Identity Theft and Data Breaches

William Roberds\textsuperscript{1}  Stacey L. Schreft\textsuperscript{2}

June, 2009

\textsuperscript{1}Federal Reserve Bank of Atlanta
\textsuperscript{2}Mutual Fund Research Center, LLC
Identity theft: some facts

- 2006 FTC Survey (Synovate 2007)
  - 3.7% of U.S. households victimized
  - Estimated annual cost:
    - FTC: $16 billion
    - Schreft (2007) adjusts up to $64 billion

- Big question: *is this a market failure?*
ID theft a market failure?

- Popular wisdom: **YES**
  - “too much” data (PII) collected & stored
  - unauthorized access (breaches) too easy
  - ID theft too common
- Legal literature: **YES**
  - Swire (2003): credit & payments industry has not delivered “efficient confidentiality” of PII, i.e., market failure has occurred
- Elected officials: **YES**, e.g.
  - U.S. 2003 FACT Act $\implies$ 30+ pages of Federal regulations
  - Breach notification laws in 36 states
On the other hand

- **Industry view:** ID theft **NOT a market failure**
- Industry
  - fraud losses "**low**" relative to usage of systems (\(> \$3 \text{ trillion card payments/year}\))
- Industry: collecting PII **deters** identity thieves
  - Surveys (e.g. Synovate 2007): much ID theft is **low tech** (stolen wallets, acquaintance fraud), does not stem from data breaches
  - (However Gordon et al. 2007: 50% of ID theft convictions result from business data theft)
- Industry
  - if there is a problem, solution is to collect **more** (e.g., biometric) data
This paper

- Theoretical examination of popular wisdom/industry view using economics of payments
- EOP: study of mechanisms that allow people to trade when
  1. People want to consume at different times
     - (intertemporal displacement of consumption/production)
  2. Limited enforcement of promises of future actions
Key tradeoff

Payment systems must deter 2 kinds of identity thieves

- *Unskilled frauds* ("opportunists"): **discouraged** by systems’ collection of PII, data security not important for deterrence
- *Skilled frauds* ("hackers"): possibly **enticed** by systems’ collection of PII, data security key for deterrence

**Efficiency** requires balance between data collection and data security
Model: basic features

- Infinite horizon, continuous time
- Large number of risk-neutral agents, congenitally split into multiple (2) groups $G_A$ & $G_B$
  - All transactions occur within a given group
- Overlapping generations: random subset of each group dies at dates 0, 1, 2, ... and is replaced by new agents
Agents also partitioned according to **legitimacy** and **type**

- **Legitimate agents**: can produce tradeable goods, no talent for fraud, measure $1 - F$
- **Frauds**: cannot produce goods, but can impersonate others, measure $F$

Agents distributed over **types** (virtual locations); many agents at each location

(Legitimate) type $y \in [0, 1]$ agent can **produce** unit of nondurable good of type $y$ at times $y, 1 + y, 2 + y, \ldots$ at cost $c$

At all other times $y' \in [0, 1], y' \neq y$, agent of type $y$ wants to **consume** goods of (randomly selected person of) other types, generating flow utility $u > c$

<consumption/production displacement>
Each agent has unique **identity**, time-invariant vector of personal data (not transactions history), effectively infinite dimension

Subset of identity (**PII**) may be assembled, stored, secured at positive cost

Agent’s group, type, legitimacy & identity are **private information** subject to

- **costly & imperfect verification** and/or
- **revelation** through agent’s behavior (not available instantaneously)

<informational frictions $\Rightarrow$ imperfect enforcement of promises>
Trade through payment networks

- No repeated interactions ⇒ unless agents’ behavior can be tracked, no agent would ever produce
- Payment networks modeled as **clubs** for sharing information on agents’ behavior
- One club for each group; **no info sharing across clubs**; club membership voluntary
- Information compiled by club
  - (1) members’ **production history** (has an agent produced goods for other members?)
  - (2) members’ PII (so as to correlate individuals with histories, distinguish new members from old)
- Production information, PII available at discrete dates 0, 1, 2, . . .
Clubs’ operation

- At 0, 1, 2, … clubs $G_A$ & $G_B$ open membership to all
- Agents wishing membership in club $i$ must submit PII of dimension $d_i > 0$ (if not already on file)
- Each club member receives uncounterfeitable credit card entitling agent to goods produced by other (legitimate) club members
- In general, clubs not viable (not IR for legit agents) if legit agents must produce for all agents, including frauds
- $\Rightarrow$ Clubs exclude nonproducers at discrete dates, when production info becomes available
- Can apply penalties to non producers (bill collection) but only if stored PII corresponds to “real identity”
Data costs

- Club \(i\) collects, stores data \(d_i\)

- **One-time cost** \(K\) when member first joins, plus proportional storage cost \(kd_i\) per unit time
  - \(K, k\) include intangibles ("loss of privacy")
  - if data not stored, initial verification cost must be incurred

- Club \(i\) applies **security level** ("skill threshold") \(s_i\) at cost \(\ell s_i\)

- Hacking skills \(s\) have some distribution \(\Phi(s)\) over population of frauds
Frauds can join clubs by impersonating a legitimate agent; frauds either skilled or unskilled

- Unskilled group $i$ frauds ($s \leq s_j$) can join club $i$ w/o revealing true identity at effort cost $\varepsilon d_i$ where $\varepsilon \geq 0$ has distribution $\Gamma$

- Skilled group $i$ frauds ($s > s_j$) lower effort cost by stealing (breaching) data held by other club $j$ (club $i$ rejects duplicate identities) at lower effort cost

$$\varepsilon \max\{d_i - \eta d_j, 0\}$$

- $\eta \in (0, 1)$ measures overlap between 2 clubs’ databases of members’ PII; determines spillover effects

- <note: successful ID theft always revealed after one period>
Identity theft: costs per incidence of fraud

1. $c$: cost to legitimate members of club $i$ of providing goods to identity thieves (e.g., FTC: median cost $1,350/ stolen ID)

2. $L$: additional cost (time, inconvenience, intangible) to club $i$ of resolving fraud (FTC: resolution time 10 hours/ stolen ID)

3. $B$: cost to club $i$ when club $j$ ID theft results from breach of club $i$’s data (Ponemon Institute 2006: $< $100 / record breached)

Model calculations assume $c + L > B$
Steady-state allocations and objectives

- **Allocation:** PII and security \((d_i, s_i)\) for each club \(i = G_A, G_B\)
- **Objectives:** Each club chooses \((d_i, s_i)\) to maximize value of legitimate membership

\[
\text{(transaction benefit)} - \text{(data costs)} - \text{(ID theft costs)}
\]
Comparison of allocations


- **Symmetric Nash equilibrium** \((d^*, s^*)\)
  - maximizes club \(i\) membership value when \(j\) also chooses \((d^*, s^*)\)

- **(Constrained) efficient** \((d_p, s_p)\)
  - maximizes steady-state value of legitimate club membership for both clubs

- Game plan: characterize “market failures” as deviations of Nash from efficient allocation
Nash equilibrium: sources of inefficiency

- Externalities present in both decision variables; each club
  1. Internalizes deterrence **benefits** of PII collection $d$ but not **costs** to other club (facilitation of future skilled ID theft)
  2. Does not internalize full **benefits** of data security $s$ (reduction in skilled ID theft to other club)
Nash equilibrium: manifestations of inefficiency

- With sufficiently high data overlap ($\eta \to 1$) and sufficiently low data costs ($k, \ell \to 0$)
  - inefficient overaccumulation of PII, inefficiently low levels of data security

- Perhaps less obviously
  1. Unskilled ID theft **inefficiently low** (because too much PII collected)
  2. Skilled ID theft **inefficiently high** (data undersecured)
  3. For $(k/\ell)$ bounded (persistent intangible privacy cost), total ID theft **inefficiently low** (first effect dominates)

- Inefficiency of Nash equilibrium consistent with stylized facts
  - “low” ID theft rates of both types
  - prevalence of unskilled ID theft
<table>
<thead>
<tr>
<th>Variable</th>
<th>Eq. vs. efficient value</th>
<th>Popular wisdom?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data length $d$</td>
<td>Higher</td>
<td>Yes</td>
</tr>
<tr>
<td>Data security $s$</td>
<td>Lower</td>
<td>Yes</td>
</tr>
<tr>
<td>Skilled ID theft rate</td>
<td>Higher</td>
<td>Yes</td>
</tr>
<tr>
<td>Unskilled ID theft</td>
<td>Lower</td>
<td>No</td>
</tr>
<tr>
<td>Total ID theft</td>
<td>Lower</td>
<td>No</td>
</tr>
</tbody>
</table>
Policy approaches
Summary of analytical / numerical results

1. Increase civil liabilities for a data breach (up to economic loss)
   - limited effectiveness; does not shut down substitution of data collection for security

2. Enforce higher security standards
   - can approximate efficient allocation but requires very high data security standard

3. Constrain PII collected
   - in welfare terms, almost as effective as (2) but may lead to unacceptably high ID theft rate
Summary of paper

1. Develops a meaningful concept of “efficient confidentiality” for PII levels of PII and security that enable beneficial exchange at minimum cost

2. Characterizes market failures
   - consistent with popular wisdom in some dimensions, not others
   - can be consistent with facts emphasized in industry discussions

3. Analyzes policy interventions

4. Provides generalizable framework