Optimal Data Security with Redundancies

Rod Garratt and Linda M. Schilling

University of California Santa Barbara and BIS
Olin Business School - Washington University in St. Louis, CEPR

The future.s of money - Paris 2022

Motivation

Data is valuable, can be monetized

- Google, Facebook, Amazon (targeted ads)
- Payment Data: Apple Pay, Libra/Diem, WeChat, Alipay
- Cambridge Analytica

Control over Information attracts attackers

- Attack:
  - Stealing of information
  - Blocking access to reading and writing the ledger
  - compromising information (alteration)
- Equifax data breach (2017, 147 mio accounts), Moscow Stock Exchange and Sberbank cyber attack (Feb 2022), JP Morgan (2014, 83 mio customer accounts)
- Double-spending attack
- Governments freezing citizen's accounts

Motivation

Firms underinvest in data security (neglect social cost)

- 2017: IBS Intelligence states that a new report from the Online Trust Alliance (OTA) claims that up to 65% of US banks are 'extremely' vulnerable to cyber attacks.
- Firms neglect (social) value of data to their customers

Motivation

Data Security involves two dimensions

- How many entities observe information I (redundancy)
- How much information does each entity observe (segmentation)?
  - complementarity vs
  - overlap of information sets

Motivation

Data Security involves two dimensions

- How many entities observe information I (redundancy)
- How much information does each entity observe (segmentation)?
  - complementarity vs
  - overlap of information sets

1. More Redundancy implies
   (Miners, RAID backup 'Redundant Array of Independent Disks'), '# information covers'
   - Less privacy: everybody (every backup) observes everything
     $\Rightarrow$ Creates free-riding w.r.t. investment in security
   - improves censorship-resistance: If one party is attacked, the ledger is robustly protected since there are backup entities
     $\Rightarrow$ full recovery possible

   (Substitutability of information entities)
2. Data Segmentation
   - More information is valuable, worth protecting more
     $\Rightarrow$ large data entities should invest more in data security
   - More information is more costly to store
   - More information increases chance of an attack

This paper

How characterize optimal data security along the dimensions

- data redundancy
- data segmentation

Contrast economics of different information systems against one another

- Should a central bank observe all payment data via a CBDC (Fedwire)?
- When observing more data, do entities invest more in data security?

Literature

- Value of information
  - Hirshleifer, 1978
  - Morris and Shin 2002
  - Angeletos and Pavan 2004, 2007
- Incentives to involved with blockchain
  - Budish (2018)
  - Biais, Bisiere, Bouvard, Casamatta (2019)
  - Hubermann, Leshno, Moallemi (2021)
  - Ebrahimi, Routledge, Zetlin-Jones (2020)
- (Monetizing) Privacy
  - Garrat and Lee, 2021
  - Garratt and v Oordt (2021)
- Computer Science literature on Data Backups
  - Littlewood and Strigini, 2004;
  - Ghaffarzadegan, 2008;
  - Jia, Xin, Wang, Guo and Wang, 2018;
  - AlZain, Soh and Pardede, 2012

Model I

- Total data (ledger, transaction data, payment data): I
- Set of entities that observe data $\mathcal{N}$
  - $i \in \mathcal{N},\ i = 1, \ldots, N$
  - financial intermediaries, miners, central bank
- Each entity $i \in \mathcal{N}$ observes a subset of the ledger $I_i \subseteq I$ (I_i is exogenous to i, information sets are common knowledge)
- Each piece of data is observed by at least one entity: All entities jointly observe the full ledger

$$I \subseteq \bigcup_{i=1}^{N} I_i \tag{1}$$

  That is, the information sets $I_1, \ldots I_N$ form a 'topological cover' of the full ledger I
- Redundancy: Multiple entities may (but do not have to) simultaneously observe the same piece of information. Potentially:

$$I_i \cap I_j \neq \emptyset, \quad i, j \in \mathcal{N} \tag{2}$$

Model II

- Information I_i
  - Information is valuable: Revenue $R(|I_i|)$ is strictly increasing in $|I_i|$
  - Value of information is symmetric: $R(|I_i|)$ only depends on $|I_i|$ (quantity of information), not I_i (content, quality of information)
  - Revenue is only earned when data I_i is 'not compromised'
- Investment in data security c_i:
  - variable cost to entity $i \in \mathcal{N}$: reduces profits by $-c_i \times f(|I_i|)$
  - reduces chance of an attack on entity i
- Attack (loss of (reading and writing of) data, loss of reliability of data): Entities are attacked with probability $\alpha(c_i, |I_i|) \in [0, 1]$
  - Entities can reduce chance of an attack: $\alpha(\cdot, \cdot)$ is strictly decreasing and convex in c_i ($i's$ investment in data security) with $\lim_{c \to 0} \alpha(c) = 1$ and $\lim_{c \to \infty} \alpha(c) = 0$.
  - Observing more data makes an entity a more likely target of an attack: $\alpha(\cdot, \cdot)$ is strictly increasing in $|I_i|$
  - Information security is costlier if more information is observed. $\frac{\partial}{\partial |I_i|} \frac{\partial}{\partial c_i} \alpha \geq 0$

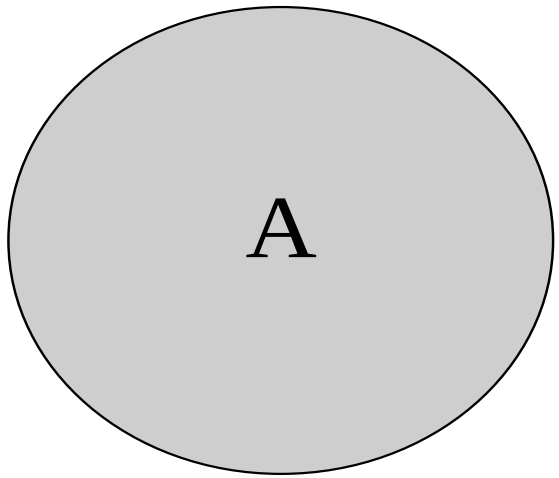
Model III - Data Recovery

- Full Ledger Backup (single cover): Entities $i = 1, \dots m \in \mathcal{N}$, form a 'validation collective' of I if their joint information sets cover the full ledger

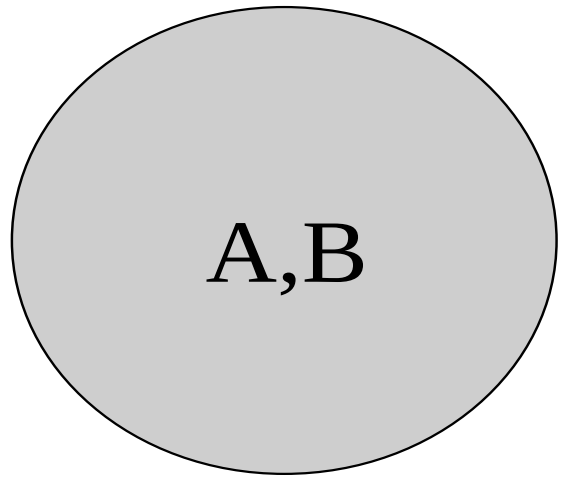
$$I \subseteq \bigcup_{i=1}^{m} I_i, \; m \leq N \tag{3}$$

- Partial Backup (single cover of I_i): 'Validation collective of $I_i \subset I$', entities that serve as backup to i's information.
  $I_i \subseteq \bigcup_{j=1}^{m} I_j, \; m < N, \; j \neq i$
- Backup depth behind entity i: Number of *mutually exclusive* validation collectives I_i
- Ass: $R(|I_i|)$ independent of the covers of I_i (no competition)
- Given an attack on i
  - data I_i is 'compromised' only if *all* validation collectives of I_i are jointly successfully attacked
  - Given an attack on i was successfull, but at least one attack on a validation collective of I_i was not, entity i can recover data I_i at zero costs.

Demonstration: Redundancy



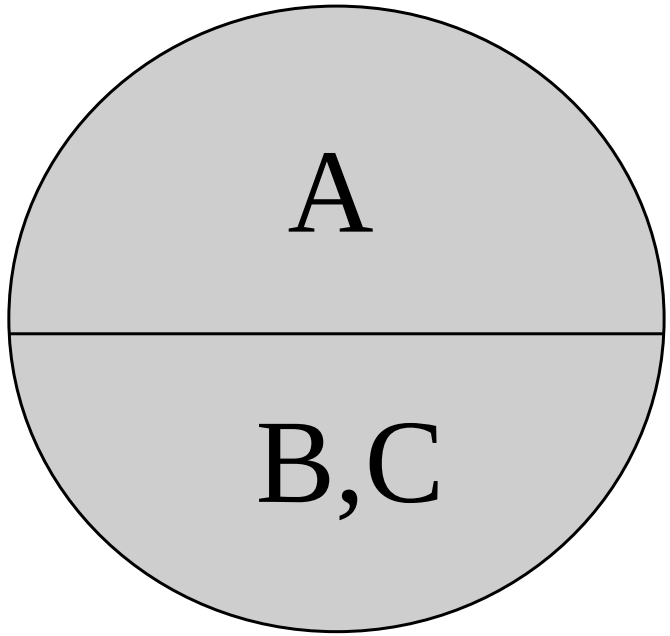
(a) Monopolistic CBDC



(b) 2 miners

- In either Information System: All entities observe the same (quantity of) information $|I_A| = |I_B|$
- In (a): A's data is uniquely observed (no Backup)
- In (b): A, B mutually serve as a back up (validation collective)

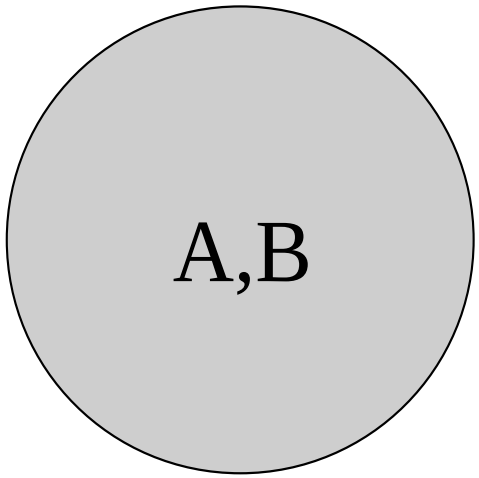
Demonstration: Redundancy



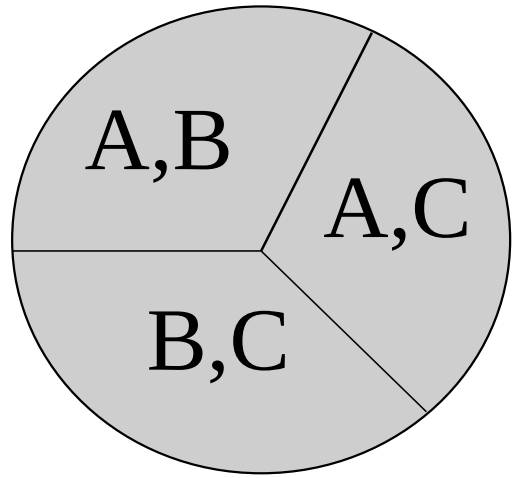
- All entities observe the same quantity of information $|I_A| = |I_B| = |I_C|$
- B, C mutually serve as a back up
- A's data is uniquely observed

$\Rightarrow$ A will behave differently from B & C

Demonstration: Information segmentation



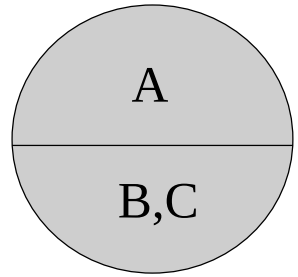
(c) perfect substitutes



(d) partial substitutes

- In either information system: redundancy (# covers) is 2
- In (a): A, B observe full ledger $|I_i| = |I|$, $i \in \{A, B\}$
- In (b): A,B,C observe less $|I_i| < |I|$, $i \in \{A, B, C\}$

Analysis: Example I - Private Optimum



Private Optimum (c_A^*, c_B^*, c_C^*):

- Profit A:

$$\pi_A(c_A) = (1 - \alpha(c_A))\ R(|I_A|) - c_A\ f(|I_A|) \tag{4}$$

- Profit B,C (symmetric)

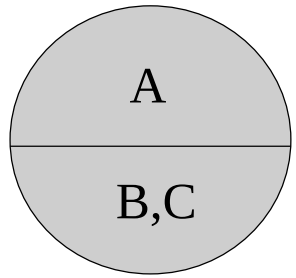
$$\pi_B(c_B) = (1 - \alpha(c_B)\alpha(c_C))\ R(|I_B|) - c_B\ f(|I_B|) \tag{5}$$

Equilibrium Conditions

- $$-\alpha'(c_A^*) = \frac{f(|I_A|)}{R(|I_A|)} \tag{6}$$

- $$-\alpha'(c_B^*)\alpha(c_C^*) = \frac{f(|I_B|)}{R(|I_B|)} = \frac{f(|I_A|)}{R(|I_A|)} = -\alpha'(c_C^*)\alpha(c_B^*) \tag{7}$$

Analysis: Example I - Private Optimum



Proposition (Example I: Private Optimum)

(i) In every private equilibrium: $c_A^* > c_B^*, c_C^*$ *with* $-\alpha'(c_A^*) = \frac{f(|I_A|)}{R(|I_A|)}$.
(ii) There exists a unique private equilibrium with symmetric costs $c_B^* = c_C^*$ *in which case* c_B^* *is given as the solution to* $-\alpha'(c_B^*)\alpha(c_B^*) = \frac{f(|I_B|)}{R(|I_B|)}$.
(iii) If $\alpha'(c)/\alpha(c)$ *is strictly monotone in the expenditure, then the private equilibrium is unique and satisfies* $c_B^* = c_C^*$.

Proof

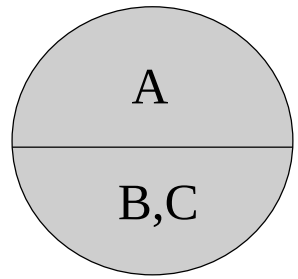
$$\alpha'(c_A^*) = \alpha'(c_B^*)\alpha(c_C^*) > \alpha'(c_B^*) \tag{8}$$

Because $\alpha' < 0$. Further $\alpha'' > 0$, thus $c_A^* > c_B^*$.
Unique symmetric equ: $\alpha'(c)\alpha(c)$ is strictly increasing.

Analysis: Example I -Social Optimum

Social planner

- has symmetric valuation for data $|I|$
- $L(|I|)$ social loss to the public when $|I|$ compromised



$$\pi_P(c_A, c_B, c_C) = \pi_A + \pi_B + \pi_C - \alpha(c_A)L(|I_A|) - \alpha(c_B)\alpha(c_C)L(|I_B \cup I_C|)$$
$$= \pi_A + \pi_B + \pi_C - (\alpha(c_A) + \alpha(c_B)\alpha(c_C))\ L(|I_A|)$$

Proposition (Socially optimal vs Private optimal expenditure)

Under the private solution there is underinvestment in data security. The social optimum $(\hat{c}_A, \hat{c}_B, \hat{c}_C)$ relative to the privately optimal solution (c_A^*, c_B^*, c_C^*) satisfies $\hat{c}_A > c_A^*$, $\hat{c}_B > c_B^*$ and $\hat{c}_C > c_C^*$. (All entities underinvest.)

Analysis: Example I -Social Optimum

Proposition (Social Optimum)

The social optimum $(\hat{c}_A, \hat{c}_B, \hat{c}_C)$ *satisfies the same ordering as the private optimum:* $\hat{c}_A > \hat{c}_B, \hat{c}_C$. *A's optimal investment is uniquely, implicitly given as*

$$-\alpha'(c_A) = \frac{f(|I_A|)}{[R(|I_A|) + L(|I_A|)]} \tag{9}$$

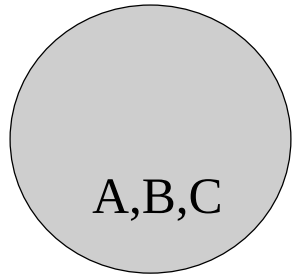
Further, there exists a unique equilibrium $(\hat{c}_A, \hat{c}_B, \hat{c}_C)$ *with* $\hat{c}_B = \hat{c}_C$ *where socially optimal investment by* B *and* C *is pinned down via*

$$-\alpha'(c_B)\alpha(c_B) = \frac{f(|I_B|)}{[R(|I_B|) + L(|I_B|)]}. \tag{10}$$

If α'/α *is strictly monotone, then the social optimum is unique and has the property* $\hat{c}_B = \hat{c}_C$.

Example II - Decentralized Ledgers

$I \equiv I_A = I_B = I_C$



Private Optimum $(c_A^{**}, c_B^{**}, c_C^{**})$:

- Profits for each entity $i \in \{A, B, C\}$ are given by

$$\pi_i(c_i) = \left(1 - \alpha(c_i) \times \prod_{j \in \{A,B,C\} \setminus i} \alpha(c_j)\right) R(|I|) - c_i \, f(|I|)$$

Equilibrium conditions

-

$$-\alpha'(c_i^*) \underbrace{\prod_{j \in \{A,B,C\} \setminus i} \alpha(c_j)}_{\text{increases in \# covers}} = \underbrace{\frac{f(|I|)}{R(|I|)}}_{\text{const. in \# covers}}, \; i = A, B, C$$

Example II - Decentralized Ledgers

- Joint equilibrium conditions

$$\alpha'(c_A)\alpha(c_B) = \alpha(c_A)\alpha'(c_B) \tag{11}$$

$$\alpha'(c_C)\alpha(c_B) = \alpha(c_C)\alpha'(c_B) \tag{12}$$

$$\alpha'(c_A)\alpha(c_C) = \alpha(c_A)\alpha'(c_C) \tag{13}$$

- or alternatively $\frac{\alpha'(c_A)}{\alpha(c_A)} = \frac{\alpha'(c_B)}{\alpha(c_B)} = \frac{\alpha'(c_C)}{\alpha(c_C)}$

Proposition

There exists a unique symmetric private equilibrium $c_A^{**} = c_B^{**} = c_C^{**}$ *where the expenditure level is pinned down via*

$$-\alpha'(c_A^*)\alpha^2(c_A^*) = \frac{f(|I|)}{R(|I|)} \tag{14}$$

The symmetric equilibrium is the unique equilibrium if $\alpha'(c)/\alpha(c)$ *is strict monotone in the expenditure* c*.*

Example II - Decentralized Ledgers- Social Optimum

The social planner maximizes

$$\pi_P(c_A, c_B, c_C) = \pi_A(c_A) + \pi_B(c_B) + \pi_C(c_C) - \alpha(c_A)\alpha(c_B)\alpha(c_C)\ L(|I|)$$

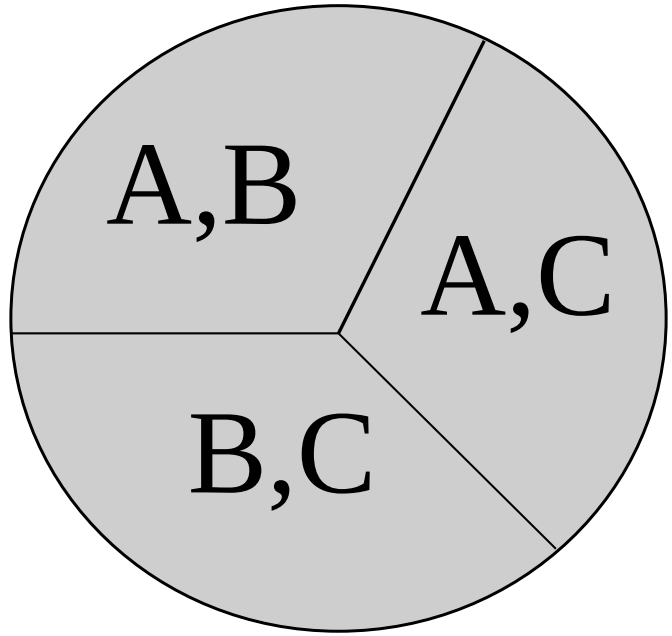
Proposition (Socially optimal investment in DLT data security)

The social optimal investment in DLT data security $(\hat{\hat{c}}_A, \hat{\hat{c}}_B, \hat{\hat{c}}_C)$ *exceeds private investment:* $\hat{\hat{c}}_i > c_i^{**}$ *for all* $i \in \{A, B, C\}$.

Proposition (Socially optimal investment in DLT data security)

There exists a unique symmetric social equilibrium with $c_A = c_B = c_C$, *where expediture is pinned down implicitly via* $-\alpha'(c_A)\alpha^2(c_A) = \frac{f(|I|)}{(R(|I|)+L(|I|))}$. *The symmetric equilibrium is the unique equilibrium if* $\alpha'(c)/\alpha(c)$ *is strict monotone in the expenditure* c.

Example III: Double Covers with Information Segmentation



- Symmetry $|I_A| = |I_B| = |I_C| = 2/3|I|$
- Data Segmentation: No entity observes everything (difference to DLT)
- Redundancy: Every two entities out of $\{A, B, C\}$ jointly recover the full ledger (2 validation collectives) $I_C \subset I_A \cup I_B$, $I_B \subset I_A \cup I_C$, $I_A \subset I_C \cup I_B$.

Example III: Double Covers with Information Segmentation - Private Optimum

Then profits are given as

$$\pi_A(c_A) = (1 - \alpha(c_A)(\alpha(c_B) + \alpha(c_C)))\ R(|I_A|) - c_A\ f(|I_A|) \tag{15}$$

The first order condition becomes

$$\frac{\partial}{\partial c_i}\pi_i(c_i) = -\alpha'(c_i)\left(\sum_{j\in\{A,B,C\}\setminus\{i\}} \alpha(c_j)\right)\ R(|I_i|) - \ f(|I_i|) = 0 \tag{16}$$

By $|I_A| = |I_B| = |I_C|$, the private equilibrium must satisfy

$$\alpha'(c_A^*)\,(\alpha(c_B^*) + \alpha(c_C^*)) = \alpha'(c_B^*)\,(\alpha(c_A^*) + \alpha(c_C^*)) = \alpha'(c_C^*)\,(\alpha(c_B^*) + \alpha(c_A^*)) \tag{17}$$

Proposition (Double Covers)

Let $|I_A| = |I_B| = |I_C|$ and $|I_{AB}| = |I_{BC}| = |I_{AC}|$. There exist a unique private symmetric equilibrium $c_A^* = c_B^* = c_C^*$ where the equilibrium expenditure level solves $-\alpha'(c_A^*)\alpha(c_A^*) = \frac{1}{2}\frac{f(|I_i|)}{R(|I_i|)}$,

Example III: Double Covers with Information Segmentation - Social Optimum

The social planner problem.

$$\pi_P(c_A, c_B, c_C) = \pi_A + \pi_B + \pi_C - (\alpha(c_A)\alpha(c_B) + \alpha(c_B)\alpha(c_C) + \alpha(c_A)\alpha(c_C))\ L(|I_{AC}|)$$

where $I_{AB} \equiv I_A \cap I_B$, and anologously I_{BC}, I_{AC}. By symmetry $|I_{AB}| = |I_{BC}| = |I_{AC}|$ and $|I_A| = |I_B| = |I_C|$ it follows:

Example III: Double Covers with Information Segmentation - Social Optimum

The social planner problem.

$$\pi_P(c_A, c_B, c_C) = \pi_A + \pi_B + \pi_C - (\alpha(c_A)\alpha(c_B) + \alpha(c_B)\alpha(c_C) + \alpha(c_A)\alpha(c_C))\ L(|I_{AC}|)$$

where $I_{AB} \equiv I_A \cap I_B$, and anologously I_{BC}, I_{AC}. By symmetry $|I_{AB}| = |I_{BC}| = |I_{AC}|$ and $|I_A| = |I_B| = |I_C|$ it follows:

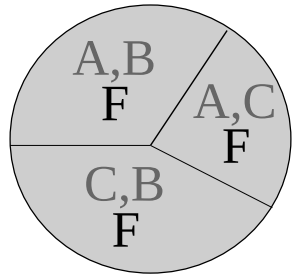
Proposition (Double Covers: Social Equilibrium)

There exist a unique symmetric social optimum $\hat{c}_A = \hat{c}_B = \hat{c}_C$ *where the expenditure level solves* $-\alpha'(c_A)\ \alpha(c_A) = \frac{1}{2}\ \frac{f(|I_A|)}{(R(|I_A|)+L(|I_{AB}|))}$

Proposition (Double Covers: Social versus Private Optimum)

In every private equilibrium, there is underinvestment in data security relative to the social optimum. $c_i^* < \hat{c}_i$, *for all* $i \in \{A, B, C\}$.

Example IV: Double Covers with Fedwire



$$\pi_A(c_A) = (1 - \alpha(c_A)\alpha(c_F)(\alpha(c_B) + \alpha(c_C)))\ R(|I_A|) - c_A\ f(|I_A|) \quad (18)$$

Via the first oder condition

$$-\alpha'(c_A)\alpha(c_F)(\alpha(c_B) + \alpha(c_C))\ = \frac{f(|I_A|)}{R(|I_A|)} \quad (19)$$

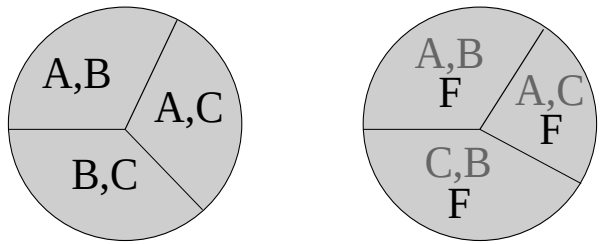
Proposition (Private equilibrium with Fedwire)

With Fedwire, there exist a unique private symmetric equilibrium $c_A^{*,F} = c_B^{*,F} = c_C^{*,F}$ *where* $c_A^{*,F}$ *solves*

$$-\alpha'(c_A)\alpha(c_A)\ = \frac{1}{2\ \alpha(c_F)} \frac{f(|I_A|)}{R(|I_A|)}$$

Data Security across Information Systems I

Adding a single backup: Compare case of Double Covers with versus without Fedwire



(e) Double Covers

(f) with Fedwire

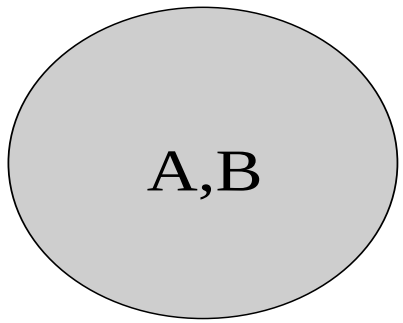
$$-\alpha'(c_A^*)\alpha(c_A^*) = \frac{1}{2}\,\frac{f(|I_i|)}{R(|I_i|)} < \frac{1}{2\;\alpha(c_F)}\,\frac{f(|I_A|)}{R(|I_A|)}$$

Proposition (Double Covers versus Fedwire)

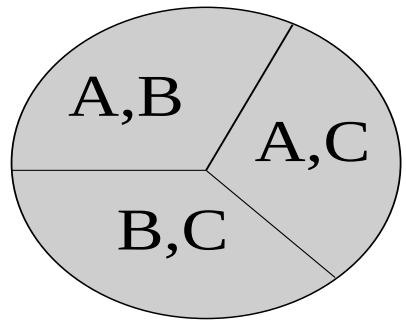
When adding Fedwire as an additional cover, all entities reduce their investment in data security in the private optimum, $c_i^* > c_i^{*,F}$ for all $i \in \{A, B, C\}$.

Data Security across Information Systems II

Effect of Data Segmentation for Fixed Covers $C = 2$



(g) DLT



(h) Double Covers

Trade-off

- Less data reduces revenue $\Rightarrow$ data protection pays off less
- Observing less information reduces chance of an attack
- Data segmentation 'diversifies', alters the chance of an attack on the ledger

Data Security across Information Systems II

Proposition (Double Covers versus DLT, $C = 2$ fixed)

Assume symmetric variable costs $f(|I_{2/3}|) = f(|I|)$. Assume $\alpha_I(c)$ is constant in $|I|$. Comparing the unique symmetric private equilibria:
(i) If $R(|I|) < 2R(|I_{2/3}|)$, then $c_A^{*,DLT} < c_A^{*,DC}$.
(ii) If $R(|I|) \geq 2R(|I_{2/3}|)$, then $c_A^{*,DLT} \geq c_A^{*,DC}$.

If $\alpha_I(c)$ const in $|I|$:

- Chance of data loss in DLT: $P(\text{attack}^{DLT}) = \alpha_I(c_A^{*,DLT})\alpha_I(c_B^{*,DLT})$.
- Chance of data loss in DC: $P(\text{attack}^{DC}) = \alpha_I(c_A^{*,DC})\alpha_I(c_B^{*,DC}) + \alpha_I(c_A^{*,DC})\alpha_I(c_C^{*,DC}) + \alpha_I(c_C^{*,DC})\alpha_I(c_B^{*,DC})$

$\Rightarrow$

- If $c_A^{*,DLT} > c_A^{*,DC}$, then

$$P(\text{attack}^{DLT}) = \alpha_I(c_A^{*,DLT})\alpha_I(c_B^{*,DLT}) < \alpha_I(c_A^{*,DC})\alpha_I(c_B^{*,DC}) < P(\text{attack}^{DC})$$

- If $c_A^{*,DLT} < c_A^{*,DC}$, possibly $P(\text{attack}^{DC}) < P(\text{attack}^{DLT})$

Data Security across Information Systems III

General Trade-off

- Less data reduces revenue ⇒ data protection pays off less
- But: Information security is cheaper if less information is observed

$$\frac{\partial}{\partial |I|}\frac{\partial}{\partial c}\alpha(c,|I|) \geq 0$$

.

Proposition (Segmentation externality: Double Covers versus DLT, $C = 2$)

Assume symmetric variable costs under double covers and DLT, $f(|I_{2/3}|) = f(|I|)$. *Comparing the unique symmetric private equilibria:*
(i) If $R(|I|) < 2R(|I_{2/3}|)$ *and* $\left(\frac{\partial}{\partial |I|}\frac{\partial}{\partial c}\alpha(c)\right)\alpha(c) > \left(-\frac{\partial}{\partial c}\alpha(c)\right)\left(\frac{\partial}{\partial |I|}\alpha(c)\right)$ *hold, then* $c_A^{*,DLT} < c_A^{*,DC}$.
(ii) If $R(|I|) \geq 2R(|I_{2/3}|)$ *and* $\left(\frac{\partial}{\partial |I|}\frac{\partial}{\partial c}\alpha(c)\right)\alpha(c) \leq \left(-\frac{\partial}{\partial c}\alpha(c)\right)\left(\frac{\partial}{\partial |I|}\alpha(c)\right)$ *hold, then* $c_A^{*,DLT} \geq c_A^{*,DC}$.

Conclusion

- Costly and risky to restructure payment and banking networks
- Analysis in this paper contributes on how to do that.

- We provide an ECONOMIC MODEL OF DATA SECURITY that features
  - Redundancies (backups)
  - Information Segmentation
- We observe
  - Redundancies cause free-riding
    $\Rightarrow$ Not necessarily increases security when all parties internalize the quantity of backups.
  - Information segmentation (keeping backups constant) can but does not have to improve security.