The Center for Advanced Studies
in Science and Technology Policy

Technical and Policy Challenges:
Implications for Emerging Business Models

K. A. Taipale

Executive Director, Center for Advanced Studies
Senior Fellow, World Policy Institute
Adjunct Professor of Law, NYLS

Presented at:
16th Annual Economic Crimes Institute Conference
Information at Risk: Challenge and Conflict
Arlington, VA • October 24, 2005

Overview

- Background and context
  - Achieve security (~accountability, trust, and risk management)
  - Control for privacy and ID theft (~identity and reputation)
- Technology-enabled opportunities and challenges
- Existing business models
  - Based on transient information economics
  - Encourage fraud
- New business models
  - Information society
  - *Proposed solution to identity theft and privacy*
- Conclusions?
- Discussion

Background, obligatory self-promotion, and caveats

- K. A. Taipale, "*The Trusted Systems Problem: Security Envelopes, Statistical Threat Analysis, and the Presumption of Innocence*," IEEE Intelligent Systems 81-83 (Sep/Oct 2005)
- My identity and reputation … "google me"
- POV: *Social informatics*
- Bias/COI: academic (cybercrime), policy (national security), business (dinosaur consulting, invest in disruptive technologies)
- Caveat/caution (abstract and general) (cf. previous work on incremental improvements through technology -- rules-based processing, selective revelation, systems security, and audit)

More prelude

- Security (~accountability) and "privacy" are *not dichotomous rivals* to be traded one for another in a zero-sum game
- Rather, they are *dual obligations*, each to be maximized within the constraints of the other
- But, this does *not mean that they are compatible*
- Indeed, there is a fundamental incompatibility (and no intrinsic reciprocal proportionality) between degrees of freedom and degrees of control in systems (thus, a *wicked problem*, not a balancing act)

Contextual overview

- Philosophical/social/political: changing notion of “security”
  - Move to *preemption* (risk management/counterparty assessment)
- Technology enabled change: changing *information economics*
  - Data interoperability and connectivity (~collection/*sharing*/access)
  - Analysis (~linking, data mining, visualization, etc.)
- Information policy issues (the two “problems”)
  - *Identity theft* (~data security or business process problem?)
  - “*Privacy*” (~intrusion, error, autonomy) (cf. secrecy)
- Emerging business models (*social informatics* trends)
  - Web 2.0 (~commodification of infrastructure and data)
  - Identity 2.0 (~portable/authoritative/selective revelation)

Philosophical/social/political context

- Unprecedented human mobility and the potential for catastrophic outcomes, see *Risk Revolution* (?), is leading to a:
- Transformation of modern societies from a notional Beccarian model of *accountability* for deviant actions after they occur,
- To a Foucauldian model of *authorization* and *preemption* through ubiquitous preventative surveillance, risk management, and control through system constraints.
- As a result, the role of private and public "security services" is changing from policing toward *risk management* through surveillance, information exchange, auditing, communication, and classification.

Underlying Security Strategies

- Accountability (DEFAULT=PERMIT)
  - Lower costs on functionality (high degrees of freedom)
  - Potentially high cost to security (~catastrophic outcomes)
  - ~ historically associated w/ freedom/liberty
- Authorization (DEFAULT=DENY)
  - Low cost of implementation, high degree of control
  - High cost to functionality (constrains degrees of freedom)
  - ~ historically associated w/ totalitarianism
- Caveat: making systems claim not political claim

Authorization strategies

- Require *conditional prediction* about behavior within a system (I.e., authentication of a “trust” attribute) (~ “reputation”)
- May or may not require “identification”
    - E.g., compare a search to deny capability (airport checkpoint) (systems constraints) with basing trust on reputation or risk assessment (CAPPS II, etc.)
- Suffer high cost to functionality (low degrees of freedom)
- Security value is directly related to reliability of trust indicator (confidence interval)
- Generally, do not scale well because of friction and implementation costs

Authorization strategies (cont)

- Suffer “trusted systems” problem:
  - Can never prove trustworthiness, only non-evidence yet of un-trustworthiness (e.g., not on watch list, not yet defaulted on financial obligation, not yet a traitor, etc.)
  - “Reputation” as conditional predictor
  - Profiling and CI (DMDS) (probative value not probabilistic nature)
- Any trust-based system will fail occasionally, thus,
  - Requires constant monitoring and adjustment (Bayesian - degree of truth in an uncertain statement) (cf. cc fraud monitoring)
  - Must be part of layered defense (~collapsing perimeter of defense)
  - And, designed for elegant failure

Accountability strategies

- Generally require authentication of one or more identity attributes that allow the imposition of a sanction on an actor (or group or third party) for deviant behavior
- But does not necessarily require a unique identifier
- The special case of surveillance and accountability
  - chilling effect (suppression or control through systems constraint)
  - evidentiary role (audit) (accountability)
  - real-time defense/response (compliment to authorization)
- Low implementation costs, but high cost to security (particularly where there is potential for critical/catastrophic outcomes from failures)

Attribute vs identifier

- Compare attribute vs. identifier
  - Both authorization and accountability strategies can use group or category attributes and don’t necessarily require unique identifiers (“find radius” is good enough)
  - Use of token or categorization (early passports describe holder only as “gentleman”) (US Persons -- arbitrary attribute)
  - Match attribute to transaction requirement. Example of poor implementation: name on credit card
- Also, compare anonymity and pseudonymity
  - Anonymous: ID cannot be attributed ~ non-accountable
  - Pseudonymous: ID cannot be attributed *in the ordinary course* ~ accountability through process)
  - Sup Ct anonymity cases - “identity leakage”

Identification and identity

- B. Traven (*The Death Ship,* 1926, 1934*)*

  “You ought to have some papers to show who you are.”

  “I do not need any papers. I know who I am.”

  “Maybe so. But others are also interested in who you are.”

- T. Gilliam (*Brazil*, 1985)

  “Do you want to see my papers?”

  “No need, sir”

  “But I could be anyone.”

  “No you couldn’t, sir, this is information retrieval.”

Identity and reputation

- What is identity
  - Who I say I am
  - Who you say I am
  - Who others say I am
  - ~Reputation
  - Multiple and variable identities (~compartmentalized)
- For our purposes:
  - Identity as a set of claims useful for counterparty assessment
  - Authenticate/evaluate claims to establish trust
  - NB: claims should be related to transactional needs (cf., credit card)
- Caveat: reputation ≠ trustworthy
  - confidence interval for “scores” validating claim (FICO, ID,etc.)

Three types of data attribution (identification as a technical matter)

- Individual authentication (~identification) bio, multifactor, etc.
  - *Confidence* that an identifier refers to a specific individual
- Identity authentication (~indexing) (>entity resolution, <DM)
  - *Confidence* that an identifier refers to an identity
  - (Use of SSN, DOB and mother’s maiden name)
  - (Persistent across observations)
  - Misuse of identifiers has led to identity theft on the one hand and restrictions on use of identifiers (less security) on the other side
- Attribute authentication (~authorization) (information retrieval)
  - *Confidence* that an attribute applies to a specific individual
  - ~ “reputation” attributes and their use to determine trust

Privacy

- What is “privacy” (right to conspire in secret?)
  - *Secrecy* - keep data unknown
  - *Anonymity* - keep data unattributed
  - *Autonomy* - keep data from constraining opportunity
- Hierarchy of legitimate privacy concerns (*for the “innocent*”)
  - Identity theft or appropriation (privacy=identity theft)
  - Intrusion (*social*) (Dyson-fashion; Smith-online cc; security clearance)
  - Compartment breach/secondary purposes (autonomy trap)
  - Information asymmetry (loss of bargaining power)
  - Other

Underlying problem with existing models: who "owns" personal information?

- First problem is the rhetoric of ownership itself doesn't fit (or is contrapose depending on POV)
  - "rivalrous-ity", "excludability"
  - information wants to be free (zero marginal cost of distribution - IP)
  - brittleness of secrecy -- no cure for making something unsecret
- Subject ("owns" identity -- R?)
  - Subject to theft
  - Issues of control (selective or compartmentalized disclosure)
- Collector/aggregator ("owns" reputation -- Non-R?)
  - Non-excludable knowledge (can't teach to others?)
  - Cf. "autonomy trap"
  - No "dossier" problem (future of data fusion) (sharing - ISE ***)
- Reconcile: "joint custody" and shared responsibility of authoritative ID

IT and new information efficiencies

- Data is no longer transient (always available)
- Data is proximate (available anywhere)
- Leading to transitional business models based on collection and aggregation to make data available for use in risk management (~services)
- Results in an *end to "practical obscurity*" of data by virtue of its physical location and an end to anonymity through data transience
- This challenges traditional notions of privacy based on inefficiencies or high cost of access

New information economics and data availability

- The cost of *data retention* is less than the cost of *selective deletion* (~regulatory effect)
- The cost of indiscriminate *data collection* is less than the cost of *selective acquisition* (e.g., Echelon vs Carnivore)
- Data is increasingly *produced in digitized form*
  - Ultimately, this will undermine business models premised on investment in collection except in specialized niches
  - Technical/digital means of collection are capital intensive not labor intensive, thus *cost per unit of information* will decrease to point of commodification
- Thus, data largely “exists” and value creation will be in services (authentication, verification, attestation) (brand strategy?)

Current (flawed) business model

- Value of PI data:
  - Data subject trades PI at *marginal cost* of initial transaction
  - Aggregators invest at *aggregate cost* of collection
  - Users purchase at *utility cost* for subsequent transaction
- Doesn’t account for externalities
  - Subjects bear concentrated harms of errors/breaches but receive only diffuse benefits
  - Aggregators bear little cost of errors/breaches on either side and receive concentrated benefits ** (~PR hit)
  - Users bear diffuse harms of errors (acceptable below some critical value?) and from differences between salience for initial transaction and for subsequent use (cf. marketing, credit, national security)

Current remediation efforts

- Force aggregators to internalize costs:
  - Data security requirements (obviously, but use liability not regulation)
  - Disclosure/notice (~PR harm)
  - Limit their business model based on enforced *secrecy*
    - Inconvenience or limit customers (e.g., large customers over small) (approved customers) (source location) (two-factor)
    - Constrain growth and business opportunities for additional services on data infrastructure (for entire economy)
  - Current regulatory efforts limit competition and reinforce concentration of asymmetries

Current remediation efforts doomed to failure

- BECAUSE -- aggregators may not be optimal mitigators (not lowest cost avoiders?)
  - For data quality -- data subject has best incentive/kn and new business models can enlist the power of distributed contribution and collective intelligence
  - For transaction fraud -- data user is lowest cost avoider (authenticate transaction)
- AND, protecting privacy through secrecy is the cause of identity theft (can’t both exchange identifiers and keep them secret)
- Thus, alternative solution …

Solution: *Identity registrar* to eliminate identity theft and protect privacy

- Government registry (FTC?) (voluntary system)
  - Make your name, SSN, and certain protected contact information publicly available -- including a designated “identity broker”
  - Service providers (credit issuers, etc.) who use the system retain their current exemption from legal liability for misidentification
  - Service providers who did not use the system would be liable for misidentification (defamation, invasion of privacy, negligence)
- “Consumer driven”
  - 60-70% consumer initiated transactions
  - OK, here is consumer driven model
- “More information” - annotation rather than restriction

Identity registrar (cont.)

- Identity Broker would be authorized repository for authentication and could offer varying levels of service --
  - automated aggregation (“google me”)
  - automated aggregation with annotation (“zoominfo” me)
  - fully verified and attested (“Verified ID [Brill] me”)
  - ??? “LexisNexis me” “ChoicePoint me”, “Equifax me” etc. ???
- Multiple identities for specific purposes. Brokers could be employers, unions, banks, data aggregators, etc.
- Controlled contact (cut-out, filter) (verified counterparty, call back, etc.)
- Subject (~ind) designated verification (and disclosure) process
  - ~fraud alert (90 days, 7 years if you prove you are victim)
  - ~discrepancy notices/flags / file freeze
  - Opt out (privacy) (disclose x info for y purpose) (notice vs easy)
  - Can be extended to any (or all) transactions
- Required legislation: liability issue, prohibit SSN as password

Information society II

- Web 2.0
  - Infrastructure and platform as commodities
  - Leverage customer self-service
    - reach out to edge and service long-tail
    - gets better the more its used
  - Harness collective intelligence
    - capture and make part of the value chain
    - Network effect of user contributions
    - Wisdom of crowds - architecture of participation
  - Branding data
    - authenticate, validate, verify, attest
    - Little value in collection or aggregation
  - Loosely coupled services designed for customer “re-mixing”

Information society II (cont.)

- Identity 2.0
  - Identity 1.0 was silo-ed, single entity user account
  - Identity 2.0 is portable, comprehensive, selectively disclosable
- Identity Commons
  - develop the framework for an open global trust network in which individuals and organizations own and have control over their identifiers and data profiles, and in which identity information can be exchanged and used in a secure trusted environment through identity broker
- Identity drop down menu on browser
  - Multiple pseudonyms
  - One time authenticators (~Amex)

Conclusions?

(The Transparent Society, 1999)

There is not a crime, there is not a dodge,
there is not a trick, there is not a swindle,
there is not a vice which does not live by secrecy.

Joseph Pulitzer

Whenever a conflict arises between privacy and accountability, people demand the former for themselves and the latter for everybody else.

David Brin