Legal Aikido: A Data-Sharing Framework to Advance Network & Security Research

Erin Kenneally, M.F.S., J.D.
Erin @ caida.org

Cooperative Association for Internet Data Analysis
University of California San Diego

Talk Map

- Defining the Issue & Solution Space
- Challenges & Motivations
  - Uncertain Legal Regime
  - Incomplete Technology Solution Models
  - Data Risks
  - Under-valued Benefits of NetSec Research
- Applying Akido:
  - Self-Reg Opportunity
  - Operational model: PS2 Framework

The Issue Space Defining the Solution

- Current posture:
  - defensive, default-deny sharing network traffic data
- (Misinformed) assumptions:
  - Privacy risks and legal restrictions >>> benefits of sharing
  - Unprecedented data availability = plethora of network infrastructure information
  - ISE directives post-911 → incent network data exchange
- Muted legislative, judicial, policy drivers
  - Threat model from NOT sharing data = vague
  - No body count / $billion losses (at least no explicit, causal)
- No widespread, standard procedures for exchange
  - Ad-hoc, nod & wink
- Dynamic and normative-deficient understanding of privacy risk and research utility
  - No cost-accounting for privacy risk
  - No ROI for investment in empirical network measurement

- **confusion = window of opportunity**

Challenges & Motivations
(1) Uncertain Legal Regime

- No legal framework explicitly prescribes, incentivizes, or forbids sharing network data for security research

- Ambiguity between tech & legal discourse re: fundamental concepts driving risk
  - PII, REP, content, URLs, IPAs, packet headers, payload … oh my!
  - Law inconsistent- functional equivalent of PII
  - E.g., is IPA 'content' and URL 'addressing' data (ECPA, 4th A. purposes)?
    - *Johnson v. Micro*soft (2008) - IPA does not identify persons
    - *State v. Reid* (2007) - REP in subscriber information attached to IPA
    - *US v. Forrester* (2007) - URLs may have REP - reveal communication content
    - *HIPAA Privacy Rule* – IPA is protected PII
    - States' data breach laws – IPA not in definition of personal information

- Social normative expectations: my IPA, URLs + search terms are digital fingerprints?
  - E.g. Tor, automated in-browser cookie and URL deletion

Challenges & Motivations
(2) Incomplete Technology Solution Models

- Point solutions fail to address context-dependent risks
  - Cases-in-point: de-anonymization attacks success
    - Prefix-preserving anonymization subject to re-identification
    - Poster cases (Netflix, Yahoo!, Traffic injection attacks)

- Trade-off: Purely technical approaches v. research utility goals
  - Data minimization intentionally obfuscate essential data (network management, countering security threats, evaluating algorithms, apps, architectures)
  - E.g., Conficker

Challenges & Motivations
(3) Data Sensitivity Risks

- Sources: legal compliance, ethical obligations, norms/court of public opinion
- Main categories
  - Disclosure risk
    - Public
    - Accidental/malicious
    - Compelled – Private (RIAA subpoenas) , Gov't (NSA, Telco releases)
  - Misuse risk
    * increasing quant & qual - Evolving tech > capabilities and < costs:
    - False inference (match linking 1st/2nd order identifiers)
    - Data confidentiality (network topology, health)
    - Privacy
      - linking network data to individuals
      - de-anon / re-identification commoditization
        - Tension- protect (law, policy) AND motivation to uncover PII (profit, avoid legal liability triggers, attribution)
        - Cat & mouse gamers = LE investigations, biz intel, legal dispute resolution, security incident response, gov't infrastructure protection

Challenges & Motivations
(4) Under-valued Benefits of Network Research

- Justify the Ask:
  - Understanding structure, function of CI
  - (topology, workload, traffic routing, performance, threats & vulnerabilities)

- Network Data sharing utility criteria
  - Objective is positively related to social welfare
  - Need for empirical research
  - Research purpose not being pursued
  - Research could not be conducted without
  - No sufficiently similar data already being collected that could be shared
  - Uses of shared data are transparent, objective, scientific, control for risk
  - Results using shared data can be acted upon meaningfully
  - Results are capable of being integrated into ops or biz processes (security improvements, situational awareness)

Solution Space: Using Aikido on Net Sec Researcher's BFF

- Options:
  - Amend law (ECPA research exception)
  - Aikido the law via self-reg regime (i.e. inform norms & legal precedent)

- /1st/ **ID the attack** – on voluntary disclosure of non-content to researchers
    - ECPA's invariables- From Whom, To Whom, What, Where
    - ECPA's variables- Consent, Provider Exception & Relevance, Gov't, Content

- /2nd/ **Blend & harmlessly redirect attack** - use the law itself to clarify the gray and embolden the exceptions
  - **Consent Exception**
    - interpreted broadly, esp if internal to Provider, so define for network performance engineering and research
    - unclear USE scope, so define specific uses viz. ToS & Privacy Policies, banners
  - **Provider Exception** (outsourcing research under cyber security justification)
    - ECPA allows EE of an ECS to "intercept, disclose, or use" communications when such activity "is a necessary incident to the rendition of his service or to the protection of the rights or property of the provider of that service."
    - ECPA does not, per se, prohibit outsourcing research
    - Internal use is largely unregulated /eg/ telcos and SE retention of traffic
    - necessary incident..protect rts & property- largely untested
    - So → define "research" explicitly, make nexus between collection and sharing for protection of rts/ property substantially clear; ensure that provider procedures are consistent with one of the statute's exceptions; SCA no apply if sharing occurs w/i service provider (ie, researcher employed by provider)
  - **Provider Relevance**-
    - Non-public provider may disclose non-content records to a governmental entity (state university researcher); or
  - **Gov't Entity**-
    - ECPA no define, unclear if State U. Researcher is "gov't"
    - Press whether public sector entity must have compulsory powers to = "gov't"
  - **Content** –
    - Does ECPA apply data if data anonymized beyond being recognizd as "content"? (substance, purport or meaning)

Implementing Akido

- **Privacy-Sensitive Sharing (PS2) model**

  = Privacy-enhancing technology + privacy-principled policies

- Risk – Benefit methodology
  - Bridges risk – utility perception gap
- Enables transparency as touchstone of data sharing
  - counter to subjective, opaque evaluations
  - engender trust, beyond "trust me"
- Considers practical challenges of stakeholders
  - network researchers, sys operators, security professionals, legal advisors, policymakers
- Bottom-up & Proactive
  - Anchor point to demo community norms, inform law & policy

PS2 Framework
Policy Components

- Core underpinnings:
  - Make risks 'contagious' (sharing= data AND responsibilities & obligations)
  - Components rooted in principles and practices of national & global laws, policies

1. Authorization
2. Transparency
3. Compliance with applicable laws
4. Purpose adherence
5. Access limitations
6. Use specifications and limitations
7. Redress mechanisms
8. Oversight
9. Security
10. Audit tools
11. Data quality assurances
12. Training
13. Transfer to 3rd parties
14. Ethical impact assessment
15. Disclosure minimization

PS2 Framework
Technology Component

- **Disclosure Minimization/Controls**
  a) Deleting all / part of sensitive data
  b) Generalizing
  c) Perturbing
  d) Pseudonomizing all/ parts
  e) Aggregation or sampling techniques
  f) Mediation techniques (sending code-to-data)
  g) Aging the data
  h) Limiting quantity
  i) Synthesizing
  j) Layering anonymization

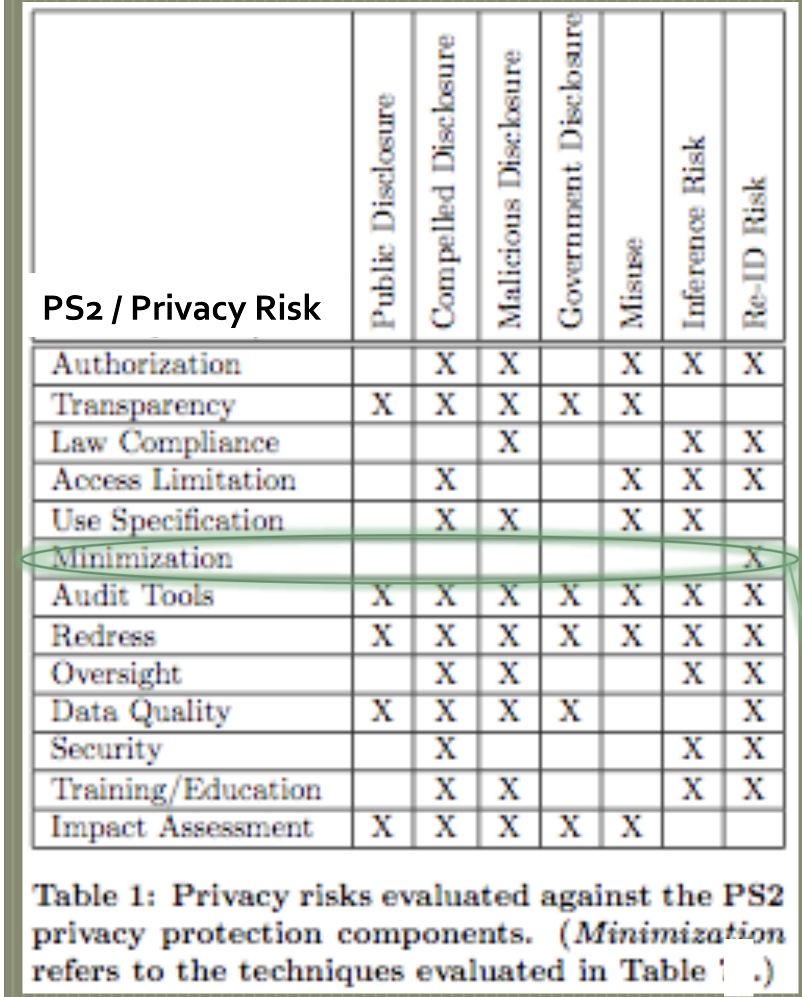
- **Implementing Vehicles :**
  - MOU/MOA/MOC, model K's, binding organizational policy, NDA, AUP

Evaluating PS2
Addressing Data Risk & Utility Goals

- **Criteria**:
  - 1. How well PS2 addresses data risks (table 1)
    - Policy controls, alone = coverage gaps
    - Tech controls, alone = seemingly control for privacy risks (? policy control components superfluous ?)
  - 2. To what extent PS2 impedes utility goal (table 2)
    - Technical controls, alone = impedes utility

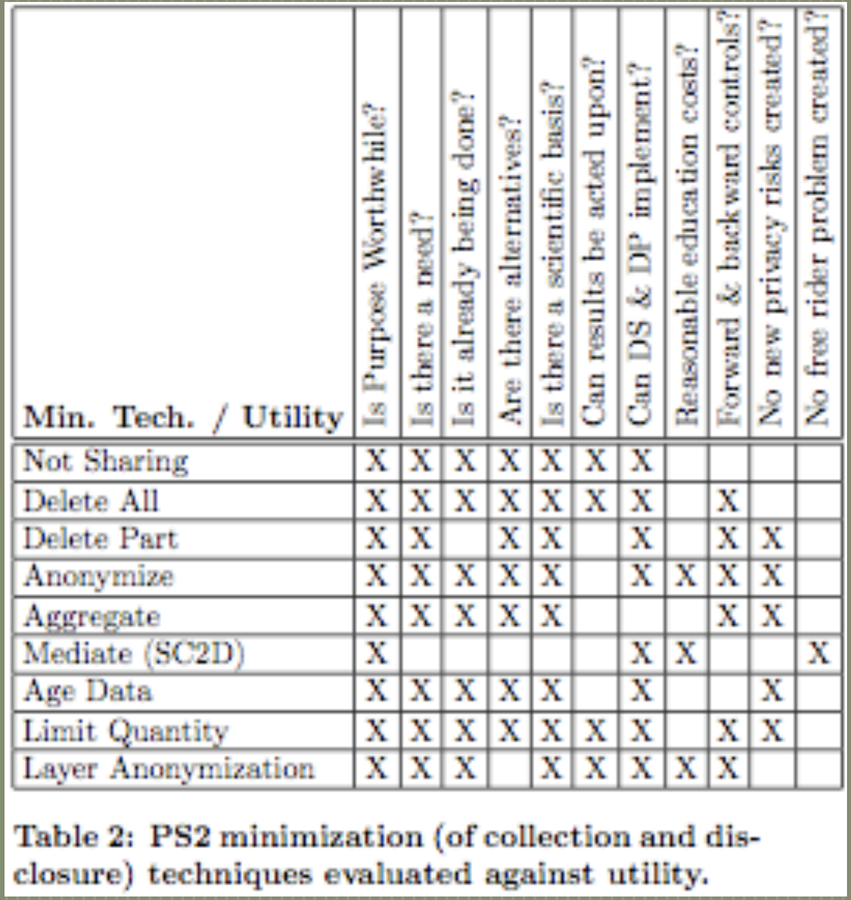
- **Conclusion**:
  - Only tech solution breaks down along utility dimension
  - Only policy solution leaves too high privacy risk exposure
  - Therefore, hybrid strategy - dial down tech controls for utility objectives AND dial up policy controls to cover risk
  - Framework is both
    - Evaluation of hybrid model
    - Policy Dev tool for data sharing

Evaluating PS2
Addressing Privacy Risk & Utility Goals



PS2 / Privacy Risk	Public Disclosure	Compelled Disclosure	Malicious Disclosure	Government Disclosure	Misuse	Inference Risk	Re-ID Risk
Authorization		X	X		X	X	X
Transparency	X	X	X	X	X		
Law Compliance			X			X	X
Access Limitation		X			X	X	X
Use Specification		X	X		X	X	
Minimization							X
Audit Tools	X	X	X	X	X	X	X
Redress	X	X	X	X	X	X	X
Oversight		X	X			X	X
Data Quality	X	X	X	X			X
Security		X				X	X
Training/Education		X	X			X	X
Impact Assessment	X	X	X	X	X		

Table 1: Privacy risks evaluated against the PS2 privacy protection components. (*Minimization* refers to the techniques evaluated in Table [illegible].)



Min. Tech. / Utility	Is Purpose Worthwhile?	Is there a need?	Is it already being done?	Are there alternatives?	Is there a scientific basis?	Can results be acted upon?	Can DS & DP implement?	Reasonable education costs?	Forward & backward controls?	No new privacy risks created?	No free rider problem created?
Not Sharing	X	X	X	X	X	X	X				
Delete All	X	X	X	X	X	X	X		X		
Delete Part	X	X		X	X		X		X	X	
Anonymize	X	X	X	X	X		X	X	X	X	
Aggregate	X	X	X	X	X				X	X	
Mediate (SC2D)	X						X	X			X
Age Data	X	X	X	X	X		X			X	
Limit Quantity	X	X	X	X	X	X	X		X	X	
Layer Anonymization	X	X	X		X	X	X	X	X		

Table 2: PS2 minimization (of collection and disclosure) techniques evaluated against utility.