Illuminating the Way for Trusted Darkspace Data Sharing

Erin Kenneally

CAIDA, Elchemy

DUST 2012
Workshop on Darkspace &
UnSolicited Traffic Analysis
La Hoya, CA
14 May 2012

Practical Disclosure Guidelines: CHALLENGES

- Difficulty bounding attack risk
  - New inference attacks being developed
  - Access to secondary data sources
  - Privacy definitions immature for network data
- Massively heterogeneous data
  - Hundreds of protocols and new ones being developed
  - Corner cases and implementation differences
- Interactions between policy and technology
  - Multiple types of policy risks, control technologies, data formats, access methods, etc.
  - Different levels of risk tolerance and strength of controls
  - Exponential number of unique scenarios to cover

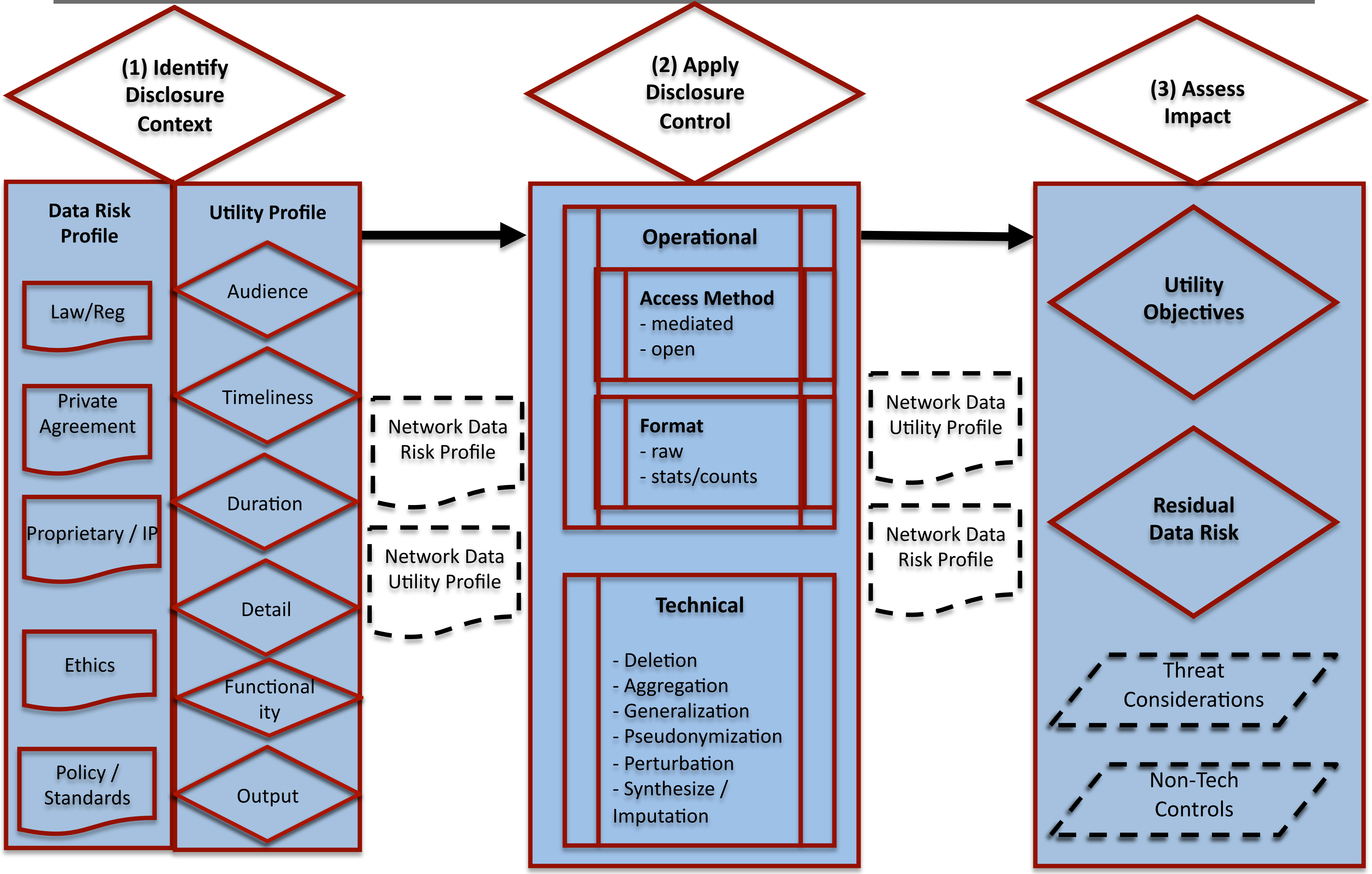
Framework for Sharing Network & Security Data

- **Purpose**: Develop reference framework for applying disclosure control technologies to network data
  - guidance on technical controls
  - enable risk-sensitive data sharing
  - consider legal constraints and utility needs
  - for both data producers and consumers
- **Audience**: researchers, analysts, operators, policymakers
- **History**: Kick-off workshop (spring 10'); Advisory workshop (spring 11'); SME workshop (winter 11'); First draft (end of May 12')
- Co-lead: Scott Coull, Redjack

It's Elementary ≠ Easy

- knowing what you want to do (utility)
- knowing what you can't do (risks)
- how to have your cake & eat it too (disclosure techniques)
- knowing who you want to play with (trust in the data recipient)

Disclosure Framework



(1) Identify Disclosure Context

Data Risk Profile

Law/Reg

Private Agreement

Proprietary / IP

Ethics

Policy / Standards

Utility Profile

Audience

Timeliness

Duration

Detail

Functional ity

Output

Network Data Risk Profile

Network Data Utility Profile

(2) Apply Disclosure Control

Operational

Access Method
- mediated
- open

Format
- raw
- stats/counts

Technical

- Deletion
- Aggregation
- Generalization
- Pseudonymization
- Perturbation
- Synthesize / Imputation

Network Data Utility Profile

Network Data Risk Profile

(3) Assess Impact

Utility Objectives

Residual Data Risk

Threat Considerations

Non-Tech Controls

Discussion: Decision Drivers

- **Discussion Goals:**
  - Gather more real-world challenges faced by data providers
  - Concrete needs of providers in terms of policy and technical guidance
  - Feedback on initial reference framework

- **(1) What are the major factors in your decision to collect and share network data?**

Risk IQ

- (2) Do you feel like you have a strong understanding of the risks (legal, contractual, etc.) of sharing network data?

- (3) Do you feel like you have a strong understanding of the available controls for mitigating those risks (both technical and policy)?

Incentives / Motivations

- (4) What (if anything) would motivate you to collect and share more network data with the research and operational community?
  - (a) Better understanding of best practices (both technical and policy)?
  - (b) A community-driven best practices document? If not a document, what form (if any) should this take?
  - (c) How detailed should the guideline be? Are general categories helpful or is it better to dive into specific implementation details?
  - (d) Would you expect the guidelines to provide a quantifiable risk score, or is general discussion of the concepts sufficient (remember: quantifiable risk assessment approach is not guaranteed to be correct)?

Components

- (5) What should a Best Practices Guide include to improve data sharing?
  - (a) Description of policy risks (e.g. laws, contracts, and ethical guidelines)?
  - (b) Description of intended utility objectives (e.g. publicly available research, private operational release)?
  - (c) Description of available disclosure controls, their benefits, and potential pitfalls? Technical? Policy?
  - (d) Description of threat considerations and how they impact how well disclosure controls will work?
  - (e) What are we missing?