Can Network Science Help Re-Write the Privacy Playbook?

Erin Kenneally, M.F.S., J.D.
CAIDA| Elchemy

W3C Data Usage & Control Workshop
MIT | 6 Oct 2010

Gameplan

- Incumbent playbook
- Problems with playbook
- Playbook fractures exposed
- Evolved playbook: Scale-free Privacy 101
- Validating the new playbook
- Operationalizing the new playbook

- Definition
  - PIA = personal information artifact
  - PC = PIA controller
  - REP = reasonable expectation of privacy
  - Control = law, regulation, policy, standard, contract

TAKEAWAY

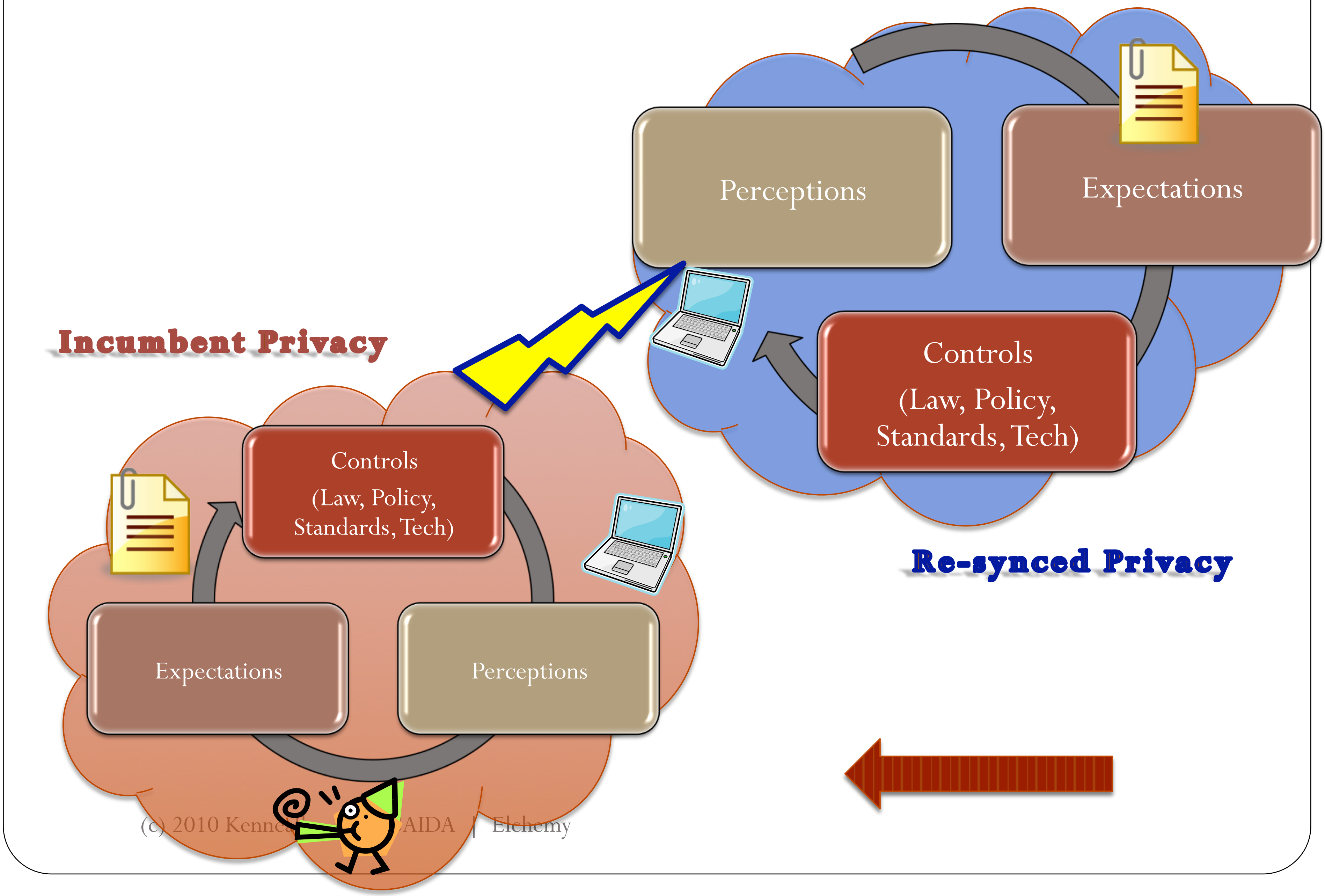
- Privacy inflection point
- Cognitive dissonance over its meaning and measurement
- Need to re-sync 3-legged stool
  - Perceptions → Expectations → Controls
- Can network science enable this phase shift?

NETWORK SCIENCE CAN DESCRIBE PRIVACY EXPECTATIONS & RISKS AS A SCALE-FREE NETWORK …

To what end?

MORE EMPIRICALLY DESCRIBE REASONABLE EXPECTATIONS OF PRIVACY AND APPLY PRIVACY CONTROLS

Re-Syncing Expectations with Controls



Perceptions

Expectations

Controls

(Law, Policy, Standards, Tech)

Incumbent Privacy

Controls

(Law, Policy, Standards, Tech)

Expectations

Perceptions

Re-synced Privacy

Incumbent Playbook

- Genl purpose of privacy controls - balance competing interests
- REP principle underpins many privacy controls
  - 4th A.: subj & obj. EOP
  - Tort: obj EOP via consent & control elements
  - K: “public” info exceptions in NDAs
  - FOIA
  - Industry self-regulations/best practices
  - Civil discovery rules
- REP draws boundaries (implemented often via public-private doctrine)

- Mechanisms for proving (current)
  - Public opinion/survey
  - Observational data
- We’ve got issues: What is REP/Public–Private in network playing field?
  - Offline = Visible to public; communicated to public; occur in public
  - Online = boundary sentience very different

Problems with Current Playbook:

- Incumbent REP presumes a scaled network model contoured around privacy perceptions
- But, privacy in networked context is different in perceived risks and threats, and resembles a scale-free network
- So what?
  - incongruous awareness and protection of rights
  - circular paradigm: privacy controls apply REP by what is deemed "private", vice versa, but what does that mean in network playing field?

Why We Need New Privacy Playbook

Network Playing Field

- PIA dynamic, temporary
- PC differentiated
- Relationships between PC matter
- Disclosures carry different relative risks

- Privacy threat model:
  - < awareness & understanding of technology underpinning PIA location and movement
  - PIA is continuous, privacy choices more intricate
  - Referential boundaries (virtual) : privacy risk more opaque

Offline Playing Field

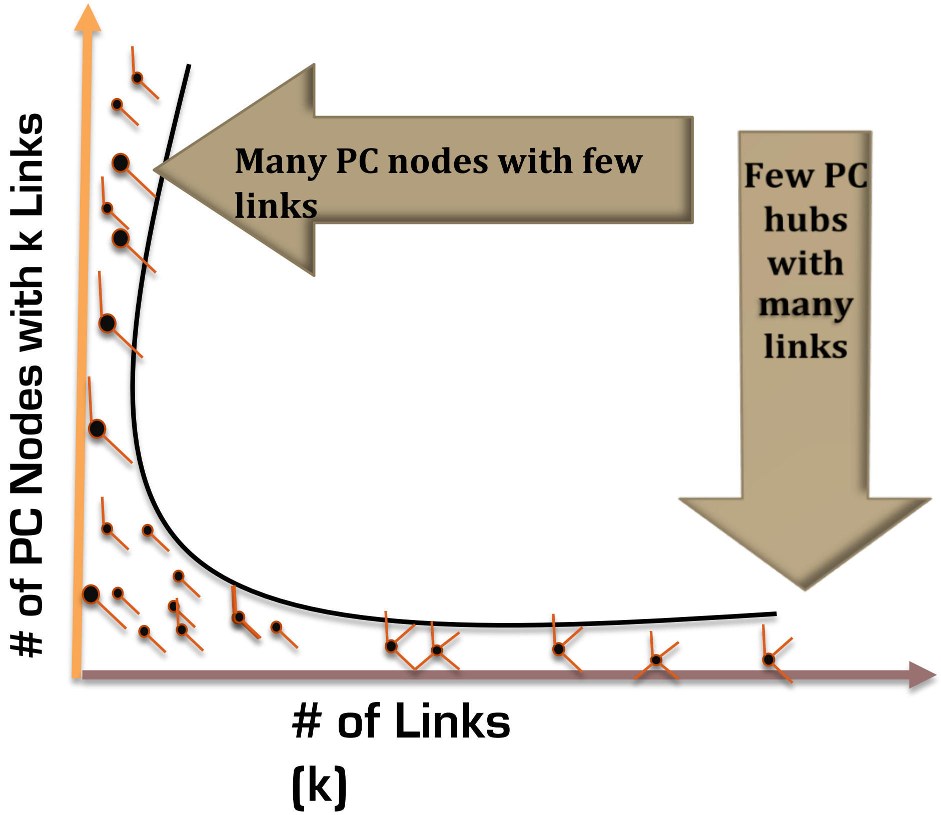
- PIA static & ~permanent
- PIA controllers (PC) equivalent
- Unit of risk was PIA itself
- PIA disclosures to all 3rd parties ~identical

- Privacy threat model:
  - Knowledge of PIA ~ known
  - Privacy-relevant data discrete & linear
  - Boundaries that inherently define privacy sentient : Privacy risks ~ transparent

Playbook Fractures Manifest

- **Industry Self-Reg / 'standards'**
  - Notice & consent inadequate
  - Too coarse
  - Capability ≠ actuality
  - "Partner" catch-all (LBS, advertiser, app developer, ___)
  - 'Trust-Us' privacy policy is a shill
  - Awareness & enforcement challenges
- **Location-based surveillance**
  - 3 US App. Cts split
  - public movement ≠ no REP; public movement across time = REP (?)
- **Google Streetview**
  - 8 class actions claiming privacy violations
  - Unencrypted data from unsecured network routers = REP(?)
  - ECPA no prohibit collection of data from networks "accessible to the public"
- **Social Networking data**
  - Is wall posting public? REP?
  - Crispin crt remand to determine if privacy settings render messages public and outside stored communication protections
- **FOIA & exceptions**
  - anonymized PIA that can be re-identified = REP(?)
  - No exempt data found on DL, but, what if same data in Internet ecosystem

Modeling Privacy As Scale-Free Network


Many PC nodes with few links
Few PC hubs with many links
# of PC Nodes with k Links
# of Links (k)

Albert-Laszlo Barabasi; http://www.macs.hw.ac.uk/~pdw/topology/

- 1. Distribution of nodes approximates a power law → few nodes have many links (aka, hubs) and most nodes have few links.
- 2. Network evolves and is dynamic → nodes added & removed throughout time.
- 3. Links exhibit preferential attachment ('the rich get richer') → new links added to nodes based # of existing links or node fitness.

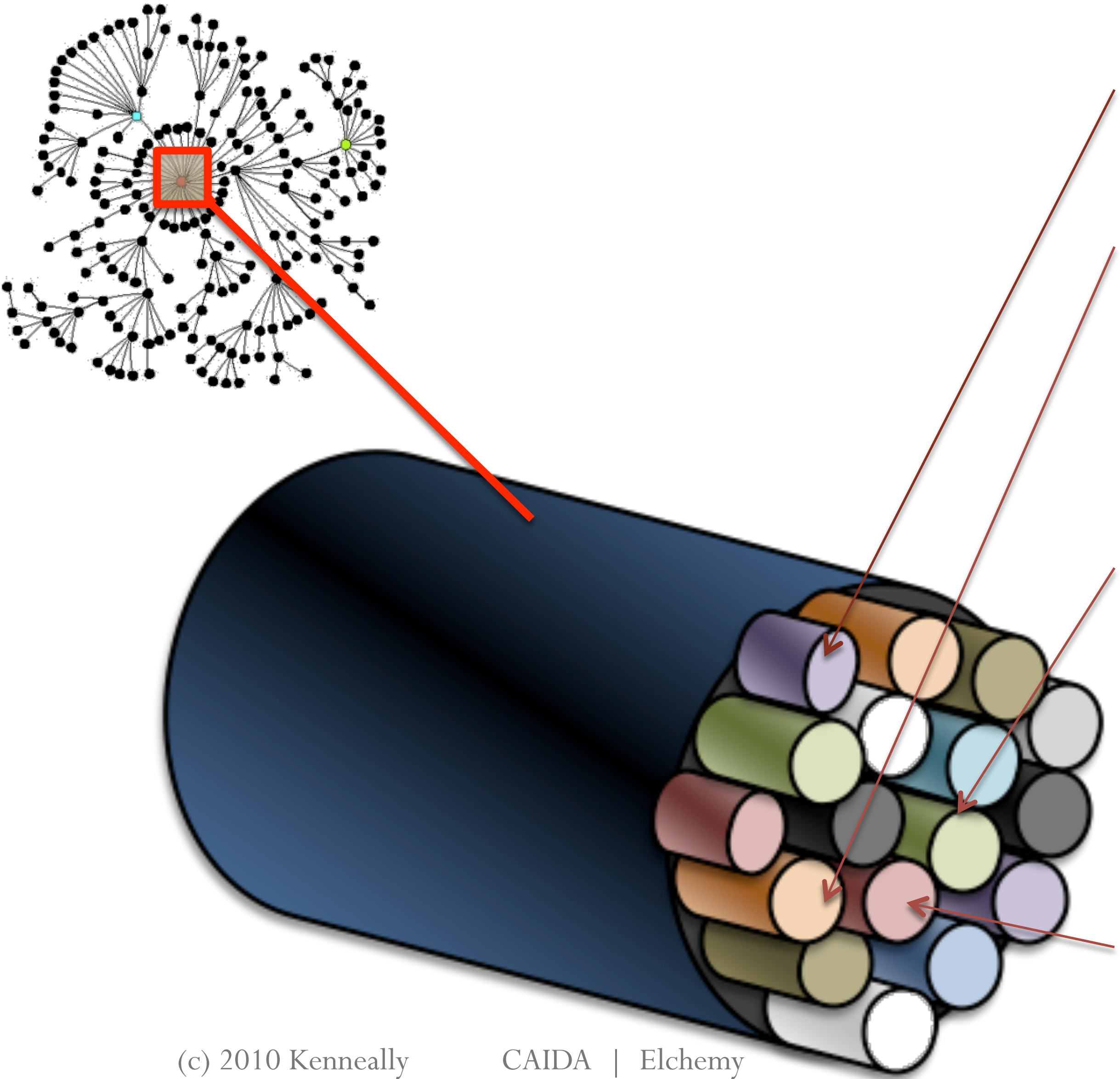
Validating the New Playbook

- Is information privacy a scale-free network?
- **Is PIA network structure and relationships (flow dynamics) similar to commodities?**
  - If so, what does it mean for describing and prescribing REP?
  - E.g., what are the possible normative implications for information privacy law, such as whether PIA exposure to 3rd parties is a de facto poor indicator of greater threat to privacy?
  - How might knowledge of PIA flows either eliminate the use of public-private standard for measuring REP; or, can it be used to re-define what we mean by public-private space with a fidelity that is more aligned with the reality of information flows?
  - How well are certain PC integrated with the whole system, such as data aggregators or online advertising networks?
  - How closely does the geo-location of PC hubs correspond to traditional public-private and 3rd party doctrines?
- **How should we apply a scale-free model to privacy controls?**
  - E.g., does knowing how PC ages enhance our understanding of how privacy evolves with time?
  - Can the PC churn rate help us understand how quickly PC accumulate links and determine the rate of collection/disclosure of PIA?
  - Should the size of PC clusters and their proliferation establish living REP or indicate failure of privacy controls?
- **Is there congruence between collection/disclosure topology and the semantic topology of PIA?**
  - E.g., do the clusters of PC link based on shared meaning of the value of a particular PIA for price discrimination or some other economic use?

? *Empiricizing* Scale-Free REP ?

- 1) Node Fitness
- 2) Structure of the PIA network (links)
- 3) PIA content
  - behavior, location, health, physical, financial, communication, other data
- 4) Relationships between PCs

What Might PC Node Fitness Mean?



* Purpose of collection (functional, advertising)

•Subject’s awareness of C/U/D

• Optional or compulsory collection

•Identify or verify

• C/U/D time: fixed or indefinite

•Where, how long PIA stored

• Who possesses the PIA

•Who accesses the PIA

• What are disclosure restrictions

•Security of PIA storage

• Security of PIA format

•Security of PIA transmission

• Type of analysis done on PIA (eg, mathematical, interpretive/inference-laden)

•Derived or original

• Sensitivity to cultural constraints (moral, legal constraints)

Operationalizing Scale-free Privacy Playbook:

- **Inform evidence-based policymaking –**
  - ensure that choice and control of the c/u/d of PIA is based on empirical reality of how it flows throughout networks;
  - inform **default privacy presumptions** for efficient K rules, e.g., should we impose implied nondisclosure obligations on certain PC for certain categories PIA? Or, should privacy settings or ToS establish default REP in web communications?
  - Can knowing structure and dynamics help **traceback derivative data to origins** in privacy/data protection litigation? Understand **match-link risks** for data protection standards (e.g., HIPPA standards for anonymization)
- Enable **better privacy risk management** for both individuals asserting privacy rights and entities handling PIA – the entities with countervailing interests—through more predictable outcomes, more certainty about REP determinations, and lower liability risk.
- Advocate **common definitional semantics** to harmonize reasonable expectations across privacy controls-
  - industry-specific and data-specific laws,
  - geopolitical authorities responsible for enforcing privacy controls
  - between and among privacy self-regulated industries.
- **Refute or validate** non-institutionalized intuitions about REP **norms.**
- Devise more **sophisticated justifications for our intuitions** about privacy (e.g., autonomy, seclusion, property).

Questions & Answers Welcome

Erin Kenneally

erin@elchemy.org

erin@caida.org