A Qualitative Risk Assessment Framework for Sharing Computer Network Data

Scott E. Coull
RedJack

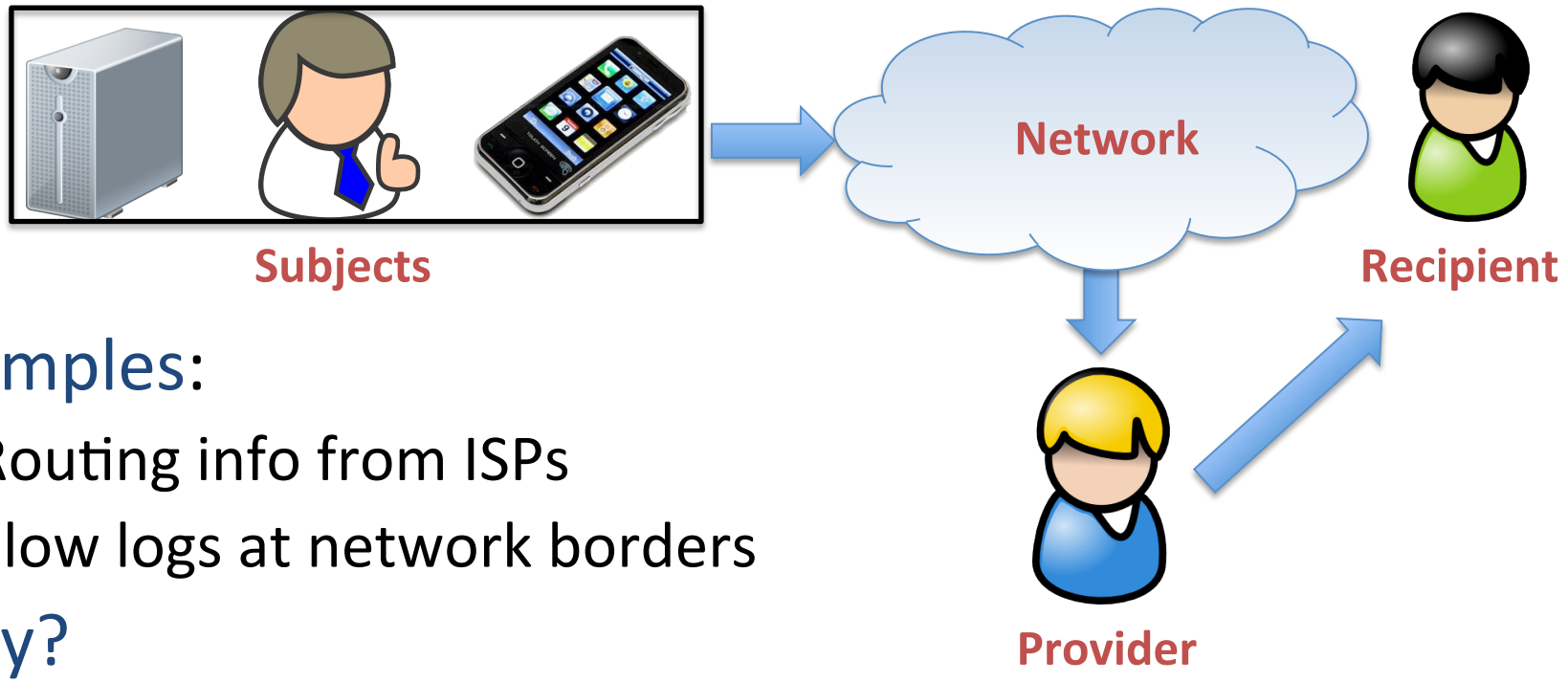
Erin Kenneally
Elchemy

23 September 2012

Game Plan

- What is Network Data Sharing?
- Distinct Challenges
- Reality of Data Sharing
- Risk Assessment Approach
  - Framework Overview
  - Case Study

Setting the Stage



- Examples:
  - Routing info from ISPs
  - Flow logs at network borders
- Why?
  - Improved innovation from real-world data
  - Faster response to network incidents
  - Data-driven policies

Pitfalls and New Trails

- Data complexity
  - Massively heterogeneous data
  - Huge volume of data
  - Many different types of actors
- Difficulty bounding attack risk
  - Cannot quantify access to secondary data sources
  - Privacy definitions are immature for network data
- Interactions between policy and technology
  - Not just PII → intellectual property, network security, etc.
  - Lack of legal precedent or guidance for network data

Reality of Data Sharing

- Uncertainty of legal risk
- Understated value of potential benefits
- One-size-fits-all approach to disclosure controls
- Implicit assumption that **any** sharing increases risk

- Results in:
  - Data rich vs. data poor
  - Sharing through ad-hoc, interpersonal relationships
  - Scarcity of scalable, transparent, sustainable sharing

Moving Forward…

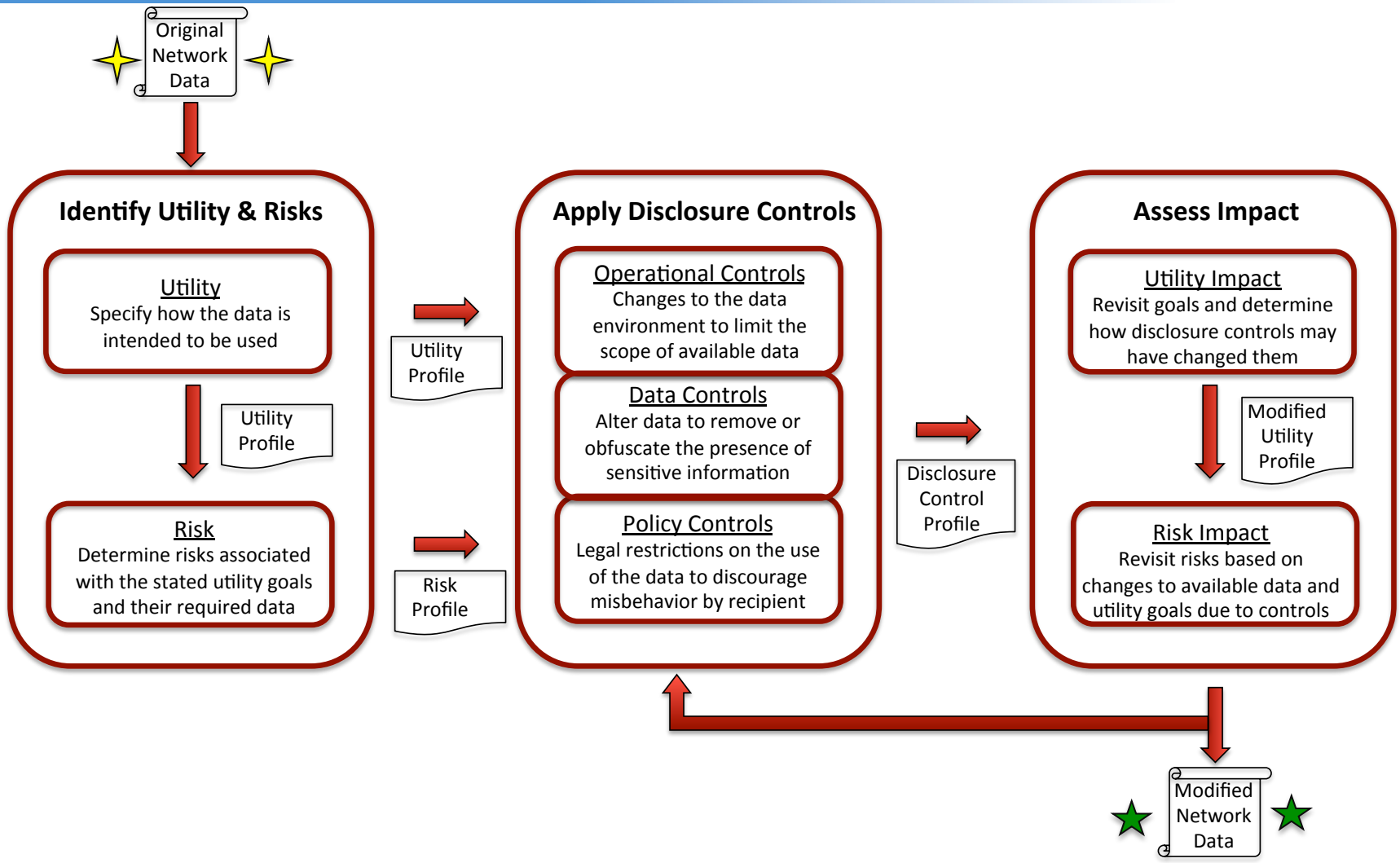
- Qualitative framework for:
  1. Identifying specific utility goals and related risks
  2. Choosing disclosure controls to address risks
  3. Assessing effects of those controls
- Generalizable across all network data & scenarios
- Enable data providers to:
  - Better understand sources of risk
  - Tailor controls to intended utility
  - Justify choices and explicitly state assumptions
- Promote the social value of shared data & process

Between the Cracks

- Does not provide yes/no answers
  - Data sharing is a risk management process
  - Appetite for risk varies significantly

- Attacks may exist or information may be leaked
  - Understand what risks exist
  - Justify disclosure control choices

In a Nutshell

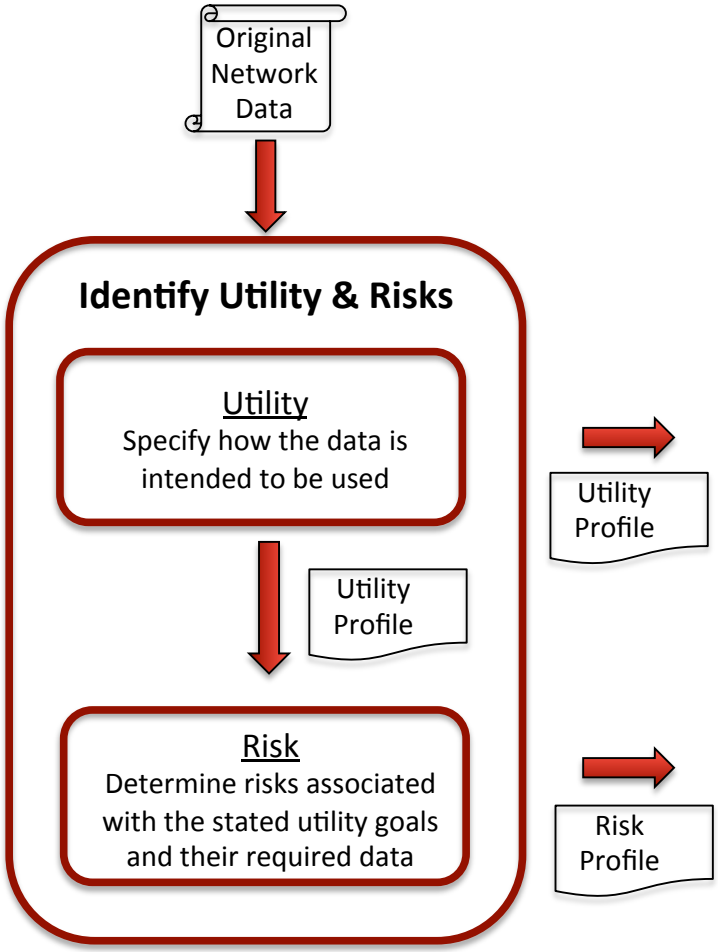


Framework in Action: DNSChanger

- DNSChanger malware:
  - Created by Estonian company Rove Digital
  - Redirected user DNS queries to malicious servers
  - Internet Systems Consortium (ISC) operated replacement servers to provide continuity to victims

- Opportunity to collect and share data about 800,000 real-world DNS clients!!

Where are we?


Original Network Data
Identify Utility & Risks
Utility
Specify how the data is intended to be used
Utility Profile
Utility Profile
Risk
Determine risks associated with the stated utility goals and their required data
Risk Profile

Utility and Risks

- Possible high-level goals:
  1. Understanding DNSChanger infection properties
  2. Analysis of general malware behaviors
  3. General DNS traffic modeling

- Translates to:
  - Infection properties → client IP address
  - Malware behavior → DNS data
  - DNS traffic modeling → DNS & client IP

Utility Requirements

Category	Score	Justification
Audience	3	Access by legitimate security and networking researchers only
Timeliness	5	Research does not require immediate access, data is useful for long period of time
Duration	2	Research studies require long-term access to data
Detail	1	DNS modeling and analysis requires traffic contents and fine-grain client info
Functionality	2	General DNS-related research studies
Output	1	Publication of research findings gleaned from data

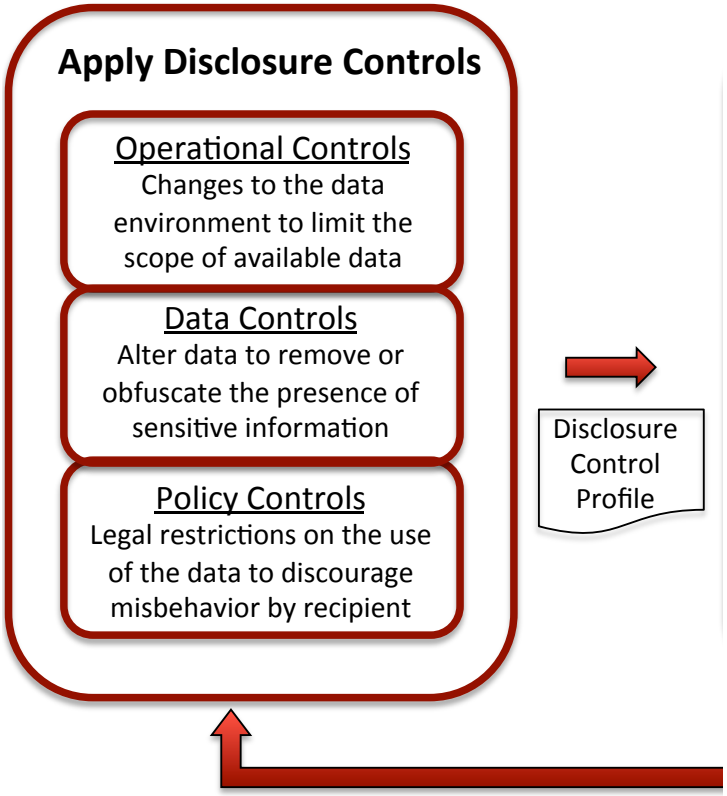
- Summarize utility requirements derived from those high-level goals into six categories
  - Low score indicates high-utility requirements
  - High scores indicate low-utility requirements
  - Examples:
    - Timeliness (5) → longitudinal collection and delayed release
    - Detail (1) → complete information about each data packet

Data Risks

Data Type	What	Who	Why	Overall	Justification
Client IP Address	4	4	1	4	Considered subscriber data under ECPA; Restricted under court order and private agreement; Indirectly identifiable under ethical precepts; Intended use is consistent with agreements and laws
DNS Data	3	4	1	3	Query name possibly considered content under ECPA; Other DNS info considered transactional under ECPA; Restricted under court order and private agreement; Confidential but not identifiable under ethical precepts; Intended use is consistent with agreements and laws

- Risk profile involves iterative exchange between technical staff and risk counsel based on risk factors derived from risk sources
  - Q&A: Who? What? Why?
  - Low score = low-risk elements of sharing
  - High score = high-risk elements of sharing
  - Examples:
    - Client IP, What (4) → explicitly restricted as indirectly identifiable
    - DNS Data, What (4) → not explicitly restricted, though confidential

Where are we?


Apply Disclosure Controls
Operational Controls
Changes to the data environment to limit the scope of available data
Data Controls
Alter data to remove or obfuscate the presence of sensitive information
Policy Controls
Legal restrictions on the use of the data to discourage misbehavior by recipient
Disclosure Control Profile

Choosing Disclosure Controls

- Necessary changes:
  - Ensure client IP address does not identify victim
    - Limit personal identifiability, prevent further exploitation
  - Restrict access to only DNS researchers and analysts
  - Limit redistribution of data by recipient

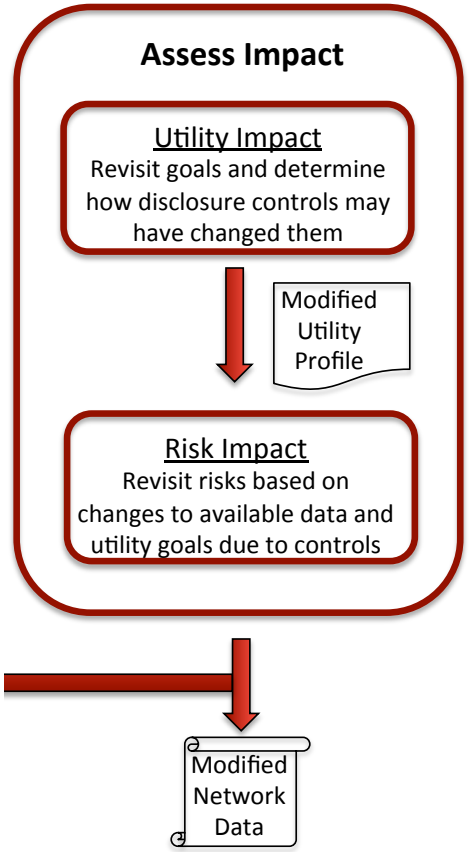
- Balanced against:
  - Relatively specific geographic information
  - Monitoring DNS queries of independent clients

Choosing Disclosure Controls

Disclosure Control	Data Type	Intensity	Description
Time	All	3	Withhold data until remediation
Access	All	3	Authenticated access by researchers
Generalization	Client IP	4	Truncate IPs at longest autonomous system prefix
Pseudonymization	Client IP	3	Replace remainder of IP with linkable pseudonyms
Policy	All	2	Verification of affiliation and private agreement

- Consider disclosure controls that mitigate the types of sensitive information identified by risks
  - Withhold data until malware remediation is likely
  - Implement limited vetting of researchers and analysts
  - Replace actual client IPs with pseudonyms
  - Lightweight policy restrictions to prevent redistribution

Where are we?


Assess Impact
Utility Impact
Revisit goals and determine how disclosure controls may have changed them
Modified Utility Profile
Risk Impact
Revisit risks based on changes to available data and utility goals due to controls
Modified Network Data

Utility and Risk Assessment

Category	Original Score	Modified Score	Justification
Audience	3	3	Access by legitimate security and networking researchers only
Timeliness	5	5	Research does not require immediate access
Duration	2	2	Research studies require long-term access to data
Detail	**1**	**2**	**Client IP granularity is restricted to high-level organizational info**
Functionality	2	2	General DNS-related research studies
Output	1	1	Publication of research findings gleaned from data

- Modified utility profile summarizes the changes to utility made by the disclosure controls

- Only significant change to utility is the resolution of client IP: changed to high-level organizations

Utility and Risk Assessment

Data Type	What	Who	Why	Overall	Justification
Client IP Address	2	3	1	3	De-sensitize data by reducing identifiability
DNS Data	3	3	1	3	Mitigate against maliciously motivated recipient

- Technical personnel again interact with legal counsel through questions about the modified utility profile

- Primary reduction in risk from replaced client IPs and limited access to vetted recipients

Summary

- Framework can be an **agent of change** for current data sharing
  - Risk uncertainty, benefit articulation
  - Vanilla sanitization, skewed risk-benefit calculation

- Enables data providers to **trust recipients and be trusted** by oversight entities
  - Provides a unified methodology for examining risk and utility
  - Explicitly states justifications and assumptions for choices
  - Facilitates interaction between technology and policy

- **Re-conceptualizes risk** by promoting the social value of shared data and processes