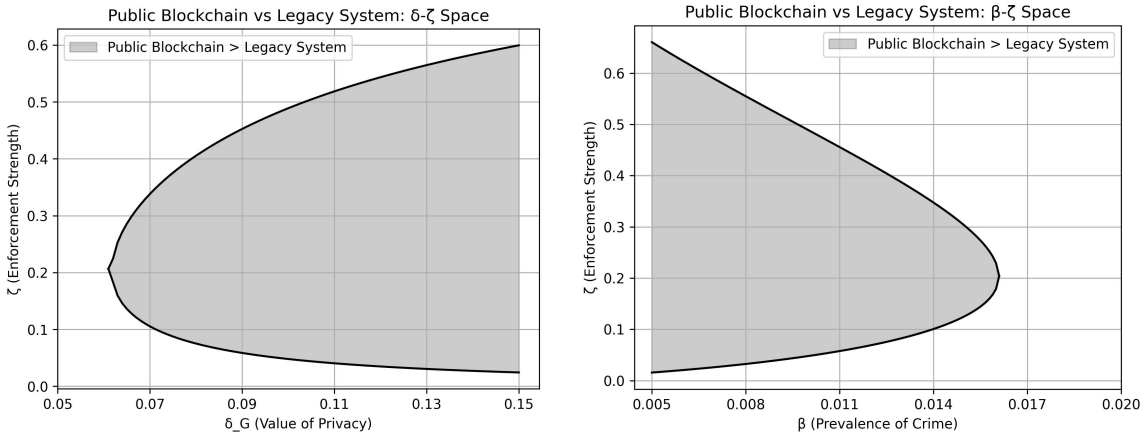

Figure 5: Shaded regions of the δ_G-ζ space (left) and β-ζ space (right) show where public blockchains are socially preferred to legacy systems. We set $\beta = 0.01$, $\delta_G = 0.1$, $\kappa = 2$, $\underline{Q} = 1$, and $\overline{Q} = 100$.

From a technical standpoint, privacy-provision is predicated on the ability for a system to distributed selective access to information across various actors of the settlement process. For IP, relaxing the permissions relating to account creation suffices. For TP, more advanced privacy-enhancements are used, such as zero-knowledge proofs (ZKPs), which are cryptographic protocols that enable one party to prove the validity of a statement (such as a transaction's validity) to another party without revealing underlying details. In the context of payments, ZKPs enhance transaction privacy by allowing verification without disclosing sensitive information.

A notable application are *zk-SNARKs (zero knowledge succinct non-interactive arguments of knowledge)*, a specific type of ZKP. These are used to enhance the privacy features for systems such as Monero or ZCash, both representing existing public blockchain systems with IP and TP. A notable drawback of this specific application of zk-SNARKs is the incompatibility with compliance functionalities that may be desirable for broader adoption: these systems allow for large-scale abuse by illicit actors and do not comply with legal regulations, bringing attention to the need for protocols that balance privacy provision with regulatory concerns.

Recent technical approaches aim to balance transaction privacy with with compliance standards. For example, the *Privacy Pools* protocol allows users themselves to provide zk-SNARKs to demonstrate their transactions are not illicit, typically achieved by proving membership in some federation of good users. Cryptographic tools can be combined with governance-centric measures to produce protocols that can flexibly implement the unhashing of encrypted data

and activity. Sahu, Gajera and Chaudhary (2023b) present a framework of *involuntary selective de-anonymization* in which transactions of interest can be ex-post decrypted to view the addresses and/or identities involved. In their framework, decisions to de-anonymize are made by a preselected quorum upon receiving a de-anonymization request. This framework can be readily applied to implement a (IR, some TP) system in our model by (i) including real-world identity in the encryption data, (ii) designating nodes of a permissioned or federated blockchain as the decision-making quorum, and (iii) automating de-anonymization requests based on the transaction-privacy rule.

6 Conclusion

We provide a foundational framework for designing privacy-enhanced payment systems that balance user privacy, enforcement efficacy, and societal welfare. By modeling how good and bad users interact with the payment system under different privacy settings, we show how the optimal design of such systems balances the strategic choices of illicit actors and the societal costs of their activities with the benefits of privacy to good users.

Our equilibrium analysis reveals that any provision of transaction privacy in an identity-private system is suboptimal. Using sophisticated technologies to mimic payment behaviors of good users, bad users can exploit transaction privacy to scale illicit activities across multiple accounts without detection, and resulting in prohibitive social costs. identity-required systems with carefully designed transaction-privacy rules can achieve superior outcomes. By tailoring transaction-privacy to disjoint ranges of payment volumes, the operator can maximize privacy benefits for good users while limiting the scope of bad user activity.

Our analysis has important policy implications, for central banks and private payment providers alike. The robust design rule given by no identity-privacy and a simple threshold on transaction size effectively caps the maximum extractable privacy value by bad users, ensuring resilience against future technological advancements or operational shifts by criminal actors. The implementation of such a system requires detailed knowledge of bad users' technologies and their operational boundaries, emphasizing the importance of strong enforcement and technological oversight. Our findings also highlight that existing designs of CDBC systems, which typically rely on identity-private systems with account balance limits, may fail to curtail high-throughput illicit transactions.

References

Acharya, Viral, Robert Engle, and Matthew Richardson, "Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks," *American Economic Review*, 2012, *102* (3), 59–64.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman, "The economics of privacy," *Journal of economic Literature*, 2016, *54* (2), 442–492.

Ahnert, Toni, Peter Hoffmann, and Cyril Monnet, "The digital economy, privacy, and CBDC," 2022.

Alnasaa, Marwa, Nikolay Gueorguiev, Jiro Honda, Eslem Imamoglu, Paolo Mauro, Keyra Primus, and Dmitriy Rozhkov, "Crypto-assets, corruption, and capital controls: Cross-country correlations," *Economics Letters*, 2022, *215*, 110492.

Auer, Raphael, Rainer Böhme, Jeremy Clark, and Didem Demirag, "Privacy-enhancing technologies for digital payments: mapping the landscape," 2025.

Becker, Gary, "5." Crime and Punishment: An Economic Approach."," *Journal of Political Economy*, 1968, *76* (2), 169–217.

Biais, Bruno, J.C. Rochet, and S. Villeneuve, "Do cryptocurrencies matter?," *Working Paper*, 2024.

Brownworth, Anders, Jon Durfee, Michael Junho Lee, and Antoine Martin, "Regulating Decentralized Systems: Evidence from Sanctions on Tornado Cash," 2024.

Buterin, Vitalik, Jacob Illum, Matthias Nadler, Fabian Schär, and Ameen Soleimani, "Blockchain privacy and regulatory compliance: Towards a practical equilibrium," *Blockchain: Research and Applications*, 2024, *5* (1), 100176.

Calvo-Armengol, Antoni and Yves Zenou, "Social networks and crime decisions: The role of social structure in facilitating delinquent behavior," *International Economic Review*, 2004, (45), 939–958.

Capponi, Agostino, Ruizhe Jia, and Kan Ye Wang, "Maximal Extractable Value and Allocative Inefficiencies in Public Blockchains," 2024.

Chainalysis, "Organized Crime Shows High Level of Professionalization, Low Level of Crypto Sophistication," *Technical Report*, 2025.

Chen, Jiakai and Asani Sarkar, "Slowed-Down Capital: Using Bitcoin to Avoid Capital Controls," 2022.

Chiu, Jonathan and Thorsten V Koeppl, "Blockchain-based settlement for asset trading," *The Review of Financial Studies*, 2019, *32* (5), 1716–1753.

Coelho, R, J Fishman, and D Ocampo, "Supervising cryptoassets for anti-money laundering," 2021.

Cong, Lin William, R. Harvey Campbell, Daniel Rabetti, and Zong-Yu Wu, "An Anatomy of Crypto-Enabled Cybercrimes," *NBER Working Paper*, 2023, (30834).

Cong, Lin William, Campbell Harvey, Daniel Rabetti, and Zong-Yu Wu, "AN ANATOMY OF CRYPTO-ENABLED CYBERCRIME," *NBER Working Paper Series*, 2023, (30834).

Cong, Lin William, Xi Li, Ke Tang, and Yang Yang, "Crypto wash trading," *Management Science*, 2023, *69* (11), 6427–6454.

Cong, Lin William, Ye Li, and Neng Wang, "Tokenomics: Dynamic Adoption and Valuation," *Review of Financial Studies*, 2021, *34* (3), 1105–1155.

Cong, Will and Simon Mayer, "The Coming Battle of Digital Currencies," *Working Paper*, 2024.

Daian, P., S. Goldfeder, T. Kell, Y. Li, X. Zhao, I. Bentov, L. Breidenbach, and A. Juels, "Flash Boys 2.0: Frontrunning in Decentralized Exchanges, Miner Extractable Value, and Consensus Instability," in "2020 IEEE Symposium on Security and Privacy (SP)" 2020, pp. 910–927.

Department of Homeland Security, "Combatting Illicit Activity Utilizing Financial Technologies and Cryptocurrencies," 2022.

Erol, Selman and Michael Lee, "Insider networks," *FRB of New York Staff Report*, 2019, (862).

European Central Bank, "Exploring anonymity in central bank digital currencies," *In Focus*, 2019, *4*.

Europol, "The Other Side of the Coin: An Analysis of Financial and Economic Crime," 2023.

Foley, Sean, Jonathan R Karlsen, and Talis Putnins, "Utility tokens as a commitment to competition," *Review of Financial Studies*, 2019, *32* (5), 1798–1853.

Garratt, Rodney and Michael Lee, "Monetizing privacy with central bank digital currencies," *Available at SSRN 3583949*, 2021.

Garratt, Rodney J and Maarten RC Van Oordt, "Privacy as a public good: a case for electronic cash," *Journal of Political Economy*, 2021, *129* (7), 2157–2180.

Ghosh, Pulak, Boris Vallee, and Yao Zeng, "FinTech lending and cashless payments," in "Proceedings of Paris December 2021 Finance Meeting EUROFIDAI-ESSEC" 2022.

Goldstein, Itay, Deeksha Gupta, and Ruslan Sverchkov, "Utility tokens as a commitment to competition," *The Journal of Finance*, 2023.

Graf von Luckner, Clemens, Carmen M. Reinhart, and Kenneth Rogoff, "Decrypting new age international capital flows," *Journal of Monetary Economics*, 2023, *138*, 104–122.

He, Zhiguo, Jing Huang, and Jidong Zhou, "Open banking: Credit market competition when borrowers own the data," *Journal of financial economics*, 2023, *147* (2), 449–474.

Hu, Maggie, Adrian D Lee, and Tālis J Putniņš, "Evading capital controls via cryptocurrencies: evidence from the blockchain," *Available at SSRN 3956933*, 2021.

Iorio, Alberto Di, Anneke Kosse, and Ilaria Mattei, "Embracing diversity, advancing together-results of the 2023 BIS survey on central bank digital currencies and crypto," 2024.

Kahn, Charles M, James McAndrews, and William Roberds, "Money is privacy," *International Economic Review*, 2005, *46* (2), 377–399.

Lee, Michael and Rodney Garratt, "Monetizing Privacy with Central Bank Digital Currencies," *FRB of New York Staff Report*, 2021, (958).

____ , Antoine Martin, and Robert M Townsend, "Optimal design of tokenized markets," *FRB of New York Staff Report*, 2024, (1121).

____ , ____ , and ____ , "Zero Settlement Risk Token Systems," 2024.

Meling, Tom G, Magne Mogstad, and Arnstein Vestre, "Crypto Tax Evasion," Technical Report, National Bureau of Economic Research 2024.

Monnet, Cyril, Jon R Frost, Leonardo Gambacorta, Raphael Auer, and Tara Rice, "Central Bank Digital Currencies: Motives, Economic Implications, and the Research Frontier," *Annual Review of Economics*, 2022, *14* (1), 697–721.

Nakamoto, Satoshi, "Bitcoin: A peer-to-peer electronic cash system," *Satoshi Nakamoto*, 2008.

Parlour, Christine A, Uday Rajan, and Haoxiang Zhu, "When fintech competes for payment flows," *The Review of Financial Studies*, 2022, *35* (11), 4985–5024.

Phua, Kenny, Bo Sang, Chishen Wei, and Gloria Yang Yu, "Don't trust, verify: The economics of scams in initial coin offerings," *Available at SSRN*, 2022, *4064453.*

Pocher, Nadia and Adnreas Veneris, "Privacy and Transparency in CBDCs: A Regulation-by-Design AML/CFT Scheme," *IEEE Transactions on Network and Service Management*, 2022, *19* (2), 1776–1788.

Rochet, Jean-Charles and Jean Tirole, "Controlling risk in payment systems," *Journal of Money, Credit and Banking*, 1996, *28* (4), 832–862.

____ **and** ____ , "Externalities and Regulation in Card Payment Systems," *Review of Network Economics*, 2006, *5* (1).

Rogoff, K., *The curse of cash*, Princeton University Press, 2016.

Saberhagen, Nicolas Van, "CryptoNote v 2.0," 2013.

Sahu, Naveen, Mitul Gajera, Amit Chaudhary, and Hamish Ivey-Law, "SeDe: Balancing Blockchain Privacy and Regulatory Compliance By Selective De-Anonymization," *arXiv Preprint 2311.08167*, 2023.

____ **,** ____ **, and** ____ , "zkFi: Privacy-Preserving and Regulation Compliant Transactions using Zero Knowledge Proofs," *arXiv Preprint 2307.00521*, 2023.

Sasson, Eli Ben, Alessandro Chiesa, Christina Garman, Matthew Green, Ian Miers, Eran Tromer, and Madars Virza, "Zerocash: Decentralized anonymous payments from bitcoin," in "2014 IEEE symposium on security and privacy" IEEE 2014, pp. 459–474.

Sockin, Michael and Wei Xiong, "Decentralization through Tokenization," *Journal of Finance*, 2023, *78* (1), 247–299.

Townsend, Robert M, *Distributed Ledgers: Design and Regulation of Financial Infrastructure and Payment Systems*, MIT Press, 2020.

Treasury, "2024 National Risk Assessments for Money Laundering, Terrorist Financing, and Proliferation Financing," 2024.

TRM Labs, "Beyond Fentanyl: Cryptocurrency's Role in the International Drug Precursor Market," 2024.

Uhligh, Harald, Mike Alonso, and Jon Frost, "Privacy in Digital Payments—Escaping the Panopticon," *Georgetown Journal of International Affairs*, 2023, *24* (2).

Villaverde, Fernandez, D. Sanches, Linda Schilling, and H. Uhlig, "Central bank digital currency: Central banking for all?," *Review of Economic Dynamics*, 2021, *41*, 225—-242.

____ , Linda Schilling, and H. Uhlig, "Central bank digital currency: When price and bank stability collide," *Working Paper*, 2020.

von Luckner, Clemens Graf, Carmen M Reinhart, and Kenneth Rogoff, "Decrypting new age international capital flows," *Journal of Monetary Economics*, 2023, *138*, 104–122.

Wang, Zijian, "Money laundering and the privacy design of central bank digital currency," *Review of Economic Dynamics*, 2023, *51*, 604–632.

White House, "Executive Order on Ensuring Responsible Development of Digital Assets," 2022.

Whited, Toni, Yufeng Wu, and Kairong Xiao, "Will Central Bank Digital Currency Disintermediate Banks?," *Working Paper*, 2022.

A Proofs

A.1 Proof of Lemma 1

Fix a target volume q and technology subset S. The bad user then faces the problem

$$M_S(q) = \max_{(Q_i)_{i\in S}} \sum_{i\in S}(\alpha \cdot y_i(Q_i) - c_i(Q_i)) - \gamma(\alpha)$$
$$\text{s.t.} \sum_{i\in S} Q_i = q$$
$$Q_i \geq 0 \; \forall \; i \in S.$$

Introduce a Lagrange multiplier λ for the equality constraint and multipliers $(\mu_i)_{i\in S}$ for the inequality constraints. The KKT conditions are then:

- Primal and dual feasibility: $\sum_{i\in S} Q_i = q$, $Q_i \geq 0 \; \forall \; i \in S$, and $\mu_i \geq 0 \; \forall \; i \in S$.
- Complementary slackness: $\mu_i Q_i = 0 \; \forall \; i \in S$.
- Stationary point: $y_i'(Q_i) - c_i'(Q_i) = \lambda - \mu_i \; \forall \; i \in S$.

Define the following:

$$f(\mathbf{q}, q) = \sum_{i\in S}(y_i(Q_i) - c_i(Q_i)),$$
$$g(\mathbf{q}, q) = \sum_{i\in S} Q_i - q.$$

Let $\mathbf{q}^\star(q)$ be the vector of payment volumes for total payment volume q. Then we can write

$$M_S(q) = f(\mathbf{q}^\star(q), q).$$

Taking derivatives with respect to q yields

$$\frac{dM_S}{dq} = \frac{\partial f}{\partial q} + \sum_{i\in S} \frac{\partial f}{\partial Q_i} \cdot \frac{dQ_i^\star}{dq}. \tag{11}$$

Note that the equality constraint must satisfy

$$g(\mathbf{q}^\star(q), q) = 0.$$

Taking derivatives with respect to q yields

$$\frac{\partial g}{\partial q} + \sum_{i \in S} \frac{\partial g}{\partial Q_i} \cdot \frac{dQ_i^\star}{dq} = 0 \tag{12}$$

Multiplying (2) by λ and adding the resulting expression to (1) yields

$$\frac{dM_S}{dq} = \frac{\partial f}{\partial q} + \lambda \cdot \frac{\partial g}{\partial q} + \sum_{i \in S} \left(\frac{\partial f}{\partial Q_i} + \lambda \cdot \frac{\partial g}{\partial Q_i} \right) \frac{dQ_i^\star}{dq}.$$

Substituting our definitions for f and g yields

$$\frac{dM_S}{dq} = -\lambda^\star + \sum_{i \in S} (y_i'(Q_i^\star) - c_i'(Q_i^\star) + \lambda^\star) \cdot \frac{dQ_i^\star}{dq}.$$

Suppose that we have a general solution. By KKT stationarity,

$$\begin{aligned} \frac{dM_S}{dq} &= -\lambda^\star + \sum_{i \in S} (\lambda^\star - \mu_i^\star + \lambda^\star) \cdot \frac{dQ_i^\star}{dq} \\ &= -\lambda^\star + 2\lambda^\star \sum_{i \in S} \frac{dQ_i^\star}{dq} - \sum_{i \in S} \mu_i \cdot \frac{dQ_i^\star}{dq} = \lambda^\star - \sum_{i \in S} \mu_i^\star \cdot \frac{dQ_i^\star}{dq}. \end{aligned}$$

We have two cases:

- For all nonbinding constraints j, complementary slackness implies $\mu_j^\star = 0$.
- For all binding constraints k, we have $Q_k^\star = 0$. By differentiating the complementary slackness condition with respect to q, we have

$$0 = \mu_k^\star \cdot \frac{\partial Q_k^\star}{\partial q} + Q_k^\star \cdot \frac{\partial \mu_k^\star}{\partial q} = \mu_k^\star \cdot \frac{\partial Q_k^\star}{\partial q}.$$

It thus follows that

$$\frac{dM_S}{dq} = \lambda^\star - \sum_{i \in S} \mu_i^\star \cdot \frac{dQ_i^\star}{dq} = \lambda^\star.$$

We then have the following:

- As the correspondence $\phi(q) = \{\mathbf{q} \in \mathbb{R}_+^n : \mathbf{1}^\top \mathbf{q} \leq q\}$ is continuous on $q \in \mathbb{R}_{++}$, by the Theorem of the Maximum, M is continuous on $\mathbb{R}_{++}$.

- Writing $M_S(q) = f(q^\star(q), q)$ and differentiating yields

$$\begin{aligned}\frac{d^2 M_S}{dq^2} &= \sum_i \left(f_{Q_i}(q^\star(q)) \cdot \frac{d^2 Q_i^\star}{dq^2} + f_{Q_i Q_i}(q^\star(q)) \cdot \left(\frac{dQ_i^\star}{dq^2} \right)^2 \right) \\ &= \sum_i \left((y_i'(Q_i^\star) - c_i'(Q_i^\star)) \cdot \frac{d^2 Q_i^\star}{dq^2} + (y_i''(Q_i^\star) - c_i''(Q_i^\star)) \left(\frac{dQ_i^\star}{dq^2} \right)^2 \right)\end{aligned}$$

  By the stationarity condition, $y_i'(Q_i^\star) - c_i'(Q_i^\star) = \lambda^\star - \mu_i^\star$. If $Q_i^\star = 0$, then $d^2 Q_i^\star / dq^2 = 0$ (except if q corresponds to a point where Q_i is exactly binding, but these points comprise a set of measure zero); if $Q_i^\star > 0$, then $\mu_i^\star = 0$, so $y_i'(Q_i^\star) - c_i'(Q_i^\star) = \lambda^\star$. Then we can write

$$\frac{d^2 M_S}{dq^2} = \lambda^\star \sum_i \frac{d^2 Q_i^\star}{dq^2} + \sum_i (y_i''(Q_i^\star) - c_i''(Q_i^\star)) \left(\frac{dQ_i^\star}{dq^2} \right)^2.$$

  Note that twice-differentiating the budget constraint with respect to q yields

$$\sum_i \frac{d^2 Q_i^\star}{dq^2} = 0$$

  so it follows that

$$\frac{d^2 M_S}{dq^2} = \sum_i (y_i''(Q_i^\star) - c_i''(Q_i^\star)) \left(\frac{dQ_i^\star}{dq^2} \right)^2.$$

  Since the y_i's are concave and the c_i's are convex, we have $d^2 M_S / dq^2 < 0$, as desired.

- Let $\underline{\lambda} = \min_i (y_i'(0) - c_i'(0)) > 0$ and $\underline{q}_i$ be the unique value of q such that $y_i'(q) - c_i'(q) = \underline{\lambda}$ for each $i \in S$. By construction, $\underline{q}_i \geq 0 \ \forall \ i \in S$ with all but one inequality being strict. Let $\tilde{q} = \sum_i \underline{q}_i$. We claim that M is increasing over $[0, \tilde{q}]$. For $q \in [0, \tilde{q}]$, there must exist some technology j where $q_j^\star > 0$. By complementary slackness, we have $\mu_j^\star = 0$, so $y_j'(q_j^\star) - c_j'(q_j^\star) = \lambda^\star$. Then $y_i'(Q_i^\star) - c_i'(Q_i^\star) \leq \lambda^\star \ \forall \ i \in S$. Suppose, towards a contradiction, that $\lambda^\star \leq 0$. By construction, it must be that $q_1^\star + q_n^\star > \tilde{q}$, which is a contradiction. Then we must have $\lambda^\star > 0$, so M is increasing on some interval $[0, \tilde{q}]$.

- Since $\lim_{Q_i \to \infty} (y_i(Q_i) - c_i(Q_i)) \to -\infty \ \forall \ i \in [T]$, it follows that for some $\tilde{q}$ large enough, we have $M(\tilde{q}) < 0$. Since M is continuous, $M(0) = 0$ and $M'(0) > 0$, it follows that there exists $q \in \mathbb{R}_+$ such that $M'(q) < 0$. Since M is strictly concave, it follows that

$\lim_{q\to\infty} M'(q) < 0$.

A.2 Proof of Lemma 2

By Assumption 1, there exists some technology i such that $\overline{Q} < \theta_i/\eta_i$. This implies that for any $q \in [\underline{Q}, \overline{Q}]$, we have $q \leq \overline{Q} < \theta_i/\eta_i$, so $\theta_i q - \eta_i q^2 > 0$ for any $q \in [\underline{Q}, \overline{Q}]$. If there exists $q \in [\underline{Q}, \overline{Q}]$ such that $p(q) = 1$, then the bad user can submit a volume of q solely through technology i on each of N accounts and take $N \to \infty$. Since $p(q) = 0$, the bad user's payoff is

$$\lim_{N\to\infty} N(\theta_i q - \eta_i q^2) = \infty$$

so this is an optimal action. The resulting welfare cost is $\lim_{N\to\infty} Nq = \infty$.

A.3 Proof of Lemma 3

Lemma 7. *Fix $N \in \mathbb{N}$ and $\bar{q} < N\zeta^{-1}$. The optimal solution to*

$$\max_{(q_a)_{a\in[N]}} \prod_{a\in[N]} (1 - \pi(q_a))$$

$$s.t. \sum_{a\in[N]} q_a = \bar{q}$$

is $q_a = \bar{q}/N$ for all $a \in [N]$.

Proof. Since $\bar{q} < N\zeta^{-1}$, an optimal solution must satisfy $(q_a)^\star < \zeta$ for all $a \in [N]$, so we can write the objective as

$$\prod_{a\in[N]} (1 - \zeta q_a)$$

We derive an upper bound on the objective subject to the constraint. Note that the function $f(x) = \log(1 - \zeta x)$ is concave for $x < \zeta^{-1}$. By Jensen's Inequality and the constraint,

$$\frac{1}{N}\sum_{a\in[N]} \log(1 - \zeta q_a) = \frac{1}{N}\sum_{a\in[N]} f(q_a) \leq f\left(\frac{1}{N}\sum_{a\in[N]} q_a\right) = f\left(\frac{\bar{q}}{N}\right) = \log\left(1 - \frac{\zeta q}{N}\right).$$

Multiplying both sides of the inequality by N and exponentiating yields

$$\prod_{a\in[N]}(1-\zeta q_a) \leq \left(1-\frac{\zeta q}{N}\right)^N.$$

This upper bound is achieved by setting $q_a = \bar{q}/N$ for all $a \in [N]$. □

Lemma 8. *Fix $N \in \mathbb{N}$, $S \subseteq [T]$ and $\bar{q} < N\zeta^{-1}$. There exists an optimal solution $\mathbf{q}^\star$ to*

$$\begin{aligned}\max_{(Q_{a,i})_{a\in[N],i\in S}} \prod_{a\in[N]}\left(1-\pi\left(\sum_{i\in S}Q_{a,i}\right)\right)\sum_{a\in[N],i\in S}\theta_i Q_{a,i} - \sum_{a\in[N],i\in S}\eta_i(Q_{a,i})^2 - \sum_{i\in S}\phi_i \\ s.t. \sum_{a\in[N],i\in S} Q_i = \bar{q} \\ Q_{a,i} \geq 0 \;\forall\, a\in[N],\; i\in S\end{aligned}$$

such that $(Q_i^1)^\star = \cdots = (Q_i^N)^\star$ for each $i \in S$.

Proof. Denote the objective by u:

$$u((Q_{a,i})_{a\in[N],i\in S}) = \prod_{a\in[N]}\left(1-\pi\left(\sum_{i\in S}Q_{a,i}\right)\right)\sum_{a\in[N],i\in S}\theta_i Q_{a,i} - \sum_{a\in[N],i\in S}\eta_i(Q_{a,i})^2 - \sum_{i\in S}\phi_i.$$

Let $(Q_{a,i})_{a\in[N],i\in S}$ be a feasible solution and $q_a = \sum_{i\in S} Q_{a,i}$. Let $\psi = \prod_{a\in[N]}(1-\pi(q_a))$ and

$$\begin{aligned}\tilde{M}_S(q) = \max_{(Q_i)_{i\in S}} \sum_{i\in S}(\tilde{\theta}_i Q_i - \eta_i Q_i^2) \\ \text{s.t.} \sum_{i\in S} Q_i = q \\ Q_i \geq 0 \;\forall\, i\in S\end{aligned}$$

where $\tilde{\theta}_i = \psi\theta_i$. For each $a \in [N]$, let $(\tilde{Q}_{a,i})_{i\in S}$ be the argument that attains $\tilde{M}_S(q_a)$. Then

$$u((Q_{a,i})_{a\in[N],i\in S}) \leq u((\tilde{Q}_{a,i})_{a\in[N],i\in S})$$

since the probability of detection remains the same while output minus costs increases. For each $a \in [N]$, let $(\bar{Q}_{a,i})_{i\in S}$ be the argument that attains $\tilde{M}_S(\bar{q}/N)$. By Lemma 1, $\tilde{M}_S$ is strictly

concave in q. Then by Jensen's Inequality and Lemma 7,

$$u((\tilde{Q}_{a,i})_{a\in[N],i\in S}) = \sum_{a\in[N]} \tilde{M}_S(q_a) - \sum_{i\in S}\phi_i \leq \sum_{a\in[N]} \tilde{M}_S(\bar{q}/N) - \sum_{i\in S}\phi_i.$$

Let $\bar{\psi} = (1-\zeta\bar{q}/N)^N$ and

$$\begin{aligned}\bar{M}_S(q) = \max_{(Q_i)_{i\in S}} \ & \sum_{i\in S}(\bar{\theta}_i Q_i - \eta_i Q_i^2) \\ \text{s.t.} \ & \sum_{i\in S} Q_i = q \\ & Q_i \geq 0 \ \forall \ i \in S.\end{aligned}$$

where $\bar{\theta}_i = \bar{\psi}\theta_i$. By Lemma 7, we have $\bar{\psi} \geq \psi$, so $(\bar{Q}_{a,i})_{i\in S}$ is a feasible solution to the problem for $\bar{M}_S(\bar{q}/N)$ with objective value greater than $\tilde{M}_S(\bar{q}/N)$. Let $((Q_{a,i})^\star)_{i\in S}$ be the argument that attains $\bar{M}_S(\bar{q}/N)$. Then

$$\sum_{a\in[N]} \tilde{M}_S(\bar{q}/N) - \sum_{i\in S}\phi_i \leq \sum_{a\in[N]} \bar{M}_S(\bar{q}/N) - \sum_{i\in S}\phi_i = u(((Q_{a,i})^\star)_{a\in[N],i\in S}).$$

Since the inequalities hold for any feasible solution, it follows that $((Q_{a,i})^\star)_{a\in[N],i\in S}$ is optimal. By construction, $((Q_{a,i})^\star)_{a\in[N],i\in S}$ is symmetric, i.e. $(Q_i^1)^\star = \cdots = (Q_i^N)^\star$ for each $i \in S$. □

Fix $S \subseteq [T]$ and $q < N\zeta^{-1}$. By Lemma 8, there exists a symmetric optimal solution for the bad user's optimization problem under zero transaction-privacy, so for N accounts, the total volume for each account is q/N. We can then write the objective as

$$\begin{aligned}\max_{(Q_i)_{i\in S},N} \ & N\left(\left(1-\frac{\zeta q}{N}\right)^N \sum_{i\in S}\theta_i Q_i - \sum_{i\in S}\eta_i Q_i^2\right) - \sum_{i\in S}\phi_i \\ \text{s.t.} \ & \sum_{i\in S} Q_i = \frac{q}{N} \\ & Q_i \geq 0 \ \forall \ i \in S.\end{aligned}$$

Let $\theta_{(1)} = \max_{i\in S}\theta_i$. We first provide an upper bound on the objective:

$$N\left(\left(1-\frac{\zeta q}{N}\right)^N \sum_{i\in S}\theta_i Q_i - \sum_{i\in S}\eta_i Q_i^2\right) - \sum_{i\in S}\phi_i \leq N\left(1-\frac{\zeta q}{N}\right)^N \theta_{(1)}\sum_{i\in S} Q_i - \sum_{i\in S}\phi_i$$

$$\leq \left(1 - \frac{\zeta q}{N}\right)^N \theta_{(1)} q - \sum_{i \in S} \phi_i \leq e^{-\zeta q} \theta_{(1)} q - \sum_{i \in S} \phi_i.$$

We can achieve this upper bound by creating N accounts, submitting q/N through technology $\arg\max_{i \in S} \theta_i$, and taking $N \to \infty$:

$$\begin{aligned} u(S, q) &= \lim_{N \to \infty} N \left(\left(1 - \frac{\zeta q}{N}\right)^N \theta_{(1)} \frac{q}{N} - \eta_{(1)} \frac{q^2}{N^2} \right) - \sum_{i \in S} \phi_i \\ &= \lim_{N \to \infty} \left(\left(1 - \frac{\zeta}{q}\right)^N \theta_{(1)} q - \frac{\eta_{(1)} q^2}{N} \right) - \sum_i \phi_i = e^{-\zeta q} \theta_{(1)} q - \sum_{i \in S} \phi_i. \end{aligned}$$

Maximizing the attainable upper bound across q yields $q^\star = \zeta^{-1}$. The resulting utility is

$$u(S) = e^{-\zeta q^\star} \theta_{(1)} q^\star - \sum_{i \in S} \phi_i = \frac{\theta_{(1)}}{e\zeta} - \sum_{i \in S} \phi_i$$

Optimizing over all $S \subseteq \mathcal{S}$ yields

$$u^\star = \max_{i \in \mathcal{S}} \left(\frac{\theta_i}{e\zeta} - \phi_i \right).$$

By Assumption 1, there exists at least one technology i such that $\theta_i > e\zeta\phi_i$, so the user receives a positive payoff from this action. Thus, the bad user's best response is specified by equally distributing payments across $N \to \infty$ accounts, while the welfare cost is $\zeta^{-1} < \infty$.

A.4 Proof of Proposition 1

Suppose, towards a contradiction, that p is optimal and provides transaction privacy on some $q \in [\underline{Q}, \overline{Q}]$. By Lemma 2, p induces an infinite welfare cost and by Lemma 3, there exist transaction-privacy rules with bounded welfare cost. Thus p cannot be optimal.

To show that the zero function is an optimal transaction-privacy rule, consider any other transaction-privacy rule $\tilde{p}$. Since $\tilde{p}$ is not the zero function, there must exist $q \in \mathbb{R}_+$ such that $\tilde{p}(q) = 1$. If there exists $S \subseteq [T]$ such that $M_S(q) > 0$, then the bad user can create N accounts, submit a total volume of q through each account using the combination of technologies and amounts that attain $M_s(q)$, and take $N \to \infty$ to induce an infinite welfare cost for the operator. Otherwise, if there does not exist such an $S \subseteq [T]$, then welfare cost remains the same. By Lemma 2, we cannot have $q \in [\underline{Q}, \overline{Q}]$ in this case, so the welfare benefit remains the same as well. Thus the zero function remains optimal.

A.5 Proof of Lemma 4

Fix $u \in \mathbb{R}_+$ and let $\underline{q}(u) = \inf\{q \in \mathbb{R}_+ : M(q) = u\}$. We claim that the optimal transaction-privacy rule such that the bad user's best-response to the rule achieves a payoff of u, denoted $p^\star_u$, is given by

$$p^\star_u(q) = \mathbb{1}(\{M(q) < u^\star\} \cup \{WC(q) \leq WC(\underline{q}(u))\}).$$

Let p be any transaction-privacy rule such that the bad user's best-response achieves a payoff of u. Note that $\{p(q) = 1\} \subseteq \{M(q) \leq u\}$ by construction. By Lemma 1, it follows that $\{M(q) \leq u\} = \{M(q) < u\} \cup \mathcal{N}$, where $\mathcal{N}$ is a set of Lebesgue measure zero. The welfare cost is $WC(\underline{q}(u))$ under $p^\star_u$ and is lower-bounded by $WC(\underline{q}(u))$ under any p, since u must be attainable under p and $\underline{q}(u)$, by construction, is the total transaction volume gives the bad user a payoff of u with the lowest welfare cost, since WC is increasing in q. Since

$$\{p(q) = 1\} \subseteq \{M(q) \leq u\} \subseteq \{M(q) < u\} \cup \mathcal{N} \cup \{WC(q) \leq WC(\underline{q}(u))\} = \{p^\star_u(q) = 1\},$$

where $\mathcal{N}$ is a Lebesgue null set, it follows that the welfare benefit under $p^\star_u$ is lower-bounded by the welfare benefit under any other p, with set equality (outside of $\mathcal{N}$) if and only if $p = p^\star_u$. The result follows.

A.6 Proof of Proposition 2

Any transaction-privacy rule p can be classified by the payoff u that is achieved under p. Since Lemma 4 asserts that for any achievable payoff level u, there is an optimal transaction-privacy rule, the operator's problem of choosing the welfare-optimal transaction-privacy rule reduces to choosing a payoff level for the bad user resulting in the greatest welfare under $p^\star_u$. The operator now solves

$$u^\star = \sup_{u \in [\underline{Q}, \overline{Q}]} (1-\beta)\mathbb{E}[(1 - \delta_b \cdot \mathbb{1}\{p^\star_u(q) = 1\})q_g] - \beta \cdot WC(\underline{q}(u)).$$

It follows that the equilibrium transaction-privacy rule in a payment system where identity is required ($\mathcal{I} = 0$) is $p^\star_u$, which in general, is a union of disjoint intervals since for each $S \subseteq [T]$, M_S is strictly concave, increases initially, and then decreases.

A.7 Proof of Theorem 3

By Proposition 1, the zero function $p_1^\star$ is an optimal transaction-privacy rule when $\mathcal{I} = 1$. Under $\mathcal{I} = 1$ and $p_1^\star$, the welfare benefit is $\int_0^{\zeta^{-1}} (1 - \delta_g \zeta q) q \, dF_g(q)$ and the welfare cost is ζ^{-1}. Under $\mathcal{I} = 0$ and p given by $p(q) = \mathbb{1}\{q \leq \zeta^{-1}\}$, the welfare benefit is $\int_0^{\zeta^{-1}} q \, dF_g(q)$ and the welfare cost is at most ζ^{-1}. Since the welfare benefit under $(1, p)$ is greater than that under $(0, p_1^\star)$, and the welfare cost under $(1, p)$ is upper-bounded by that under $(0, p_1^\star)$, it follows that $W(0, p_1^\star) < W(1, p)$. Let $p_1^\star$ be the optimal transaction-privacy rule when $\mathcal{I} = 0$. Then $W(1, p) < W(1, p_0^\star)$. The result follows.

A.8 Proof of Lemma 5

Fix a target volume q and technology subset S. The bad user then faces the problem

$$\begin{aligned} M_S^{inc}(q) = \max_{(Q_i)_{i\in S}} & \sum_{i\in S}(\alpha \cdot y_i(Q_i) - c_i(Q_i)) - \gamma(\alpha) \\ \text{s.t.} & \sum_{i\in S} Q_i = q \\ & Q_i \geq 0 \; \forall \; i \in S \\ & \alpha \geq 1. \end{aligned}$$

Introduce a Lagrange multiplier λ for the equality constraint, multipliers $(\mu_i)_{i\in S}$ for the Q_i inequality constraints, and a multiplier ν for the α inequality constraint. The KKT conditions are then:

- Primal and dual feasibility: $\sum_{i\in S} Q_i = q$, $Q_i \geq 0 \; \forall \; i \in S$, $\alpha \geq 1$, $\mu_i \geq 0 \; \forall \; i \in S$, and $\nu \geq 0$.
- Complementary slackness: $\mu_i Q_i = 0 \; \forall \; i \in S$ and $\nu(\alpha - 1) = 0$.
- Stationary point:

$$\begin{aligned} \alpha \cdot y_i'(Q_i) - c_i'(Q_i) &= \lambda - \mu_i \; \forall \; i \in S \\ \sum_{i\in S} y_i(Q_i) - \gamma'(\alpha) &= -\nu. \end{aligned}$$

Define the following:

$$\begin{aligned} f(\mathbf{q}, q) &= \sum_{i\in S}(\alpha \cdot y_i(Q_i) - c_i(Q_i)), \\ g(\mathbf{q}, q) &= \sum_{i\in S} Q_i - q. \end{aligned}$$

Let $\mathbf{q}^\star(q)$ be the vector of payment volumes for total payment volume q. Then we can write

$$M_S(q) = f(\mathbf{q}^\star(q), q).$$

Taking derivatives with respect to q yields

$$\frac{dM_S}{dq} = \frac{\partial f}{\partial q} + \sum_{i\in S} \frac{\partial f}{\partial Q_i} \cdot \frac{dQ_i^\star}{dq} + \frac{\partial f}{\partial \alpha} \cdot \frac{d\alpha^\star}{dq}. \tag{13}$$

Note that the equality constraint must satisfy

$$g(\mathbf{q}^\star(q), q) = 0.$$

Taking derivatives with respect to q yields

$$\frac{\partial g}{\partial q} + \sum_{i \in S} \frac{\partial g}{\partial Q_i} \cdot \frac{dQ_i^\star}{dq} + \frac{\partial g}{\partial \alpha} \cdot \frac{d\alpha^\star}{dq} = 0 \tag{14}$$

Multiplying (2) by λ and adding the resulting expression to (1) yields

$$\frac{dM_S}{dq} = \frac{\partial f}{\partial q} + \lambda \cdot \frac{\partial g}{\partial q} + \sum_{i \in S} \left(\frac{\partial f}{\partial Q_i} + \lambda \cdot \frac{\partial g}{\partial Q_i} \right) \frac{dQ_i^\star}{dq} + \left(\frac{\partial f}{\partial \alpha} + \lambda \cdot \frac{\partial g}{\partial \alpha} \right) \frac{d\alpha^\star}{dq}.$$

Substituting our definitions for f and g yields

$$\frac{dM_S}{dq} = -\lambda^\star + \sum_{i \in S} (\alpha^\star \cdot y_i'(Q_i^\star) - c_i'(Q_i^\star) + \lambda^\star) \cdot \frac{dQ_i^\star}{dq} + \left(\sum_{i \in S} y_i(Q_i^\star) - \gamma'(\alpha^\star) \right) \frac{d\alpha^\star}{dq}.$$

Suppose that we have a general solution. By KKT stationarity,

$$\begin{aligned} \frac{dM_S}{dq} &= -\lambda^\star + \sum_{i \in S} (\lambda^\star - \mu_i^\star + \lambda^\star) \cdot \frac{dQ_i^\star}{dq} - \nu^\star \cdot \frac{d\alpha^\star}{dq} \\ &= -\lambda^\star + 2\lambda^\star \sum_{i \in S} \frac{dQ_i^\star}{dq} - \sum_{i \in S} \mu_i \cdot \frac{dQ_i^\star}{dq} - \nu^\star \cdot \frac{d\alpha^\star}{dq} = \lambda^\star - \sum_{i \in S} \mu_i^\star \cdot \frac{dQ_i^\star}{dq} - \nu^\star \cdot \frac{d\alpha^\star}{dq}. \end{aligned}$$

We have two cases:

- For all nonbinding constraints j, complementary slackness implies $\mu_j^\star = 0$ or $\nu^\star = 0$.
- For all binding constraints k, we have $Q_k^\star = 0$ or $\alpha^\star = 1$. Then

$$\begin{aligned} 0 &= \mu_k^\star \cdot \frac{\partial Q_k^\star}{\partial q} + Q_k^\star \cdot \frac{\partial \mu_k^\star}{\partial q} = \mu_k^\star \cdot \frac{\partial Q_k^\star}{\partial q}, \\ 0 &= \nu^\star \cdot \frac{d\alpha^\star}{dq} + (\alpha^\star - 1) \frac{\partial \nu^\star}{dq} = \nu^\star \cdot \frac{d\alpha^\star}{dq}. \end{aligned}$$

It thus follows that

$$\frac{dM_S}{dq} = \lambda^\star - \sum_{i \in S} \mu_i^\star \cdot \frac{dQ_i^\star}{dq} - \nu^\star \cdot \frac{d\alpha^\star}{dq} = \lambda^\star.$$

We claim that the constraint for α never binds for $q > 0$. Suppose, towards a contradiction, that $\alpha^\star = 1$ at some fixed total volume q and technology set S. Let $\mathbf{q}^\star$ denote the optimal transaction volumes given q and S. Then the stationarity condition for α is

$$\sum_{i \in S} y_i(Q_i^\star) - \gamma'(1) = \sum_{i \in S} y_i(Q_i^\star) > 0$$

which is a contradiction, since dual feasibility requires $\nu^\star$ to be nonnegative. Then, under our assumptions, the stationarity conditions for all $i \in S$ such that $Q_i^\star > 0$ and for α are

$$\alpha^\star \theta_i - 2\eta_i Q_i^\star = \lambda^\star$$
$$\sum_{i \in S} \theta_i Q_i^\star - 2\gamma(\alpha^\star - 1) = 0.$$

Note that that if $q_j^\star = 0$, then technology j does not contribute to the total output $\sum_{i \in S} y_i(Q_i^\star)$. Let $Y^\star = \sum_{i \in S} \theta_i Q_i^\star$. We solve the second condition to get

$$\alpha^\star = 1 + \frac{Y^\star}{2\gamma}.$$

Multiply each first condition by $\theta_i / 2\eta_i$ on both sides and sum them to get

$$\alpha^\star \sum_{i : Q_i^\star > 0} \frac{\theta_i^2}{\eta_i} - 2Y^\star = \lambda^\star \sum_{i : Q_i^\star > 0} \frac{\theta_i}{\eta_i}.$$

Letting $R_1 = \sum_{i : Q_i^\star > 0} \theta_i^2 / \eta_i$, $R_2 = \sum_{i : Q_i^\star > 0} \theta_i / \eta_i$, and plugging the expression for $\alpha^\star$, we have

$$(R_2 - 4\gamma) Y^\star = 4\gamma R_1 \lambda^\star - 4\gamma R_2.$$

We solve for $Y^\star$ in terms of $\lambda^\star$ and vice versa to get

$$Y^\star = \frac{4\gamma R_1}{R_2 - 4\gamma} \lambda^\star - \frac{4\gamma R_2}{R_2 - 4\gamma}$$
$$\lambda^\star = \frac{R_2 - 4\gamma}{4\gamma R_1} Y^\star + \frac{R_2}{R_1}.$$

Taking derivatives with respect to q yields

$$\frac{dY^\star}{dq} = \frac{4\gamma R_1}{R_2 - 4\gamma} \cdot \frac{d\lambda^\star}{dq}.$$

Let $R_0 = \sum_{i:Q_i^\star>0} 1/\eta_i$. Divide each first-order condition by η_i on both sides and sum them:

$$\alpha^\star R_1 - 2q = \lambda^\star R_0.$$

Plugging in the expressions for $\alpha^\star$ and $\lambda^\star$ yields

$$R_1 + \frac{R_1}{2\gamma} Y^\star - 2q = \frac{R_0(R_2 - 4\gamma)}{4\gamma R_1} Y^\star + \frac{R_0 R_2}{R_1}.$$

Solving for $Y^\star$ yields

$$\frac{2R_1^2 - R_0(R_2 - 4\gamma)}{4\gamma R_1} Y^\star = 2q + \frac{R_0 R_2 - R_1^2}{R_1}$$

Note that $R_0 R_2 - R_1^2 \geq 0$ by the Cauchy–Schwarz inequality. Since $Y^\star > 0$ for any $q > 0$ and the RHS is positive, it follows that

$$\frac{2R_1^2 - R_0(R_2 - 4\gamma)}{4\gamma R_1} > 0$$

so it follows that $dY^\star/dq > 0$ and consequently, $d\alpha^\star/dq > 0$. Then the actual output $Y = \alpha^\star Y^\star$ is also increasing in q. Suppose that $R_2 - 4\gamma > 0$. Since $dY^\star/dq > 0$, we must have $d\lambda^\star/dq > 0$. Conversely, suppose that $R_2 - 4\gamma < 0$. Since $dY^\star/dq > 0$, we must have $d\lambda^\star/dq < 0$. When $R_2 = 4\gamma$, then $d\lambda^\star/dq = 0$ and the indirect utility function is linear in q.

A.9 Proof of Theorem 4

Fix $S \subseteq [T]$ and $q < N\zeta^{-1}$. By Lemma 8, there exists a symmetric optimal solution for the bad user's optimization problem under zero transaction-privacy, so for N accounts, the total volume for each account is q/N. We can then write the objective as

$$\begin{aligned}
\max_{(Q_i)_{i\in S}, \alpha, N} \; & N\left(\left(1 - \frac{\zeta q}{N}\right)^N \sum_{i\in S} \alpha\theta_i Q_i - \sum_{i\in S} \eta_i Q_i^2\right) - \gamma(\alpha - 1)^2 - \sum_{i\in S} \phi_i \\
\text{s.t.} \; & \sum_{i\in S} Q_i = \frac{q}{N} \\
& Q_i \geq 0 \; \forall \, i \in S \\
& \alpha \geq 1.
\end{aligned}$$

Let $\theta_{(1)} = \max_{i \in S} \theta_i$. We first provide an upper bound on the objective:

$$\begin{aligned}
N \left(\left(1 - \frac{\zeta q}{N}\right)^N \sum_{i \in S} \theta_i Q_i - \sum_{i \in S} \eta_i Q_i^2 \right) - \gamma(\alpha - 1)^2 - \sum_{i \in S} \phi_i \\
&\leq N \left(1 - \frac{\zeta q}{N}\right)^N \alpha \theta_{(1)} \sum_{i \in S} Q_i - \gamma(\alpha - 1)^2 - \sum_{i \in S} \phi_i \\
&\leq \left(1 - \frac{\zeta q}{N}\right)^N \alpha \theta_{(1)} q - \gamma(\alpha - 1)^2 - \sum_{i \in S} \phi_i \\
&\leq e^{-\zeta q} \alpha \theta_{(1)} q - \gamma(\alpha - 1)^2 - \sum_{i \in S} \phi_i.
\end{aligned}$$

We can achieve this upper bound by creating N accounts, submitting q/N through technology $\arg\max_{i \in S} \theta_i$, and taking $N \to \infty$:

$$\begin{aligned}
u(S, q) &= \lim_{N \to \infty} N \left(\left(1 - \frac{\zeta q}{N}\right)^N \alpha \theta_{(1)} \frac{q}{N} - \eta_{(1)} \frac{q^2}{N^2} \right) - \gamma(\alpha - 1)^2 - \sum_{i \in S} \phi_i \\
&= \lim_{N \to \infty} \left(\left(1 - \frac{\zeta}{q}\right)^N \alpha \theta_{(1)} q - \frac{\eta_{(1)} q^2}{N} \right) - \gamma(\alpha - 1)^2 - \sum_{i \in S} \phi_i \\
&= e^{-\zeta q} \alpha \theta_{(1)} q - \gamma(\alpha - 1)^2 - \sum_{i \in S} \phi_i.
\end{aligned}$$

Maximizing the attainable upper bound across q and α yields

$$q^\star = \zeta^{-1}$$

$$\alpha^\star = 1 + \frac{\theta_{(1)}}{2e\zeta\gamma}$$

The resulting utility is

$$u(S) = e^{-\zeta q^\star} \alpha^\star \theta_{(1)} q^\star - \sum_{i \in S} \phi_i - \gamma(\alpha^\star - 1)^2 = \frac{\theta_{(1)}}{e\zeta} + \frac{\theta_{(1)}^2}{4e^2\zeta^2\gamma} - \sum_{i \in S} \phi_i$$

Optimizing over all $S \subseteq \mathcal{S}$ yields

$$u^\star = \max_{i \in \mathcal{S}} \left(\frac{\theta_i}{e\zeta} + \frac{\theta_i^2}{4e^2\zeta^2\gamma} - \phi_i \right).$$

By Assumption 1, $u^\star > 0$, so the bad user's best-response in an identity-private system with no transaction privacy results in a bounded welfare cost. Since the baseline model corresponds to $\alpha = 1$, the bad user can attain at least the baseline model's payoffs if transaction privacy is provided at some $q \in \mathbb{R}_+$. Using an argument similar to that in the proof of Proposition 1, the first result follows.

For the second result, we use an argument similar to that in the proof of Proposition 2. By Lemma 5, M_S^{inc} is increasing for any $S \subseteq [T]$, so $M^{inc}(q) = \max_{S\subseteq[T]} M_S^{inc}(q)$ is increasing as well. Since WC is also increasing in q, the form of the welfare-optimal contract given a payoff level u in Lemma 4 reduces to a single interval from 0 to a threshold $\bar{q}$.

Under $\mathcal{I} = 1$ and $p_1^\star$, the welfare benefit is $\int_0^{\zeta^{-1}} (1 - \delta_g \zeta q) q \, dF_g(q)$ and the welfare cost is ζ^{-1}. Under $\mathcal{I} = 0$ and p given by $p(q) = \mathbb{1}\{q \leq \zeta^{-1}\}$, the welfare benefit is $\int_0^{\zeta^{-1}} q \, dF_g(q)$ and the welfare cost is at most ζ^{-1}. The third result then follows from an argument similar to that in the proof of Theorem 3.

A.10 Proof of Lemma 6

Fix a target volume q and technology subset S. The bad user then faces the problem

$$\begin{aligned} M_{S,\tau}(q) = \max_{(Q_i)_{i\in S}} & \sum_{i\in S\setminus\{\tau\}} (y_i(Q_i) - c_i(Q_i)) + (y_{\tau(\alpha)}(q_\tau) - c_{\tau(\alpha)}(q_\tau)) \\ \text{s.t.} & \sum_{i\in S} Q_i = q \\ & Q_i \geq 0 \; \forall \; i \in S \\ & \alpha \geq 1. \end{aligned}$$

Introduce a Lagrange multiplier λ for the equality constraint, multipliers $(\mu_i)_{i\in S}$ for the Q_i inequality constraints, and a multiplier ν for the α inequality constraint. The KKT conditions are then:

- Primal and dual feasibility: $\sum_{i\in S} Q_i = q$, $Q_i \geq 0 \; \forall \; i \in S$, $\alpha \geq 1$, $\mu_i \geq 0 \; \forall \; i \in S$, and $\nu \geq 0$.
- Complementary slackness: $\mu_i Q_i = 0 \; \forall \; i \in S$ and $\nu(\alpha - 1) = 0$.
- Stationary point:

$$\begin{aligned} \alpha \cdot y_i'(Q_i) - c_i'(Q_i) &= \lambda - \mu_i \; \forall \; i \in S\setminus\{\tau\} \\ \theta'(\alpha) \cdot q_\tau - \eta'(\alpha) \cdot q_\tau^2 &= \phi - \nu. \end{aligned}$$

Define the following:

$$f(\mathbf{q}, q) = \sum_{i \in S \setminus \{\tau\}} (y_i(Q_i) - c_i(Q_i)) + (y_{\tau(\alpha)}(q_\tau) - c_{\tau(\alpha)}(q_\tau)),$$

$$g(\mathbf{q}, q) = \sum_{i \in S} Q_i - q.$$

Let $\mathbf{q}^\star(q)$ be the vector of payment volumes for total payment volume q. Then we can write

$$M_S(q) = f(\mathbf{q}^\star(q), q).$$

Taking derivatives with respect to q yields

$$\frac{dM_S}{dq} = \frac{\partial f}{\partial q} + \sum_{i \in S} \frac{\partial f}{\partial Q_i} \cdot \frac{dQ_i^\star}{dq} + \frac{\partial f}{\partial \alpha} \cdot \frac{d\alpha^\star}{dq}. \tag{15}$$

Note that the equality constraint must satisfy

$$g(\mathbf{q}^\star(q), q) = 0.$$

Taking derivatives with respect to q yields

$$\frac{\partial g}{\partial q} + \sum_{i \in S} \frac{\partial g}{\partial Q_i} \cdot \frac{dQ_i^\star}{dq} + \frac{\partial g}{\partial \alpha} \cdot \frac{d\alpha^\star}{dq} = 0 \tag{16}$$

Multiplying (2) by λ and adding the resulting expression to (1) yields

$$\frac{dM_S}{dq} = \frac{\partial f}{\partial q} + \lambda \cdot \frac{\partial g}{\partial q} + \sum_{i \in S} \left(\frac{\partial f}{\partial Q_i} + \lambda \cdot \frac{\partial g}{\partial Q_i} \right) \frac{dQ_i^\star}{dq} + \left(\frac{\partial f}{\partial \alpha} + \lambda \cdot \frac{\partial g}{\partial \alpha} \right) \frac{d\alpha^\star}{dq}.$$

Substituting our definitions for f and g yields

$$\frac{dM_S}{dq} = -\lambda^\star + \sum_{i \in S} (\alpha^\star \cdot y_i'(Q_i^\star) - c_i'(Q_i^\star) + \lambda^\star) \cdot \frac{dQ_i^\star}{dq} + (\theta'(\alpha^\star) \cdot Q_i^\star - \eta'(\alpha^\star)(Q_i^\star)^2 - \phi) \frac{d\alpha^\star}{dq}.$$

Suppose that we have a general solution. By KKT stationarity,

$$\frac{dM_S}{dq} = -\lambda^\star + \sum_{i \in S} (\lambda^\star - \mu_i^\star + \lambda^\star) \cdot \frac{dQ_i^\star}{dq} - \nu^\star \cdot \frac{d\alpha^\star}{dq}$$

$$= -\lambda^\star + 2\lambda^\star \sum_{i \in S} \frac{dQ_i^\star}{dq} - \sum_{i \in S} \mu_i \cdot \frac{dQ_i^\star}{dq} - \nu^\star \cdot \frac{d\alpha^\star}{dq} = \lambda^\star - \sum_{i \in S} \mu_i^\star \cdot \frac{dQ_i^\star}{dq} - \nu^\star \cdot \frac{d\alpha^\star}{dq}.$$

We have two cases:

- For all nonbinding constraints j, complementary slackness implies $\mu_j^\star = 0$ or $\nu^\star = 0$.
- For all binding constraints k, we have $Q_k^\star = 0$ or $\alpha^\star = 1$. Then

$$0 = \mu_k^\star \cdot \frac{\partial Q_k^\star}{\partial q} + Q_k^\star \cdot \frac{\partial \mu_k^\star}{\partial q} = \mu_k^\star \cdot \frac{\partial Q_k^\star}{\partial q},$$
$$0 = \nu^\star \cdot \frac{d\alpha^\star}{dq} + (\alpha^\star - 1)\frac{\partial \nu^\star}{dq} = \nu^\star \cdot \frac{d\alpha^\star}{dq}.$$

It thus follows that

$$\frac{dM_S}{dq} = \lambda^\star - \sum_{i \in S} \mu_i^\star \cdot \frac{dQ_i^\star}{dq} - \nu^\star \cdot \frac{d\alpha^\star}{dq} = \lambda^\star.$$

Suppose that $\alpha^\star = 1$. Then we are in the baseline case where $\lambda^\star \geq 0$ initially. Now suppose that $\alpha^\star > 1$. The stationarity condition with respect to α is $\eta(\tau, \alpha) \cdot (q_\tau/\alpha)^2 = \phi_\tau$, so we must have $q_\tau > 0$. The stationarity conditions are then

$$\theta_i - 2\eta_i Q_i = \lambda - \mu_i$$
$$\theta_\tau - 2\frac{\eta_\tau}{\alpha} q_\tau = \lambda$$
$$\frac{\eta_\tau}{\alpha^2} q_\tau^2 = \phi_\tau.$$

Solving this system of equations yields

$$Q_i^\star = \frac{(\theta_i - \theta_\tau) + 2\sqrt{\eta_\tau \phi} + \mu_i}{2\eta_i}$$
$$q_\tau^\star = \bar{q} - \sum_i Q_i^\star$$
$$\alpha^\star = \sqrt{\frac{\eta_\tau}{\phi_\tau}} \left(\bar{q} - \sum_i Q_i^\star \right)$$

By complementary slackness, if $Q_i^\star > 0$, then $\mu_i^\star = 0$. Then, if $(\theta_i - \theta_\tau) + 2\sqrt{\eta_\tau \phi_\tau} \geq 0$, then we cannot have $\mu_i > 0$. Similarly, if $(\theta_i - \theta_\tau) + 2\sqrt{\tilde{\eta}_\tau \phi_\tau} < 0$, then we must have $\mu_i = -((\theta_i -$

$\theta_\tau) + 2\sqrt{\eta_\tau \phi_\tau})$. Thus, whether or not $Q_i^\star = 0$ depends entirely on exogenous parameters, so each $Q_i^\star$ is constant. From solving the system of equations, we have that $\lambda^\star = \theta_\tau - 2\sqrt{\eta_\tau \phi_\tau}$, which is equivalent to $\theta_\tau^2/4\eta_\tau - \phi_\tau \geq 0$. This is true since the quantity is simply the maximum output minus cost attainable by technology τ, which we assume to be positive (otherwise, it could just be discarded from the technology set). Thus, the bad user's payoff is increasing in $\bar{q}$ as well.

We also need to show that the constraint on α becomes non-binding when the baseline M_S is still increasing in q. Let $\tilde{q}$ denote when the constraint on α becomes non-binding. Note that we must have $q_\tau^\star > 0$ at $\tilde{q}$, otherwise $\tilde{\eta}_\tau q_\tau^2 = 0 < \phi_\tau$ implies $\nu^\star > 0$. Then

$$q_\tau^\star = \frac{\tilde{\theta}_\tau - \lambda^\star}{2\tilde{\eta}_\tau}$$

so the constraint on α is binding if and only if

$$\tilde{\eta}_\tau q_\tau^\star = \frac{(\tilde{\theta}_\tau - \lambda^\star)^2}{4\tilde{\eta}_\tau} \leq \phi_\tau$$

Solving for $\lambda^\star$ yields

$$\lambda^\star \geq \theta_\tau - 2\sqrt{\tilde{\eta}_\tau \phi_\tau},\ \lambda^\star \leq \theta_\tau + 2\sqrt{\tilde{\eta}_\tau \phi_\tau}.$$

We know that $\theta_\tau - 2\sqrt{\tilde{\eta}_\tau \phi_\tau} \geq 0$, so the constraint on α is non-binding only at points where M_S is increasing with $\alpha^\star = 1$.

A.11 Proof of Theorem 5

Fix $S \subseteq [T]$, $\tau \in S$, and $q < N\zeta^{-1}$. By Lemma 8, there exists a symmetric optimal solution for the bad user's optimization problem under zero transaction-privacy, so for N accounts, the total volume for each account is q/N. We can then write the objective as

$$\max_{(Q_i)_{i\in S}, \alpha, N} N\left(\left(1 - \frac{\zeta q}{N}\right)^N \left(\sum_{i\in S\setminus\{\tau\}} \theta_i Q_i + \theta(\tau, \alpha)\cdot q_\tau\right) - \left(\sum_{i\in S\setminus\{\tau\}} \eta_i Q_i^2 - \eta(\tau,\alpha)\cdot q_\tau^2\right)\right) - \alpha\phi_\tau - \sum_{i\in S\setminus\{\tau\}} \phi_i$$

$$\text{s.t.} \sum_{i\in S} Q_i = \frac{q}{N}$$

$$Q_i \geq 0\ \forall\ i \in S$$

$$\alpha \geq 1.$$

Let $\theta_{(1)} = \max\{\max_{i \in S \setminus \{\tau\}} \theta_i, \theta(\tau, \alpha)\}$. We first provide an upper bound on the objective:

$$
\begin{aligned}
N \left(\left(1 - \frac{\zeta q}{N}\right)^N \left(\sum_{i \in S \setminus \{\tau\}} \theta_i Q_i + \theta(\tau, \alpha) \cdot q_\tau \right) - \left(\sum_{i \in S \setminus \{\tau\}} \eta_i Q_i^2 - \eta(\tau, \alpha) \cdot q_\tau^2 \right) \right) - \alpha \phi_\tau - \sum_{i \in S \setminus \{\tau\}} \phi_i \\
\leq N \left(1 - \frac{\zeta q}{N}\right)^N \theta_{(1)} \sum_{i \in S} Q_i - \alpha \phi_\tau - \sum_{i \in S \setminus \{\tau\}} \phi_i \\
\leq \left(1 - \frac{\zeta q}{N}\right)^N \theta_{(1)} q - \alpha \phi_\tau - \sum_{i \in S \setminus \{\tau\}} \phi_i \leq e^{-\zeta q} \theta_{(1)} q - \alpha \phi_\tau - \sum_{i \in S \setminus \{\tau\}} \phi_i.
\end{aligned}
$$

We can achieve this upper bound by creating N accounts, submitting q/N through technology $\arg\max_{i \in S} \theta_i$, and taking $N \to \infty$:

$$
\begin{aligned}
u(S, q) &= \lim_{N \to \infty} N \left(\left(1 - \frac{\zeta q}{N}\right)^N \theta_{(1)} \frac{q}{N} - \eta_{(1)} \frac{q^2}{N^2} \right) - \alpha \phi_\tau - \sum_{i \in S \setminus \{\tau\}} \phi_i \\
&= \lim_{N \to \infty} \left(\left(1 - \frac{\zeta}{q}\right)^N \theta_{(1)} q - \frac{\eta_{(1)} q^2}{N} \right) - \alpha \phi_\tau - \sum_{i \in S \setminus \{\tau\}} \phi_i \\
&= e^{-\zeta q} \theta_{(1)} q - \alpha \phi_\tau - \sum_{i \in S \setminus \{\tau\}} \phi_i.
\end{aligned}
$$

Maximizing the attainable upper bound across q and α yields $q^\star = \zeta^{-1}$ and $\alpha^\star = 1$, since $\theta(\tau, \alpha)$ is actually constant in α. The resulting utility is

$$
u(S) = e^{-\zeta q^\star} \theta_{(1)} q^\star - \sum_{i \in S} \phi_i = \frac{\theta_{(1)}}{e\zeta} - \sum_{i \in S} \phi_i
$$

Optimizing over all $S \subseteq \mathcal{S}$ yields

$$
u^\star = \max_{i \in \mathcal{S}} \left(\frac{\theta_i}{e\zeta} - \phi_i \right).
$$

By Assumption 1, $u^\star > 0$, so the bad user's best-response in an identity-private system with no transaction privacy results in a bounded welfare cost. Since the baseline model corresponds to $\alpha = 1$, the bad user can attain at least the baseline model's payoffs if transaction privacy is provided at some $q \in \mathbb{R}_+$. Using an argument similar to that in the proof of Proposition 1, the first result follows.

For the second result, we use an argument similar to that in the proof of Proposition 2. By

Lemma 5, $M^{rad}_{S,\tau}$ is increasing for any $S \subseteq [T]$, so $M^{rad}(q) = \max_{S \subseteq [T]} M^{rad}_{S,\tau}(q)$ is increasing as well. Since WC is also increasing in q, the form of the welfare-optimal contract given a payoff level u in Lemma 4 reduces to a single interval from 0 to a threshold $\bar{q}$.

Under $\mathcal{I} = 1$ and $p^\star_1$, the welfare benefit is $\int_0^{\zeta^{-1}} (1 - \delta_g \zeta q) q \, dF_g(q)$ and the welfare cost is ζ^{-1}. Under $\mathcal{I} = 0$ and p given by $p(q) = \mathbb{1}\{q \leq \zeta^{-1}\}$, the welfare benefit is $\int_0^{\zeta^{-1}} q \, dF_g(q)$ and the welfare cost is at most ζ^{-1}. The third result then follows from an argument similar to that in the proof of Theorem 3.

A.12 Proof of Proposition 6

Privacy has benefits to the good user when $\zeta \leq \underline{Q}^{-1}$ and our assumption that $\underline{Q} > (\max \theta_i / e\phi_i)^{-1}$ implies that Assumption 1 holds for all $\zeta < \underline{Q}^{-1}$, which is our region of interest. Then $\Delta WC = \zeta^{-1}$ on this region and

$$\begin{aligned}
\Delta WB(\zeta) &= \delta_g \int_{\min\{\underline{Q}, \zeta^{-1}\}}^{\min\{\overline{Q}, \zeta^{-1}\}} (1 - \zeta q) q \, dF_g(q) \\
&= \delta_g \int_{\min\{\underline{Q}, \zeta^{-1}\}}^{\min\{\overline{Q}, \zeta^{-1}\}} (q - \zeta q^2) q^{-\kappa} \, dq = \delta_g \left(\frac{q^{2-\kappa}}{2-\kappa} - \frac{\zeta q^{3-\kappa}}{3-\kappa} \right) \Bigg|_{\min\{\underline{Q}, \zeta^{-1}\}}^{\min\{\overline{Q}, \zeta^{-1}\}}
\end{aligned}$$

If $\kappa > 1$, it follows that $\Delta WB(\zeta)$ is convex in ζ. Similarly, ζ^{-1} is convex. As $\zeta \to 0$, $\Delta WB(\zeta)$ is bounded while ζ^{-1} is unbounded, and there exists some ζ where $\Delta WB(\zeta) = 0$ while ζ^{-1} is always positive. From convex analysis, it follows that $\Delta WB(\zeta)$ intersects ζ^{-1} nowhere, at exactly one point (for which the interval $[\zeta^\star_L, \zeta^\star_H]$ collapses to a point), or at two points.