A Qualitative Risk Assessment Framework for Sharing Computer Network Data

Scott E. Coull
RedJack, LLC.
Silver Spring, MD
scott.coull@redjack.com

Erin Kenneally
Elchemy
San Diego, CA
erin@elchemy.org

Abstract

The availability of computer network data is critically important for network operations, the development of next-generation systems, and creation of the evidence-based policies that drive them. Collection and sharing of this network data, which may range from detailed packet traces to aggregated security alerts, is effectively the only way to concretely understand the complex structure and function of real-world networks. The information gleaned from such data sharing efforts is key to enabling innovation and resolving formidable issues in electronic crime forensics, infrastructure security, Internet governance, and intellectual property protection.

The *data provider's* decision to share network data is ultimately anchored in trust. The nature and extent of that trust is a confluence of the capacity to satisfy relevant legal requirements, the value placed on potential benefits, and confidence that the *data recipient* will not increase the provider's risk of disclosing sensitive information. What is most notable about the current state of practice is that most would-be data providers have a relatively weak understanding of these individual issues and their interplay, particularly in the context of computer network data. Risk is fueled by uncertainty about the application and interpretation of legal restrictions and obligations (e.g., regulations and privacy laws) related to network data disclosure, and exacerbated by unfamiliarity with disclosure control methods. At present, most efforts related to network data sharing focus on defining common data formats and exchange procedures for information interoperability. However, little work has been done to address the need for generalizable and scalable guidance that helps data providers understand and reason about these data sharing issues, thereby enabling risk-sensitive data disclosures that consider both legal constraints and utility needs.

We propose a reference risk assessment framework comprised of three phases that are designed to be generalizable across a wide range of data sharing scenarios. The first phase establishes a disclosure control continuum along two primary axes: the intended utility objectives and the disclosure restrictions imposed by legal, contractual, and ethical concerns. In the second phase, we describe operational, technical, and policy-oriented disclosure control methods, along with how they may be applied to adjust the balance between utility and risk mitigation. Finally, the third phase assesses the chosen controls with respect to the stated utility objectives and disclosure restrictions in the context of a qualitatively defined threat model. Significantly, our approach is unique among similar data sharing efforts (e.g., health data) because there is no generally accepted quantitative approach for assessing the risks associated with network data sharing, nor a practicable framework for addressing the growing number of related threats. Instead, we focus on educating data providers toward more effective, balanced, and pragmatic decisions that build a level trust and understanding necessary for productive network data sharing.

1 Introduction

The pervasive nature of modern technology, rapid expansion of Internet based services, and collective migration of nearly all aspects of civil society to the web make it easier than ever to generate, collect, use and share information. Data shared about and from individuals helps in the development and delivery of new services and products– from personalized search results and shopping recommendations, to automatic categorization of email and social networks, and on through advances in scientific discovery. These benefits, however, often come with a cost to individuals' privacy and thereby create the need to strike a balance between the risks and utility associated with disclosing data. This disclosure tension exists similarly at the organization level with data collected about and from computer networks and devices such as packet traces, network flow logs, or intrusion detection alerts. For organizations, the intended operational, security, legal, and economic benefits that are derived from the shared data by all data recipients must be considered in parallel with the probable risks that disclosure presents to the data provider. From an operational

perspective, organizations that engage in bi- or multi-lateral data sharing gain access to a more complete perspective on network and systems events, which informs their tactical and strategic efforts to avoid or mitigate threats and disruptions. Meanwhile, academic and industrial research benefits from access to real-world data that helps inform collective understanding of threats, validate new technologies and services, and inform the direction of innovations. Without access to this shared data, researchers and network operators alike are left with only relatively narrow views on network events and small datasets that may not adequately reflect real-world conditions, thereby hindering the development and application of new technologies and services. Beyond organization self-interest, the collective benefits of enhanced sharing are unsurprisingly similar if not grossly underestimated. They range from improved understanding of network public health threats and effectiveness of countermeasures, to enhanced data quality that in turn legitimizes decisions and allocations of resources, on through to the strengthened trust among partners, competitors, policymakers, regulators and the public that flows from information accountability. Importantly, the long term impact of broader disclosure by providers and access by recipients is an evolution of collective norms about 'risks of first impression' and the relative value of sharing, thereby lowering the real and perceived barriers to considered disclosures.

Widespread sharing is tempered by real and perceived risks associated with the data to all participants, although they are born primarily by the *provider* of the data. Since network and security data encompasses information about a range of human and device communications, it may contain directly sensitive information about individuals such as passwords, mailing addresses, financial transactions, and behavioral information. The data may reveal information considered sensitive to the disclosing organization itself such as intellectual property, security procedures, and business relationships restricted by non-disclosure agreements and other policy restrictions. Complications arise when the sensitive information is encoded within the statistical properties of the data itself. This can add uncertainty about its sensitivity and make it difficult to control for the risks associated with the information disclosure. While managing these privacy and confidentiality-related challenges falls on a spectrum of difficulty, all effective approaches require understanding the risks and the available methods for controlling those risks.

We propose a risk-utility model – *the Disclosure Control Framework* – tailored to help data providers understand and take action in the face of this tension, thereby enabling risk-sensitive data sharing that considers both risk constraints and utility needs.

1.1 Motivations and Approach

There is express and tacit dissatisfaction with the prevailing models for considering and managing the risks and benefits of information sharing. Discourse revolves around course-grained criteria for the disclosure of data that necessarily shoehorns data *providers* and *recipients* to accept disjointed and binary trade-offs between utility and risk – for example, can we disclose network traces to entities from China? Can we disclose flow records if we scrub all Internet Protocol Addresses? This 'state-of-the-art' reflects a demand for a more sophisticated approach to disclosing data that allows participants to address data sharing according to the reality of nuanced disclosure contexts.

Decisions by data providers to share network data are ultimately anchored in trust, the nature and extent of which is a confluence of: the capacity to satisfy relevant legal requirements; the value placed on potential benefits; and confidence that the data recipient will not increase the provider's risk of disclosing the data. What is most notable about the current state of practice is that most would-be data providers have a relatively weak understanding of these individual issues and their interplay, particularly in the context of computer network data. Risk is fueled by uncertainty about the application and interpretation of legal restrictions and obligations (e.g., regulations and privacy laws) related to network data disclosure, and exacerbated by unfamiliarity with disclosure control methods. Trust then is largely a function of ad hoc, bilateral, and interpersonal relationships between individuals within the organizations providing and receiving the data. This is substantiated by the fact that network data providers and recipients lack a scalable and transparent process by which providers– be they researchers, operators or technology vendors– can make data available for operational or research purposes in a sustainable manner. Consequently, what exists is a self-perpetuating cycle that does nothing to improve the lack of trust in organizational data sharing, instead reinforcing a risk-averse posture that precludes all but the most restrictive forms of sharing between organizations. At present most efforts related to data sharing focus on defining common data formats and exchange procedures for information interoperability, but very little work has been done to address the need for generalizable and scalable guidance that helps data providers understand and reason about these data sharing issues, thereby enabling risk-sensitive data disclosures that consider both legal constraints and utility needs.

Our proposed risk-utility framework addresses this systemic deficiency by conceptualizing risk, articulating utility needs, and helping to operationalize data disclosure controls. This approach allows data providers to more efficiently manage risk and achieve desired utility in the face of legal speculation and indeterminate outcomes. The significant

merit of this framework is that it arms data providers with an objective and defensible *process* for considering, designing and communicating how to balance risk and utility when making data disclosure decisions. This is the prerequisite to engendering the discourse and shared reasoning that is fundamental to raising the quality and scope of trust that drives data sharing.

1.2 Goals and Overview

Toward the objective described above, the goals of this document are to provide a practical reference for data providers, both at the technical personnel and risk decision-maker levels, to:

- Identify common sources of data risk, consider risk factors, and create a data risk profile
- Specify intended utility for data recipients and create a data utility profile
- Understand common operational and technical disclosure controls
- Assess the impact of the disclosure controls on the initial data risk and utility
- Understand how to modify the application of the disclosure controls to achieve different risk and utility outcomes
- Consider the cumulative utility of more widespread sharing as a countervailing force to speculative risk

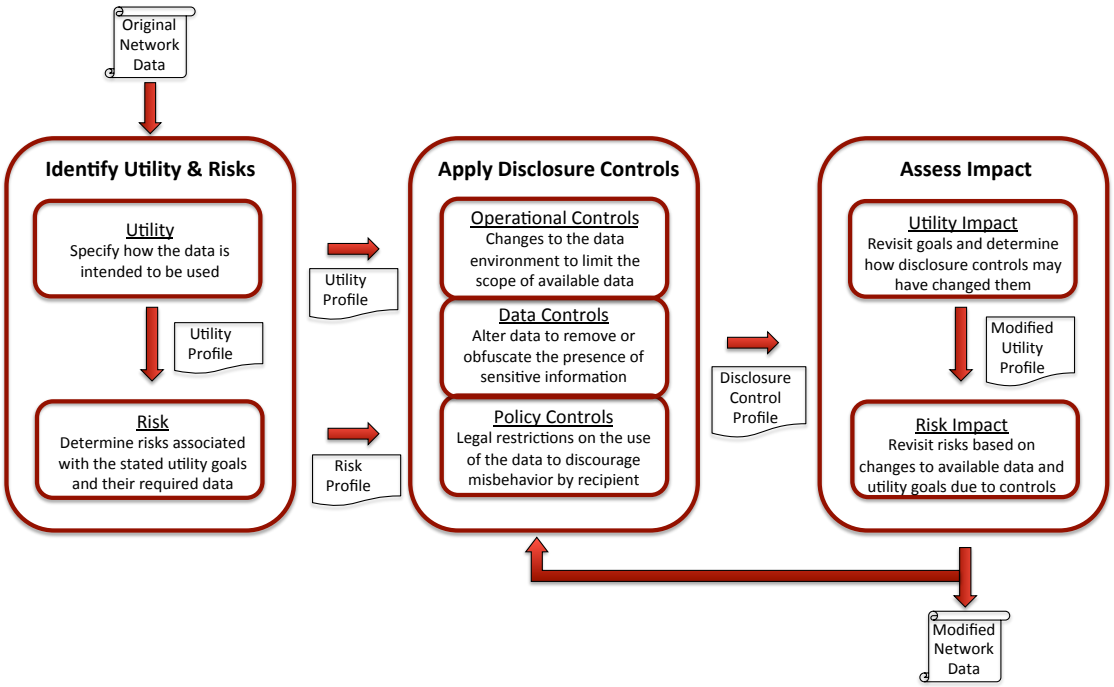

Figure 1: Overview of risk-utility framework for network and security data.

Our Disclosure Control Framework outlines a process for both understanding the objective factors (the "what") that define the myriad of risks and utility objectives, and also, for applying them according to the subjective choices (the "how") of the participants in any *data sharing scenario*.[1] Our framework is broken into three phases– analysis, application and assessment (Figure 1). Each phase in the framework discusses, in necessarily general terms, the primary considerations essential to developing a data provider's data sharing scenario: the desired utility objectives, relevant risks, options for disclosure controls, and the impact of those controls on the chosen risk and utility determinations. This framework is intended to address our normative and empirical expectations about data sharing risks and utility by embedding a process for decision-making that forces the provider to consider risks drawn from authoritative sources

[1] We use this term to describe both the broad, enterprise-wide efforts such as CRAWDAD, PREDICT, DShield, or CAIDA, as well as specific dataset instances.

alongside conscious choices about data utility, given a comprehensive set of options that can be tailored to its appetite for accepting certain risks and altered intended outcomes, tied together by an evaluative loopback that reinforces the evidence-based deliberateness and of the providers chosen actions.

The initial phase examines the two components around which the sharing scenario is anchored– the risks involved with sharing the data, and the intended utility for the provider and/or Consumer. The input for this phase is the data corpus that the provider wishes to make available to the Consumer(s), such as network flow records, packet traces, or application logs. Risk considerations and utility determinations are applied to the data, thereby producing a data *risk profile* and a *utility profile*, respectively. The risk profile helps the provider understand what common statutory, contractual, proprietary, ethical, policy, and best practices obligations and restrictions are implicated when releasing network data. Importantly, this phase considers another aspect of risk, the *threat factor*, which addresses the capability and motivation of the intended recipient(s) to enhance the provider's risk. By embedding consideration of the threat environment as part of the risk assessment, this framework allows all stakeholders– providers, recipients, and third-party oversight authorities– to assess and defend the reasonableness of a provider's choice of disclosure controls based on adherence to applicable performance criteria – utility and risk. The risk profile is scorecard of risk factors for ten (10) common data types based on sensitivity and context. The risk profile is manifest qualitatively according to a High (5)- Medium (3)- Low (1) scale for each data type contained within the disclosed dataset, accumulating into a Phase 1 output that is number-coded set of risk indicators for each of the N-data types present, prior to the application of any disclosure controls. Similarly, the utility profile is a confluence of the choices for six (6) utility choices for the entire dataset that also exist on a continuum from highly restrictive data uses (5) to completely unhindered use (1), where use is proportional to chosen utility. The output is a number-coded qualitative depiction of the initially-desired outcome for the data sharing scenario.

The second phase of the framework considers the Phase 1 risk-assessed data (*risk profile*) and utility-ascribed (*utility profile*) data in parallel and offers a menu of disclosure control options for the provider to apply to the shared data to achieve appropriate risk and utility since each option has measurable impact on each of the two components. These controls are organized as operational– how the Consumer may interact with the shared data, technical– how the data can be altered to prevent sensitive data leakage, and policy– how a provider can address the identified data risks ex post via contractual and policy-oriented agreements concerning the access, use and secondary disclosure of the data by the Consumer. The notional application of the disclosure control(s) allows the provider and Consumer to foresee how a chosen control will enhance or decrease both risk and utility. The output of this application phase is a *modified risk profile* and a *modified utility profile* that reflect the provider's acceptable risk and expected outcome.

The final phase of the framework addresses the impact of the choices made in the previous phase by mapping back to the identified risks and stated utility objectives from the first phase to determine how the *risk profile* and a *utility profile* has changed. This Phase directs the provider to compare the *modified risk profile* to the original *risk profile* and assess whether the recalibrated risk is acceptable, or whether the provider needs to loopback to Phase 2 and apply different disclosure control(s) that will pre-empt unacceptable risk exposure. In parallel, the provider is instructed to map the *modified utility profile* back to the original *utility profile* in order to evaluate whether the desired properties of the disclosure-controlled data (*modified utility profile*) satisfy the original utility objectives beyond the point where it undermines the purpose of sharing the data.

This Disclosure Control Framework uses qualitative metrics to enable data providers to practically assess and communicate risk and utility decisions related to the data disclosure. The component parts–identification of utility and risk, application of operational and technical disclosure controls, and assessment of their impact on utility and risk– are described using a simple rating scale, where 1 indicates low risk, 3 indicates medium risk, and 5 indicates high risk. The numbers are not used in a quantitative sense, but rather, as familiar means to symbolize the risk continuum. While true quantitative methods (*e.g.,* [6, 20, 14, 1]) provide the data publisher with a specific value indicating the relative severity of the risk of leaking sensitive information, these methods are effective when applied to a very specific threat scenario with a well-defined notion of what the sensitive information is. In the case of network data, however, sensitivity defies precise definition, as it is often context and fact-specific. Instead we are more concerned about general qualities of the data and what they might imply based upon our knowledge of how computer networks operate. In particular, this approach reflects the philosophy that the data publisher is better served by understanding general properties of the disclosure control process and then applying that knowledge to the intricacies of the situation at hand. Doing so ensures that the framework is general enough to be applied to new technologies and data sharing scenarios, and avoids the problem of trusting the results of quantitative analysis when they may not be applicable. For those wishing to gain a deeper technical understanding, excellent surveys are available for the general problem of privacy-preserving data sharing [3], as well as sharing network data [4, 25].

2 Identifying Utility and Risk Considerations

Before selecting the disclosure controls appropriate for the data release being considered, the data publisher must first understand the underlying utility goals and applicable risks inherent in sharing that data. In particular, the publisher must have an understanding of how they expect the data to be used and what type of information must be retained within that data to enable that use. Furthermore, it is important to define the restrictions on this information specified in relevant legal, contractual, and other policy guidelines. To aid the publisher in achieving the necessary level of understanding, we break down these issues into canonical categories which guide the publisher through the steps of quantifying the important utility and risk factors related to the data release. At the end of the first phase, the publisher will be able to develop a *utility profile* and *risk profile* that roughly quantify the general aspects of the data that must be preserved for utility reasons or protected due to risk concerns. Together, these utility requirements and risk-based restrictions serve to inform the later choice of disclosure control methods in the next stage of the disclosure control process.

2.1 Utility Objectives

The notion of *data utility* is perhaps the most important aspect of the data release since it defines the intended use of the data that motivates the entire process. By mapping out the data utility objectives in the early stages of the data release process, we explicitly identify certain constraints that our later choices of disclosure control must conform to. Moreover, the utility objectives help to specify the threats that we must protect the sensitive information in the data from, since they define how the adversary may interact with the data. As we will show in the next section, these utility goals are balanced against a number of risks that must be mitigated through the use of disclosure control methods.

While data utility is an important aspect of releasing shared network data, it is also exceptionally difficult to concretely specify. When thinking of utility, it is often easiest to consider general classes of functionality, like detecting botnets or measuring bandwidth usage. However, those activities can be achieved using dozens of different approaches, each with their own requirements on the information that must exist in the shared data, and consequently the applicable disclosure control methods. In addition, new classes of functionality are added as the state of computer networking and security develop over time. Certainly, a complete classification of these functionalities and their respective requirements would be a daunting task and would add unnecessary complexity to the already difficult task of releasing shared network data.

Instead, we put forth a set of six utility criteria that specify the most important aspects of how the data will be used by recipients. These criteria include: who accesses the data, whether there is a specific functionality or use case, the necessary level of data detail, and the output of the data recipient's use of the data. These criteria define the aspects of the data that must be maintained when applying disclosure controls. Generally speaking, the settings for these criteria represent a continuum of utility choices from those that are highly restrictive on the use of the data to those that allow for completely open ended use. The six utility criteria are defined as follows:

- **Audience.** The audience of the data release describes the type of community that is intended to act as the recipient of the data. The audience continuum can be roughly divided into individuals at the most restrictive end, consortiums in the middle, and public release at the most relaxed end. *Individuals* simply represents one-off data sharing approaches between two parties, which may include independent researchers, corporations, academic institutions, or other organizations. *Public release*, as the name suggests, indicates that the intention of the release is to make the data available to the general public. Finally, *consortiums* are a middle ground where groups of cooperating data producers and recipients freely share data within the confines of the group. Of course, there are many settings for the audience criterion between these fixed points, such as the approach taken by the CRAWDAD data repository [7], where membership in the data sharing consortium is relatively easy to obtain and data is freely available once membership is granted. By contrast, the PREDICT repository [18] lies between individual and consortium release since it maintains a closed community of recipients and producers, but only facilitates sharing among group members instead of allowing unfettered data access within the group.

- **Timeliness.** Certain uses of the shared network data may place requirements on how quickly the data is made available to the recipient. Here, the continuum of setting choices ranges from real-time access to longitudinal data collection. *Real-time* data access allows the recipient to analyze events as they occur, which is likely to be quite useful in the context of network and security operations. *Longitudinal* data collection, on the other hand, gathers the data over relatively long periods of time and releases that data in batches. While some data sharing efforts, such the Spamhaus spam blacklist [23], have obvious (near) real-time requirements, there are many examples of data sharing efforts that fall in the middle of this continuum. For example, data collected by

CAIDA [2] from network telescopes may contain details of new worm outbreaks and botnet infestations that are only useful when shared in a relatively timely manner, but which does not need to be made available in real time.

- **Duration.** Once the data recipient is given access to the data, the duration criteria indicates how long the recipient may access that data. Data access durations may span from short-term access to indefinite access. The Lawrence Berkeley National Laboratory (LBNL) Enterprise Traffic data release [16, 17], for instance, implements an effectively indefinite access period since the data is made available with no concrete usage restrictions. This type of long-term access is best for research purposes where the time scale of the project cannot be accurately scoped or new discoveries may extend the legitimate use of the data. By contrast, many data sharing programs used by operational groups, such as REN-ISAC [19], share data about specific incidents that is tightly scoped to the surrounding investigation, which may be as short as a few hours.

- **Detail.** Another important aspect of utility is the level of detail required by the data recipient. Some operational tasks, such as investigations of specific security incidents or troubleshooting network outages, require detailed data about individual network events. Other uses, like general research on network technologies, are often most interested in overall trends that manifest themselves in the data. Therefore, one may consider a continuum of data detail from *event-specific* data to *general trend* information derived from the data. Most current sharing efforts tend toward the event-specific portion of the spectrum by releasing exact event information or slightly altered network traffic logs. Collaborative security efforts, such as the DShield firewall service [8] and the Spamhaus black list, require participants to provide specific attributional information about security events. Similar detailed information sharing occurs with measurement-related projects, like the Electronic Freedom Foundation's (EFF) Panopticlick [10] or SSL Observatory [11] projects. At the same time, the research-oriented data made available through the PREDICT and CAIDA data repositories often consists of truncated packet captures or aggregated network flow logs. At the far end of the spectrum, it is also possible to simply release general statistics about the data, as CAIDA does for the general public.

- **Functionality.** In determining the utility properties necessary of the data, we may also consider whether or not there is some concrete functionality that is the target of the data release. On one end of the functionality spectrum are data releases tailored to specific *concrete tasks*, while on the other are *open-ended tasks*. Data releases that are tailored to specific tasks, like the example of spam black lists, have the benefit of allowing the provider to tightly restrict the information that is kept in the data release since there is a well-defined functionality that is expected. Open-ended use of the released data, which is typified by the CRAWDAD repository and LBNL data release, can be more problematic because there is no clear definition of what the recipient may be using the data for or what functionalities are expected. Some existing data sharing efforts also take a middle ground by allowing the provider and recipient to negotiate functionality expectations, which is the case with the PREDICT repository.

- **Output.** The final aspect of data utility revolves around the intended outcome of using the data, or the output of the recipients' interaction with the data. Operational data sharing efforts by REN-ISAC, DShield, and Spamhaus, for example, focus on producing a certain level of private or semi-public knowledge that can be used to address specific security threats or network problems. Research-oriented sharing from the PREDICT and CRAWDAD repositories, however, aims at broad publication of knowledge gained from interacting with the data. Therefore, we may consider *private knowledge* to be the most restrictive form of output, while *publication* would be the most broad form of dissemination. Again, there are many approaches that fall between these two extremes. Spamhaus uses the shared data to create tightly-scoped and operationally-relevant public knowledge with clear limitations, for instance.

The settings chosen for these six utility properties make up the provider's *utility profile*. In later sections, we will refer to this utility profile to determine the appropriate types of disclosure controls to achieve these goals. At the same time, this utility profile must be balanced against the inherent risks of sharing the data. In some cases, the utility profile and the identified risks may actually strongly conflict with each other, which we will identify during the assessment component of the framework. When such conflicts arise, it may require restricting or loosening the utility settings identified herein. These profiles illustrate the utility criteria for the respective data repositories, which allows us to understand and compare their utility goals at a high-level. Specifically, we can see that the utility profile of the two are roughly equivalent, except that CRAWDAD has a goal of enabling broader access to the shared data and contains data with detailed wireless mobility information.

2.2 Data Risk Profile

The other component of this first phase involves developing a *risk profile* for the data to be shared. This data risk profile is a function of the overall data sensitivity and the disclosure threat context, and is derived from various *sources* of disclosure restrictions, obligations and considerations. Data sensitivity is determined by *risk factors* intended to help a provider to gauge the relative data sensitivity and associated disclosure risk along qualitative continua from low to medium to high. These risk continua correspond to general thresholds for whether the disclosure will be permissible, restricted, or prohibited, respectively. The threat context considers threat factors that indicate the capability and motivation of the recipient to re-sensitize the data by undoing a disclosure control(s) (via reverse-engineering the disclosure control or link attacking the disclosure-controlled data) or by inferring sensitive data from that which is disclosed.

There are variable degrees of certainty with each risk and threat factor, so unsurprisingly the overall risk profile is the product of a range of options available to the provider and reflects its appetite for risk. It is necessary to identify the data sensitivities and threat prior to data sharing because those will impact the disclosure control strategy by its own accord, as well as for the purpose of balancing utility objectives which may conflict or be in tension. The data risk profile component of this framework is designed to facilitate effective communication and assessment of risk by helping the provider translate network level concepts to non-technical notions of sensitivity and threat. It is unreasonable to expect that personnel who have intimate technical familiarity with the data will be capable of rendering ultimate determinations related to the risk sources (e.g., law, regulation, contract). Therefore, the development of the risk profile involves an iterative exchange of information between the provider's technical and risk management personnel (e.g., legal counsel):

1. Technician provides the risk manager with factual information related to the risk sensitivity and threat factors
2. Risk manager applies those facts to the risk criteria and renders a risk profile
3. Technician uses the risk profile to inform his Phase 2 application of disclosure controls

2.2.1 Risk Factors and Criteria

Each risk factor is defined by criteria that reflect the relative sensitivity of the data along a continuum from low, to medium, to high. Each data risk profile is specific to the context of the provider and recipient relationship. Nevertheless, there are generalizable requirements and obligations that can be abstracted from the sources, and from which we filter risk factors that are common across all of the legal sources, a "data specification" of sorts. These factors are characterized as follows:

- **Nature of the Data (What).** The type of data that is made available may impose a standard of care that creates disclosure risk. For example, the disclosure of financial or medical data are highly restricted both for financial services and health industries, respectively, but are also constrained for entities outside of those spheres. In general, these data-centric factors turn on disclosure related to the privacy interests in the data– namely "personally identifiable information" (PII)[2] The criteria is thus anchored around *identifiability* and *confidentiality* and spans the continuum according to whether: (a) the data is explicitly labeled confidential or directly identifies an individual, to (b) data that is an indirectly sensitive because it provides a reasonable basis to infer confidential information or to identify an individual/entity, to (c) data that does not identify an individual/entity or from which confidential information cannot be reasonably inferred. The combination or quantity of indirect identifying data can alter where a certain type of data lays on this continuum.

  To facilitate translation dialogue, the technician should identify the network data type(s) within the disclosed data and may map it to one or more of ten common data categories[3] according to his intuitive understanding of how they are organized. This categorization is neither new nor authoritative, but rather, offered as a way to taxonomize the network level data according to concepts familiar to general society and explicit or implicit in various laws, agreements, ethical precepts and policies, as illustrated below:

[2] "[A]ny information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information, or data explicitly labeled confidential because it is proprietary or privileged." NIST Special Publications 800-122- Guide To Protecting the Confidentiality of Personally Identifiable Information. http://csrc.nist.gov/publications/nistpubs/800-122/sp800-122.pdf.

[3] Each data type includes both individual and organizational examples.

Name- first and surname (individual), business name.

Credential- government-issued and other account identifiers such as social security number, medical identifier (UPI- universal patient identifier), healthcare identifier (Medicare/caid), financial identifier (investment account), virtual identifier (email address), professional certificate/license number, alias, DUNS number.

Biometric- photograph (rendered or digital), fingerprint, DNA/genetic profile.

Location- Narrow (residence at street level or below), Broad (residence at geographic level broader than city subdivision, such as zip code).

Relation- names of family and friends, group affiliations, B2B/B2C names (customer, contractor, client).

Health- dates of birth and death, status of network/system infections (malware/spam), network/system vulnerabilities.

Behavioral/Transactional- Internet search terms, browsing behavior (page/ad views, site visits), purchase data and habits, other preference/opinion data, cost/price/billing data.

Demographic- gender, marital status, socioeconomic status, race, political affiliation.

Extension- vehicle identification and serial number, device or virtual identifiers (URL, MAC address, IP address), device fingerprint, other property identifiers (intellectual or physical).

Communications- human-to-human content, device-to-device level messages.

- **Nature of the Participants (Who).**

  Data Provider. This factor reflects the familiar and established data protection regulations that impose obligations on disclosure providers based on industry sector such as health and medical, financial, critical infrastructure, employment, education, insurance, credit, cable and telephone, and government. The criteria runs the gamut between entities that are highly regulated or bound by private agreements at the high end of the sensitivity continuum, to entities that are not covered by a regulatory regime at the low/no sensitivity end. Translation dialogue may not be necessary here since presumably the risk manager is familiar with the legal categorization of the provider. However, to the extent that such interpretation is disputable the technician may need to provide details about operations. For example, whether or not a company is considered an "Electronic Communications Service" provider to the public has nontrivial implications under electronic communications privacy law, and may require the risk manager to understand how the company provides services in order to make that assessment.

  Data Recipient. Consideration of the data recipient as a risk factor impacts the overall risk profile by addressing requirements that define the type of entities that may receive certain data, the likelihood that the recipient will nullify the disclosure controls implemented by the provider by re-sensitizing the data or inferring sensitive data from that which it receives. Also, considering the potential threat posed by the recipient addresses the vague or catch-all authoritative requirements discussed in 2.2.2. Since these types of requirements presume consideration of factors apart from the data itself, they necessarily hold the provider accountable for such assessments. This can be a daunting task, leaving some providers to believe they are being asked to predict all the possible threats that could expose sensitive data. At a time where data availability, analytic capacity and incentives are rapidly advancing, this could lead a provider to conclude that the exception swallows the rule and *everything* is potentially identifiable or confidential. This threat factor, however, organizes that threat into manageable model that enables a reasoned approach to assessing the *probable* inference threats, thus eschewing the impracticable endeavor to measure all *possible* threats that will render sensitive data. It also captures the provider's legal, contractual and/or ethical obligations to perform due diligence on downstream recipients by reasonably measuring the likelihood that improper use by a recipient may expose the provider itself and/or individuals implicated in the data to financial, physical, or other material harm. To facilitate translation dialogue, the technician should articulate to the risk manager details related to the following threat factors:

  The **recipient's capability factor.** This is defined by the following KSA criteria:

  1. Knowledge– the availability of internal background or external information that can re-sensitize the disclosure controlled data. Knowledge can be publicly-available (e.g., located on the Internet with no/few access restrictions), of limited availability (e.g., private database that does not require extraordinary financial resources or qualifications, such as is available from data aggregator services), or not reasonably available (e.g., data can only be obtained by exerting significant financial or legal process);

2. Skills– the recipient's degree of technical proficiencies or insight. Settings range from a highly sophisticated network domain expert, to a recipient that has skills equivalent to the provider, to the proficiency of the average public or lower with regard to network level information; and
3. Abilities– the level of computational resources, time and effort, and/or financial capacity needed to re-sensitize the data. These settings range from abundant computational capacity, time or economic resources at one end to scarce resources at the other.

The confluence of these KSA criteria allows the provider to gauge the capability of the recipient to derive sensitive data or recover the data that was "de-sensitized" via the technical and operational controls.

The **recipient's motivation factor.** This considers the intent of the recipient to re-sensitize the disclosure-controlled data and is defined along the following criteria:

1. Malicious– the recipient intends to deliberately attack the disclosure control or otherwise derive sensitive data from the shared data for the purpose of causing harm or damage;
2. Extra-Purposeful– the recipient is inclined to subvert the disclosure control or increase the inference risk for purposes beyond the scope of why the data was collected or shared;
3. Negligent– the recipient does not intend to re-sensitize the data but is careless in complying with data policy controls imposed by the provider;
4. Unintentional– the recipient is not interested in re-sensitizing the data.

As with the other framework components, the threat context also exists on a continuum that is proportional to risk, from high threat and risk, to medium threat and risk, and finally to low threat and risk. The combined KSA is characterized along a continuum from Expert to provider Equivalent, to general public. It intersects with the Motivation continuum that span from high threat malicious actors down to low threat disinterested lay persons. Note that the threat context is in some ways tacitly defined by the provider's chosen utility posture since it determines the boundaries for how an adversarial recipient may interact with the data. For example, analytical output that is intended to be disclosed publicly (see, 2.1) exposes the data to the full range of probable threat environments, whereas output restricted to private consumption via disclosure controls will reasonably eliminate the malicious threat factor, thereby reducing the risk exposure.

- **Nature of the Use (How/Why)** Laws, policies, contracts or ethics may specify for what purposes the data may be applied. How the data is being used, or whether it has availability restrictions are context factors that impact the sensitivity associated with its disclosure. To facilitate translation dialogue, the technician should articulate to the risk manager details related to the purpose for which the data is being disclosed and how the recipient intends to use it. For example, will the data be used for internal or external business purposes? Will it be used consistent with the purpose for which it is collected or for a secondary purpose? Is it prohibited from public release?

2.2.2 Applying Risk Factors and Criteria

These continuum correspond to general thresholds for whether the disclosure will be permissible, restricted, or prohibited, respectively. In short, sensitivity is proportional to risk and is a function of data privacy and confidentiality derived from the risk source(s). This data characterization neither imposes nor assumes a static and predetermined judgment of the comparative degrees of sensitivity between the data types. Rather, this characterization permits any data type to fall on a spectrum of sensitivity as dictated by the authoritative source(s) relevant to the provider-recipient relationship and as interpreted by the provider. Enumerating what data type has sensitivity restrictions when it is explicitly called out in the relevant authoritative source is straightforward as with expressly labeled confidential or directly identifiable data. Recognizing what data should be considered sensitive when faced with obligations for inferable data can be confounding when there are restrictions on indirectly identifiable data or ambiguous"catch-all" obligations.

The Data Risk Profile addresses the risk continuum from obvious to tacit restrictions by inferring relationships from explicitly sensitive data types based on inherent characteristics of the data types. For example, a Contract between provider and recipient that explicitly restricts the disclosure of individual Names as well as data from which Names can be reasonably inferred, would mean that if Credential data is contained in the disclosed dataset, there would be high risk that Name data will also be disclosed.

The output from this initial identification of risk phase is a Data Risk Profile which is a composite of the *policy risk profile* (the risk for each data type contained in the data to be shared) mapped to the *network data risk profile*, and

which specifies where risks lay in the disclosed network data. That mapping is done at the complete discretion of the provider based on its understanding and interpretation of the dataset. Further, that mapping between policy data type and network data type is not 1:1, rather, it can be many-to-one or one-to-many. For example, provider may determine that *extension policy data* can map to *IP addresses, MAC addresses and URL network data types* in a given dataset. Or, *extension, location, and credential policy data types* can map to *IP address network data type* in another dataset. The point is that the source of the risk allows the provider to identify *what* it should be concerned with, and then the provider determines *how* that translates to network data. In some cases it may be a direct mapping, such as client IP addresses from active DNS data where a contract specifies those to be prohibited, or it may require inference on the part of the provider when dealing with law that prohibits the disclosure of data that can be used to identify a person. In the face of vague or catch-all provisions that necessarily call for discretion, this approach allows the *how* to be applied by the provider. The current state of the art in translating policy data risk imposed by binding disclosure restrictions and obligations to network data risk is an unsophisticated and underinclusive. The network data risk is characterized along two axes- packet header and packet payload. This disclosure control framework is unique in presenting a more granular and nuanced interpretation that facilitates more efficient and effective avoidance of risk.

2.2.3 Sources of Risk

The preceding risk factors describe the general criteria that determines disclosure risk. These are derived from authoritative sources: law and regulation, private agreements, ethical obligations, policy and standards/best practices. The legally-binding requirements associated with the data are a composite of the explicit and implicit obligations from one or more authoritative sources that concern data protection obligations and information rights and assurance. The following are sources of risk that inform the disclosure restrictions related to the shared data.

- **Public Law and Regulation**.

  Organizations and persons are accountable for complying with applicable laws and regulations, and face civil or criminal liability for violations. Notwithstanding laws protecting proprietary and privileged data discussed below, there is a smattering of laws that prohibit, restrict or otherwise regulate the collection, use, or disclosure of personal information (identity or attribute data) and include a duty of care to safeguard data. Data protection and stewardship in the U.S. is governed by a patchwork of case law, and federal and state laws, regulations and statutes that create disclosure risk based on several factors (discussed below), unlike the overarching data protection model used throughout the European Union[4]. How network level data disclosure is implicated within the purview of these regimes can be ambiguous since although network data may not be a direct target of data oversight protections it is seminal to understanding threats to data, systems and entities that are expressly addressed by law, namely those that impose a duty of care to provide reasonable or appropriate security to protect information. Reasonableness determinations consider factors such as the value and sensitivity of the information, the threats that data is wrongfully disclosed, and the existence of legal and technical safeguards. Therefore, the standard of care for disclosing network data is an evolving product of the interaction between legal expectations and technology capabilities. For example, definitions of personally identifiable information (PII) and proprietary information are fundamental to interpreting and applying many laws protecting privacy and confidentiality, respectively. Translating those constructs to network data artifacts is relatively immature and evolving, so knowing the nature and extent of legal risk that attaches to its disclosure can be confounding. For example, network data such as Wi-Fi device signatures can indicate low level machine geolocation that poses no threat to users' privacy and no restrictions on data disclosure in one context, yet sharing the same data can trigger any number of data protection laws if the device signature is mappable to an identifiable person.

  Insofar as disclosure risk assessment turns on interpretation of fundamental concepts from the *human* layer names and locations to network layer protocols and encodings, the laws that implicate legal risks when sharing network data are fact- and application-dependent. The federal law that does conspicuously target network data, viz. expressly prohibiting the disclosure of electronic communications unless specifically authorized, is the Electronic Communications and Privacy Act (ECPA)[5]. Examples of other laws germane to network data are the Family Educational Rights and Privacy Act (FERPA), the Common Rule, the Federal Trade Commission Act Sec. 5, and common law invasion of privacy torts[6]. While there is currently no legal framework that writ large prescribes or expressly incentivizes the sharing of network data, there is growing momentum for the enactment

[4]Data Protection Directive (95/46/EC)

[5]ECPA, Pub. L. No. 99-508, 100 Stat. 1848 (1986).

[6]FERPA, 20 U.S.C. 1232g; 34 CFR Part 99; Common Rule, 45 C.F.R. part 46; FTC Act, 15 U.S.C. §§41-58.

of federal cyber security legislation that specifically authorizes the sharing of network cyber threat information[7]. Furthermore, some laws that restrict use and disclosure of data may include exceptions for research use of the data, permit sensitive/identifiable data to be shared with specific authorization/consent of the entity with rights associated with the data, or allow a waiver of authorization from an oversight authority such as an Ethics Research Board (ERB).

- **Private Agreements** The existence of contracts with the entity that originally collects the data or funds the generation of the data to be shared that restrict, limit or prohibit disclosure of the data are commonplace. This can be a prudent approach when mandatory laws/regulations allow modifications or exceptions to default provisions if expressly agreed by the entity(s) the law is meant to protect. Sources of such terms and conditions may be grants, contracts, memoranda of understanding/agreement/cooperation (MOU/A/C), non-disclosure agreements (NDA), service level agreement (SLA), or other bilateral agreements. Disclosure requirements may be nuanced, such as temporal limits on sharing or prohibiting the sharing of data derived from data provided by a third party. Disclosure may be restricted to certain entities (e.g., non-profits, educational institutions), for specific purposes (e.g., non-commercial), with the consent of a specific party, or if certain conditions are met (e.g., information security protections). And, terms may authorize disclosure if done via a license regime or accompanied by source attribution.

- **Proprietary Rights or Privilege**. The provider may have proprietary or intellectual property rights or privilege obligations in the shared data which impact its sensitivity and the nature and extent of the disclosure strategy. These rights could derive from an individual researcher and/or organization's protectable interests in data that: pertains to a patentable invention, has current or prospective intrinsic commercial value, has professional reputation value, constitutes a proprietary trade secret or privileged communication, or is restricted by copyright or license provisions. The quality, quantity, timing and target of the data disclosure can determine whether or not information is deemed to have been published, is considered to be in the public domain, or amounts to a waiver of protected data or communications, thus impacting intellectual property status or proprietary value.

- **Ethical Obligations**. Ethical obligations can impact disclosure restrictions or prohibitions if sharing the data potentially harms individuals or organizations by exposing private identifiable or confidential information that can result in financial, physical, legal, or psychological harm. Ethical principles and applications are reflected in various laws, treaties and guides covering industry trade groups, government-funded organization, or private corporations[8], and in general, implore provider to ensure that disclosure does not contravene respect for persons, balances potential harms and benefits, and regards fairness and equity. Ethics may restrict provider from disclosing data unless it obtains ethical review board (ERB) authorization, the consent of the entity that is the source of the data, or waiver of consent by an appropriate oversight entity. Ethical principles may prohibit disclosure if its purpose is inconsistent with either the purpose for which is was collected or the terms of the consent, or if the original authorization is no longer valid or is withdrawn. If data is obfuscated (coded, encrypted or otherwise anonymized or de-identified) such that potential harm is pre-empted, ethics may permit disclosure.

- **Unilateral Policy**. Data providers themselves may have privacy or confidentiality policies that restrict or limit data disclosure beyond the requirements of federal or state law. Sources can be internal Ethical Review Boards/ethical policies, or website, system or network Terms of Use (ToU), Acceptable Use Policy (AUP), or a privacy policy between the organization and its customers, clients, users or partners. These provisions typically describe how the entity collects, uses and discloses personally identifiable information. In contrast to bilateral agreements discussed previously, these policies are unilaterally established by the original data collector or the provider and self impose obligations to abide by data privacy and confidentiality promises when engaging in data sharing.

- **Standards or Best Practices**. Industry or professional standards and best practices guidelines are another source of disclosure limitation. While these may not wield the same enforceability as other sources of disclosure prohibitions, nonetheless, they can provide valuable accountability in self-regulation regimes (e.g., Network Advertising Initiative Code of Conduct for Online Behavior Advertising), as well as an anchor for public trust in

[7] S. 2151, the Strengthening and Enhancing Cybersecurity by Using Research, Education, Information, and Technology Act of 2012 (SECURE IT Act) (introduced March 1, 2012); S. 2105, the Cybersecurity Act of 2012 (introduced introduced on February 14, 2012); H.R. 3523, the Cyber Intelligence Sharing and Protection Act of 2011 (CISPA), (introduced on November 30, 2011).

[8] For example, Common Rule, 45 CFR part 46, the federal law governing research involving human subjects for institutions that accept federal funding; the Declaration of Helsinki; the Nuremberg Code; Hippocratic Oath.

unregulated industries or informal interest groups that galvanize around sensitive data supply and demand (e.g., responsible disclosure guidelines for software vulnerabilities).

3 Applying Disclosure Controls

To achieve the risk mitigation and utility goals laid out in the previous section, it is necessary to find ways to limit the types of sensitive information found within the collected network data. In this section, we describe a variety of disclosure control options that may be applied before data collection begins, after the data has been collected and stored, and after it is released to the recipient. The overarching goal is to create a *disclosure control profile* for each of the specific types of sensitive information identified in the previous step in the data sharing process, which in turn help the provider to create *modified risk and utility profiles* representing the state of the data after controls have been applied. As with other parts of the framework, the provider may choose from among a variety of controls, each with its own spectrum of utility and risk trade-offs. Here, the green portion of the spectrum represents low-risk controls with associated low utility, and red represents high-risk controls that maintain their utility. While the trade-off between risk and utility is rarely equal, the spectrum does provide a rough notion of the implications of the choices made while applying disclosure controls. The disclosure control profiles, therefore, explicitly state the provider's assumption about the protections offered by the applied controls and their impact on the stated utility goals.

Specifically, we consider three classes of disclosure controls: *operational controls* that restrict the data collection environment, *data controls* that transform the data to remove or hide the presence of sensitive information, and *policy controls* that mitigate risk once the data is accessed by recipients and other parties. For each category of disclosure control, we describe their high-level operation, several example implementations, and their respective benefits and weaknesses in protecting sensitive information and maintaining data utility. As mentioned earlier, although some of these disclosure controls may be used to achieve certain theoretical guarantees of privacy (e.g., k-anonymity [24], differential privacy [9]), those guarantees are not necessarily applicable to network data in any meaningful way [6]. Therefore, we do not discuss those theoretical definitions of privacy here and instead refer interested readers to related survey papers [3].

Before presenting the controls, we specify some basic terminology used throughout this section to indicate how the controls may be applied to the data. First, we define *first-order network data* to be a table of m columns and n rows, where a cell in the table is a specific column and row intersection in the table. In practical terms, this first-order network data represents the general format of raw data collected directly from the network in packet trace or flow record format, where each column is a field in the data (e.g., source IP address), each row is a packet or flow, and a cell is a specific field value for a packet or flow. From the first-order data, we may derive *second-order statistics*, including counts and averages across the available columns. As the disclosure controls are introduced, we will point out how they may be applied to the first- and second-order network data. The risk and utility trade-offs associated with each disclosure control is summarized as a spectrum running from low risk and utility to high risk and utility, similarly to the representations found in the previous section.

3.1 Operational Controls

Before any data is collected, the data provider may choose from among many types of operational controls that restrict the sensitive information being collected and how the recipient may interact with the data. In essence, these operational controls impose a baseline on the usage of the collected data upon which other disclosure controls in this section may build upon. The operational controls include the location in the network where the data is collected, the length of time the data is collected and made available for, the format of the stored data, and the ways the recipient interacts with the data. Here, we describe each of these operational controls, along with their impact on risk and utility. The choices for each of these controls should be made in the context of the risk and utility goals outlined in the previous section.

- **Location.** One way to inherently restrict the type of information collected from computer networks is to consider the placement of monitoring devices and the traffic they collect. The notion of location in this context could refer to either the physical location within the network infrastructure or the portion of the address space being monitored. Depending on the locality choices, monitoring devices that collect network data for sharing purposes may collect detailed, user-specific communications, communications to only external entities, or even strictly malicious traffic. Generally speaking, the provider can choose to implement monitoring on client workstations, within local area networks (LANs), at network gateways, or within portions of the network allocated specifically

for measurement purposes (e.g., darknets, honeypots). Each of these location choices has their own unique trade-off between the risks involved with the data and the general utility of the data for sharing purposes. The most detailed (and potentially risky) data would be collected on client workstations where all user activity can be monitored, and consequently can be most easily attributed to an individual. As we move from workstations to LANs, the connection to specific individuals becomes more difficult due to the diversity of traffic and the potential for obfuscation by network devices (e.g., network address translation), however this traffic still includes intra-organization communication. Collection at network gateways again increases the difficulty in extracting user-specific traffic and reduces the possibility of capturing sensitive internal communications since all traffic is, by definition, bound for external organizations where monitoring practices may also exist. At the same time, the gateway data can still reveal general properties of the publishing organization and their relationships with other entities. Finally, the safest location for collecting data would be within specified network address ranges that contain only malicious or unsolicited traffic, where no legitimate user communication should ever exist.

- **Time.** The time-related properties of the data collection effort affect risk and utility in two ways. First, the length of the data collection effort controls the breadth of activities captured. Short collection windows during low-utilization periods, for instance, necessarily have a smaller chance of capturing sensitive information than longitudinal data collection efforts. Similarly, long-term network data provides a higher-fidelity look into general properties of the network traffic, while several short collection periods will likely only provide useful information about very specific activities within those windows. Second, the length of time before the data is made available to the user, or the data age, may also mitigate certain risks due to changes in the network, its users, and the context of the potentially sensitive information contained in the data. Again, increasing data age is at odds with certain utility goals that may require timely access to the data. In combination, the length and age of the network data can be used to tune the collection process from collection of long-term network activities that are released long after their potential for risk has dissipated to short windows of activity that are released immediate for use in network and security operations.

- **Format.** Another operational control available to data providers is the format that the data collected from the network is initially stored as. Here, the choices range from detailed packet traces that may contain application-layer content to flow longs containing only summary information about network traffic to summary statistics that capture only the broadest notions of the activity on the network. Certainly, there is an obvious and strict trade-off here between the utility and risk of the captured data. Packet traces, for instance, may contain a variety of sensitive and personally-identifiable information within the packet payloads that is difficult to effectively mitigate. On the other hand, packet traces have strictly better utility than flow logs or statistics since they can, in fact, be used to generate those other types of data formats and do much more.

- **Access.** The final operational choice available to data providers lies in how the recipient will be allowed to interact with the collected data. The choices may span from completely unrestricted access to the data to strongly controlled access via controls on physical or virtual environments. Unrestricted access obviously allows the recipient a great deal of flexibility in performing analysis on whatever equipment they see fit, while also minimizing the administrative burden on the data provider. At the same time, this access choice does little to mitigate risks imposed by recipient usage of the data or their poor security practices, particularly because true open access data sharing imposes no auditing on the recipient's use of the data. Another option might be mediated interaction with the data that is controlled via virtualized environment or software-based registration and access restrictions, as is the case with the CRAWDAD and PREDICT data sharing environments. In a so-called virtually mediated environment, the provider has the ability to audit and control the recipient's activities, though it may still be possible for the recipient to extract some types of (previously unknown) sensitive information and reveal it to the public. A stronger form of mediation would be a physically mediated environment, such as those imposed by the U.S. Census Bureau [27, 26], where recipients are physically restricted to certain equipment and locations, and where any usage of the data may be subject to review by the provider. Of course, increased levels of mitigated access impose potentially significant overhead on the provider, but also provide better risk mitigation properties without compromising the utility of the data.

3.2 Data Controls

Once the data has been collected, there are a number of data-altering controls that may be applied to try and remove specific instances of sensitive information and mitigate other types of risk. Unlike the operational controls, the data-altering controls cannot be chosen solely based on perceived risk and utility objectives. Instead, some cursory analysis

of the data is necessary to first identify what exactly is contained within the specific network data being considered, and how risk and utility objectives manifest themselves. In other words, data controls are decided for each dataset independently based on its contents rather than generically for an entire data sharing effort. To understand why we must perform at least a cursory analysis of the data, we need only point out that the level of sensitive information in a given network dataset is entirely a function of the activities of the users and devices on the network at the time of collection. Consequently, some datasets may require significant data transformation to sufficiently mitigate the identified risks, while another dataset collected with the same operational controls may require no changes at all.

The data controls presented in this section can be considered to be transformation functions that act as basic building blocks that are applied to the data in a variety of ways to achieve a certain goal for risk mitigation or to preserve utility. As such, the transformations can be (and often are) combined to tightly target only the most sensitive data for removal while otherwise maintaining utility. Moreover, these transformations can be applied to individual columns and rows, or combinations thereof. In some cases, it is most appropriate to apply the transformation to only specific cells or even specific portions of the value contained within the cell. Below, we describe six basic transformation methods, their known weaknesses, and their impact on risk and utility.

- **Deletion.** The most basic transformation that can be applied to network data is to simply delete unwanted or sensitive portions of the data. The deletion transformation may be applied in a variety of ways to entire columns, rows, individual cells, or any combinations thereof. In practice, there are several ways to achieve deletion, including explicitly removing the relevant pieces of data, replacing the data with a fixed value that indicates deletion, or replacing the data with a value generated uniformly at random from among all potential values of the field. From a risk perspective, deletion is the best way to mitigate against the presence of obviously sensitive information found within the first-order network data, such as names, social security numbers, passwords, and detailed location information. It is important to note, however, that it may be possible to infer the true value of deleted values based on the surrounding data or properties found within second-order statistics of the data. Of course, the strong risk mitigation is balanced by complete destruction of utility for both the specific values removed and their relationships to other entities in the data. As one example of this strategy, consider removing the presence of a specific network device by deleting all related rows from the network dataset. The obvious communications of the device have been removed, as has the ability to explain the devices impact on the remaining network traffic. At the same time, that remaining network traffic may, in fact, have implicit indicators that the device was present and so it may be possible to infer some general information about the device that was removed.

- **Aggregation.** The aggregation transformation combines several records from the original network data into a single derivative record based on a set of key fields or properties. These derivative records may contain fields from the original data or new summary fields that store counts, averages, and other second-order statistics about the object represented by the derivative record. For instance, the data provider may use aggregation to combine individual packet trace records into flow logs, or flow records into summary records that describe the general properties of web pages or workstations present within the data. This transformation allows the data provider to reduce the granularity of the data available to the recipient, just as the format operational control did. The key difference, however, is that aggregation transformations can be applied dynamically and can be adjusted to changing recipient needs, while the format controls sets a baseline granularity that cannot be improved. As the aggregation increases in generality, fine-grained user and workstation information is likely to become more difficult to extract, though certain organizational properties will still exist even at the highest levels of aggregation.

- **Generalization.** In some cases, it is desirable to generalize specific values found in the network data into broad classes. This is similar to the concept of aggregating records based on certain properties or values, however, in this case we do not actually combine the records, but instead simply replace the value with the identifier of the class it belongs to. As an example, we may use generalization to replace complete IP addresses with just the network portion to create classes relating to each of the subnets in the data. Notice that the individual records and their baseline level of granularity remain unchanged. The generalization transformation helps to blend together traffic from several individual records into a single general class that may make it more difficult to associate a given row (or its values) to a specific user or workstation. By maintaining the original records (rather than aggregating them), the generalization transformation enables broader use of the data, but also opens up the possibility that the remaining, unaltered values in the rows can be used to distinguish records from different entities despite the generalized classes.

- **Pseudonymization.** Perhaps the most well-known data transformation used for disclosure control purposes is to replace original values in the data with randomly generated but consistent pseudonyms. That is, each instance of the value is replaced with the exact same pseudonym value. The prefix-preserving IP address pseudonymization scheme proposed by Xu et al. [28], for example, replaces each IP address with a new address where the prefix relationships among the addresses remain in tact. The use of pseudonyms has the benefit of allowing more detailed analysis of the rows associated with individual entities, like users or workstations, without explicitly revealing their identities. As with many of the other transformations, though, the other values within the original data, and even the second-order statistics for each of the entities, may reveal information about the identities despite the uses of pseudonyms, which is typically known as a re-identification attack [5, 20].

- **Perturbation.** For some types of numeric values and second-order statistics, it is possible to add uncertainty about the original values by randomly perturbing the original values. Oftentimes, perturbation is achieved simply by adding a random value taken from a particular type of probability distribution to the original value in order to ensure that the amount of change is restricted, but still adds sufficient levels of uncertainty about the original value. As an example, perturbation-based transformations may be used to slightly alter the value of timestamps found within the data to prevent identifying information about workstations from being revealed [13]. Differential privacy also uses perturbation methods to provably hide the effects of individual records (e.g., packets or flows) on second-order statistics of the data. Although it may be tempting to apply perturbation methods to values that are represented as numbers, like ports or IP addresses, these changes will have little semantic meaning and likely do little to protect the original values. Therefore, the perturbation transformations are obviously limited to values or second-order statistics that can be meaningfully mapped into the set of integers or real numbers, and the utility (and risk mitigation) provided by perturbation is contingent entirely upon the chosen probability distribution that bounds the amount of noise added.

- **Synthetic Data.** The use of synthetic data allows the provider to replace individual values in the network data with new ones generated according to a probabilistic model that captures the general properties of that type of value. One example would be to calculate the probability distribution over the port numbers seen across the entire network dataset, and then replace the original ports with ones sampled from that distribution. The probabilistic models can be refined to take into account other values within the same row. To expand on the port numbers example, we might create a distribution of source ports that is conditioned upon the value of the destination port. When we apply synthetic data transformations, the original source port is replaced by one sampled from the distribution of source ports conditioned upon the destination port found within the same row. The specifics of this transformation can be quite technically involved, and the quality of the output is dependent entirely on how the probabilistic model of the data is defined. In fact, the complexity of the probabilistic model provides another obvious utility-risk tradeoff since improved fidelity (i.e., more conditional fields) necessarily reveals more fine-grained information about the original data, while broader models are likely to capture only basic properties of the data. Due to the technical depth of synthetic data generation techniques, interested readers should reference appropriate surveys to understand the potential applicability of synthetic data in mitigating identified risks [22].

3.3 Policy Controls

While operational and data controls help the data provider set concrete boundaries on the recipient's access to the collected data, they may be unable to completely hide the presence of sensitive information within the shared data. In practice, sensitive information may remain despite the application of the technical disclosure controls discussed above because of a desire for enhanced utility or new network technologies that carry unexpected types data. Once those controls are applied and the recipient gains access to the data, the provider has relatively little direct control over its resultant use or the security environment of the recipient, which may lead to compromise of the data. The most effective way to mitigate against these ex post facto risks is to implement contractual and policy-oriented controls on the recipient through the use of data sharing agreements (e.g., contracts, memorandum of understanding). In effect, these policy-based controls should clearly communicate the risk mitigation and utility goals identified during the data sharing process, and specify incentives and penalties to ensure recipients conform to those goals.

Generally speaking, the policy controls can be used to address three aspects of recipient interaction with the data: *access*, *use*, and *secondary disclosure*. Policy-based controls on access require the recipient to conform to the specified technical (or other) methods for interacting with the data, including those specified by the operational controls. These types of access controls prevent the recipient from finding alternative means of gaining access to the data, which may

be easier for the recipient but impart more risk on the provider. Restrictions on data use cover the recipient's use of information gleaned from analysis of the shared data, including pre-publication review of results and restrictions on actively subverting disclosure controls to learn sensitive information. Finally, controls on secondary disclosure specify how the recipients responsibilities in the safe keeping of any of the shared data they access, or information obtained from that data. The controls on secondary disclosure may specify specific security measures, auditing procedures, and reporting processes to mitigate the impact of lapses in security.

Overall, these controls can mitigate many of the most fundamental concerns associated with sharing data with third parties without affecting the utility of that data significantly. However, the utility benefits of policy-based controls are balanced by potentially significant effort on the part of both the provider and recipient to monitor and enforce the restrictions. Moreover, the policy controls can only incentivize good behavior on the part of the recipient and may not deter malicious entities. Consequently, the policy controls are less focused on prevention of disclosure and more on mitigation of harmful situations that are difficult to control via technical measures.

3.4 Disclosure Control Profiles

From the above categories of disclosure control, the data provider should choose a combination of controls that they believe sufficiently mitigates the presence of sensitive information identified within the data. For each of these pieces of sensitive information, the provider creates a disclosure control profile that contains the set of controls they have applied to mitigate the risk, along with the color on the spectrum representing the severity of the chosen controls. There are potentially many ways of applying disclosure controls to mitigate risk mitigation. For instance, it may be possible to mitigate the risk of revealing sensitive information related to customer names by either thoroughly scrubbing the data or by implementing a series of weaker technical controls along with strong policy safeguards. The former solution clearly provides a better chance of concretely mitigating the risk of the sensitive information from being revealed, while the latter ensures that some level of utility surrounding use of names is maintained, perhaps due to a particular utility objective.

Although the respective disclosure control spectrums provide a notion of the relative risk/utility trade off, it is important to note that in all cases the purpose is to give the provider a means of understanding their options for disclosure control and forcing them to provide justification for their choices. In the absence of all-encompassing utility or risk definitions, the choices made and their anticipated impact on risk are subjective judgments made by the provider with sufficient understanding of their data and the relative benefits of the controls described in this section.

To this end, the disclosure control profiles are simply a concrete way for the data provider to illustrate their assumptions (i.e., where their chosen controls fall on the risk-utility spectrum) about how well they are m eating their utility goals and mitigating their identified risks. Therefore, the provider must go one step further in this phase of the framework, and map the assumed disclosure control efficacy back to the risk and utility profiles. For instance, if the original utility profile created in phase one reflects a desire for a particular functionality (e.g., botnet detection), then they must decide how the IP address controls profile impacts that specific utility goal and adjust the utility profile accordingly. The original risk and utility profiles are adjusted by the provider in light of the changes described by the disclosure control profiles, which creates *modified risk and utility profiles* that are used in the final phase of the framework to assess the efficacy of the choices made for this data sharing effort.

4 Assessing Disclosure Control Impact

The third and final phase of the disclosure framework qualitatively describes the impact of the chosen disclosure control methods with respect to the stated utility objectives and disclosure restrictions. This step allows all stakeholders– providers, recipients, third-party oversight authorities– to assess and demonstrate the reasonableness of a provider's choice of disclosure controls based on adherence to transparent and relevant performance criteria – utility and risk. The impact assessment serves as an accountability mechanism for addressing well-defined risks and utility objectives, as well as for resolving more conjectural risk and indeterminate outcomes.

There is no authoritative threshold or referential standard for disclosing network data compared to health data vis-a-vis the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, the familiar Federal law that defines type of entities that may receive personal medical data and for what purposes the data may be applied. It defines two methods to create *de-identified data* which consequently permit the Provider to disclose health information without the subject's prior consent: the removal of certain "identifiers" (safe harbor test), and the certification by a person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable (statistical test). That standard's relevance and applicability

as a model for network data is debatable at best. Notwithstanding HIPAA's inaptness to entities sharing network data, it focuses only on identifiability from a personal privacy value standpoint (ignoring data confidentiality risks), characterizes data from a human-level privacy perspective (failing to address the tenuous mapping between network level data and data sensitivity), and is wrought by increasing debate over its efficacy among stakeholders within that community [15, 21].

The implicit justification for all risk frameworks rings true for network data disclosure risk. Data disclosure defies black-and-white, pre-defined formulas as to whether provider A can share dataset M with recipient Z without concern for legal risk. The fact-specific nature of legal risk assessment, where a myriad of case variables (*data sharing scenarios*) intersect with indeterminate application and interpretation of rules, defies enumeration of inputs and unassailable prediction of outcomes. Defensible decisions often turn on whether actions are "reasonable". This begs the question of what is the test for *reasonableness* when disclosing network data? The answer is the *process* to understand, apply, assess and evaluate the choices comprising a disclosure decision, more so than the output of a disclosure decision.

This framework helps providers meet the reasonableness standard by virtue of its composition of widely accepted factors (risk, utility, controls, evaluation), its functioning (the coordination between the factors), and its allowance for subjective implementation based on objective performance criteria. It embeds adaptability to deal with yet unknown risks/scenarios, and enables providers to observe risks in parallel with intended outcomes and to respond with multiple solution options. This framework is not engineered to produce perfect, all-encompassing solutions, but to support defensible and pragmatic data disclosure actions. As such, the process established by this framework allows the provider to be prepared *for* rather than *against* risk. Also, its value-driven approach– legal rights and obligations, utility objectives– further influences internal and external determinations of reasonableness.

4.1 Residual Data Risk

The threat context informs the final disclosure profiles by adjusting the modified risk profile that resulted from the Phase 2 application of the disclosure controls.

The "residual risk" is the data risk profile that remains after the provider qualitatively maps the modified risk profile created in Phase 2 back to the data sensitivities identified in the initial Phase 1 risk profile. This evaluation illuminates the extent to which the notionally-applied disclosure controls meet the authoritative data restrictions and obligations from our initial phase. For example, if a pseudonymization is applied to packet header data to obfuscate IPA because the provider is prohibited by law (e.g., ECPA) from disclosing addressing information, the residual risk is minimal. If, however, the provider is obligated under contract to not disclose data that is linkable, via public records or other reasonably available external records, in order to re-identify the data, the residual risk would be high because the de-sensitized data may be vulnerable to reversal or inference attacks that de-anonymize the IPA. If the provider concludes the the residual risk is not acceptable, it can iterate back to Phase 2 and apply alternative disclosure controls until it reaches a satisfactory residual risk.

4.2 Utility

Our performance evaluation entails a similar loopback analysis for the other prong of our framework- utility objectives. This impact is gauged by qualitatively mapping the low level pros and cons associated with the specific technical and non-technical controls from Phase 2 back to the utility objectives identified in the initial Phase 1 risk profile. We conceptualize the impact of a given disclosure control on a utility objective according to how much, if any, of the desired utility is not achievable after the control has been applied. In other words, utility is described proscriptively in relation to the original intended outcome. For example, if the provider desires to perform detailed topology and geolocation analysis on packet header data and a pseudonymization disclosure technique is used, despite preservation of the subnet structure, that method fails to achieve the objective to maintain institutional granularity and therefore has a negative impact. As with residual risk, the provider can tune the utility to reach an acceptable level of efficacy for the data sharing effort.

5 Case Study

To illustrate how this disclosure control framework may be applied, we walk through a theoretical data collection and sharing scenario inspired by the recent events related to Operation Ghost Click, an FBI investigation into an international cybercrime ring that infected millions of computers, that involved remediation of the DNSChanger malware

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

Figure 2: Initial utility profile for theoretical ISC DNS data sharing effort.

[12]. This malware redirected the Domain Name System (DNS) queries of infected victims from legitimate name servers to malicious ones run by an Estonian company named Rove Digital which distributed the malware. Those malicious name servers redirected victims to advertisers associated with Rove Digital, rather than pointing to the actual web page being requested. As part of a court order against Rove Digital, the Federal Bureau of Investigation (FBI) seized the malicious name servers located within the United States. In order to ensure that infected computers maintained Internet access after the seizure, the Internet Systems Consortium (ISC), a non-profit organization responsible for developing and maintaining core Internet technologies, was tasked with running temporary DNS name servers for the victims in place of Rove Digital's servers.

The circumstances of ISC's involvement in maintaining the DNS infrastructure for infected clients presents an interesting opportunity to advance network and security research with data collected from those servers. In particular, by sharing information about the queries made to the ISC name servers and the IP addresses of the clients who make those queries, we may learn more about the prevalence of the infection, how it spread, and what other viruses or malware may be on the infected computers. Furthermore, the more general information about DNS protocol exchanges between clients and servers would provide additional operational insight into the servers' real-world functioning and help researchers to develop better models of user activities. Balancing these opportunities are risks involving the disclosure of identifiable and behavioral information about users, as well as the the exposure of their computers' infection status and potential for re-compromising those computers by targeted attacks. We next steer through each of the three phases of our disclosure control framework and provide a detailed description of how and why we make certain choices within this data sharing scenario.

5.1 Utility and Risk

Our first step in this case study is to outline the primary utility goals associated with disclosing the DNS queries of infected hosts, and the risks associated with those goals. As touched upon earlier, there are three primary goals that may be associated with collection and subsequent release of ISC's DNS query data:

1. **Understanding DNSChanger infection properties:** Examining client IP addresses, their geographic location, and their associated volume of queries over time to better understand who was affected by the malware and how quickly the infection is being remediated.

2. **Analysis of malware behavior:** Examining the contents of DNS requests or other DNS properties that indicate specific behaviors and operation of DNSChanger or other malware.

3. **General DNS modeling:** Using complete transcripts of DNS queries for individual clients to better understand DNS query volumes, the distribution of domain names requested, and other properties to help network researchers evaluate DNS-related technologies.

The first goal only requires access to victim client IP addresses, while the other two require DNS traffic that indicates the web sites and other services the clients use. The third goal further requires that the DNS queries be grouped by client IP and that those IPs maintain the geographic and network properties (e.g., originating organization) in order to provide the most realistic modeling information possible.

From these general goals, we can extract specific utility requirements from among the six categories of the utility profile (Figure 2). The utility profile can be summarily described as a desire to provide a limited community of established security and network researchers access to longitudinal DNS data containing highly-detailed records of traffic from infected clients to the ISC DNS servers. These utility objectives do not impose any particular requirements on how quickly the data must be released to the researchers, but does require a relatively open-ended research path with respect to functionality and research output, particularly when modeling DNS client traffic.

After having articulated the desired utility, we evaluate the potential areas of risk for such a data sharing effort. In particular, we focus on the two potentially sensitive types of data being shared: client IP address and the domain name queried by the client (i.e., qname). Another informational component of the DNS transactions is dismissed from further assessment since it is generally available to the public (e.g., accessible by any Internet user) and does not implicate identification of individuals. The important part of this assessment is understanding that sensitivity arises from the connection between DNS activity, infection status, and the client identifier (i.e., IP address). As described in Section 2, the appropriate way to gauge the risk is for the technician and legal counsel to engage in a question-and-answer dialogue that helps translate the technical details of the data sharing goals into terms that sensibly map to risk factors, as illustrated below.

- **What:** The data being considered for release is a simplified set of packet trace data, containing only the client IP address and all fields from the DNS query and response.

  The **client IP address** can be characterized as an extension identifier that may be associated to one or more individuals (i.e., computer users) and that may change frequently over time.

  The **DNS data** denotes machine-level transactional activity, and generally consists of information that is openly available to all users of the DNS system. The query name field (i.e., qname), however, is not publicly-available. It indicates the familiar name of the web site or other service being accessed by the user, and from which it may be possible to infer the preferences or general web browsing behaviors of the users who visit the site.

- **Who:** ISC and the researchers are the data provider and recipients, respectively, and are considered the data sharing *participants*. Their impact on risk assessment is influenced by what we colloquially refer to as important stakeholders in this scenario– the United States government and the the users/owners of the infected client systems. The nature of these participants in relation to the stakeholders are as follows:

  ISC acts as the designated replacement DNS service provider to the infected client victims of DNSChanger, by providing an electronic communications infrastructure service under the authority of a Federal court order related to the seizure of Rove Digital computer equipment that directed it to work under the guidance of the U.S. government. Due to their role as a DNS service provider, ISC access to the victims' communications was necessary incident to the rendition of its service. The **infected clients** likely have no reasonable understanding that their Internet activity and information is being proxied by a party other than their Internet service provider, yet their rights and interests associated with the traffic impact the risk assessment.

  The **data recipients** consist of reputable researchers in computer security or network measurement who seek to expand general understanding of DNSChanger infections, malware properties, and/or the operational reality of DNS server infrastructure query workload. These researchers are likely to have a high level of knowledge, skills, and abilities with this data. However, their motivation for uncovering sensitive information is relatively low since their goals (e.g., malware or DNS-related research) are not dependent on extracting identifiable or confidential information related to specific Internet users.

- **Why:** The purpose of the data sharing effort is to promote understanding of malware infection patterns and real-world DNS usage, which will in turn improve antivirus and network technology research and development efforts. Specific uses of the data include:

  Analysis of client IP patterns of communication to DNS servers, geographic locations, and changes in communication over time to understand DNSChanger-related activity.

  Evaluation of DNS query and response properties including query names, answer records, and other DNS fields for evidence of additional infection or previously unknown malicious activity, potentially relating to DNSChanger malware.

  General examination of DNS query and response fields for the purpose of understanding how the DNS service operates in a live deployment, and building better models of DNS activity for evaluation and testing of new technologies.

The feedback from the legal counsel is a risk profile and justification for each of the sensitive pieces of information and derived from relevant risks within each of the three categories above. The a summarization of the resultant risk profile for our data sharing scenario is shown in Figure 3. In general, this risk profile indicates that client IP address

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

Figure 3: Initial risk profile for theoretical ISC DNS data sharing effort.

is a relatively risky piece of information to share without significant changes due to ethical restrictions to safeguard clients and their organizations, as well as contractual obligations to the U.S. government to restrict the disclosure of confidential or identifying information. The DNS data is somewhat less risky since, in and of itself, it is not uniquely identifiable or inherently confidential, however the same ethical and contractual restrictions exist in limiting potential disclosure[9].

5.2 Disclosure Control Choices

The utility and risk profiles provide concrete descriptions of the requirements and constraints placed upon our choices of disclosure control. Now, we must translate those requirements and constraints into alterations to the data, and functionality that must be maintained for the goals of our data sharing scenario. Specifically, the risk profile indicates that the primary concerns lay in ethical obligations to not expose the IP address of known-infected clients to potentially malicious entities, and private agreement requirements to protect the privacy and confidentiality of the end user. Since the DNS data isolated from other context is relatively low risk, we can focus our disclosure control efforts on methods that limit access by malicious recipients and ensure that client IP addresses are not easily linked to Internet users.

From the available controls across the three disclosure control categories, the following set of techniques are pertinent to mitigate the identified risks. Initially, we choose these techniques independently based on potential applicability, and later we decide how to best combine them to achieve our goals.

Time: Potential harm from exposing the identity of infected clients is reduced by waiting to release the data until after remediation procedures are complete. This aging process may also make it more difficult to connect old DNS activity to particular users or clients due to changes in DNS infrastructure and user behavior.

Access: It may be possible to provide mediated access to the data through a virtually-mediated system that limits access to authenticated researchers. This will not prevent the leakage of sensitive information, but it restricts that leakage to a set of researchers with little or no malicious intent.

Deletion: The client IP address or domain name can be completely removed, though this clearly renders all utility goals unattainable.

Aggregation: Distributions of unique client IPs or DNS query names aggregated by geographic locations to provide some location information with much coarser detail. These counts can be further aggregated by time to examine trends overtime, or they can be aggregated by other fields to examine broad trends in DNS activity.

Generalization: IP addresses and domain names may be generalized to reflect general network or organizational information. IP addresses can be truncated at an appropriate subnet to only provide information about general geographic location or the Internet service providers hosting the infected clients. Likewise, domain names may be truncated by removing specific subdomains that may indicate particular services or user behaviors.

Pseudonyms: The client IP addresses and domain names in the released data may be replaced with pseudonyms to hide the original sensitive and confidential information while still allowing researchers to separate queries by client. Geographic locality and other network-related information will be removed, however.

[9] In reality, the private agreement between the U.S. government and ISC required that all 'content' be removed. In order to provide a broader data sharing scenario, we ignore this directive.

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

Figure 4: Profile of chosen disclosure controls for DNS data sharing.

Policy: In addition to technological access controls, we can further restrict and disincentivize malicious activities through the use of simple policy controls that require researchers to be associated with established research organizations and place penalties on misuse of the data.

From this set of potential disclosure controls, there are relatively few that will meet the needs described in our utility profile, particularly the need for unfettered access by researchers and detailed records of DNS activities. It is possible to address some of the utility requirements, such as monitoring the infection over time, by creating aggregate statistics and counts of the number of unique client IP addresses that connect to the DNS servers each day. It may also be possible to extract some general DNS traffic information by providing counts for the number of unique domain names queried or even the top-10 most queried domain names. The aggregated statistics can be further protected by using a perturbation strategy, such as differential privacy, to hide the effect of any particular client on the statistics. However, these aggregation methods necessarily limit the functionality of the data and almost certainly makes the goal of general DNS user modeling impossible to achieve.

To push the boundaries of this data sharing and attempt to meet all of our goals, we instead propose to generalize client IPs to the largest subnet that corresponds to the organization that owns the IP address. The remainder of the IP (i.e., the host portion) is replaced by a consistent, linkable pseudonym. In effect, the client IP can be used to learn information about general geographic areas and queries can be separated by the independent clients that made them, but the specific client IP address is not easily learned. Additionally, we wait to release the data until after the DNSChanger remediation period has ended to minimize the potential for those infected clients to be exploited again. Finally, we enact virtual mediation methods with policy controls to restrict access to established researchers and reduce the potential temptation to use the data for purposes beyond those defined in the utility goals. Obviously, the use of mediation and policy controls requires some effort on the part of the data provider, however in this case the effort can be relatively minimal. It may be sufficient to execute a simple Memorandum of Understanding between ISC and the researchers that describe acceptable uses and obligations, and to authenticate the researcher's identity and reputation with community attestation[10] The overall disclosure control profile that summarizes these choices is shown in Figure 4.

One potential problem with this approach is that if a malicious entity who gains access to the data and has significant a priori knowledge about a user or clients typical DNS query activities, it may be possible for them to learn the real identity of that client. This situation is improbable, though, since it requires the attacker to already have significant access to the user's network data through other means. Moreover, the potential for this type of attack on a large scale with more than one/two users is even more improbable given that the standard DNS query data, which would be used to re-identify the client, is not publicly accessible or easily obtained. This may not be the case for other situations where identifying information about users is available from public and easily accessible sources of information.

5.3 Assessment

To wrap up the case study, we must now re-evaluate our risks and utility goals in light of our chosen disclosure control strategy. Evaluating the change in utility is a relatively simple procedure of comparing the utility profile derived directly from our high-level data sharing goals with the utility profile that remains after application of disclosure controls. As shown in Figure 5, the major changes in utility stem from the loss of some detail due to generalization of the client IP addresses. By and large, however, the chosen disclosure controls perfectly match the utility requirements set forth at the beginning of our framework, including limited access to the data by researchers, unfettered use of the data by those researchers, and functionality related to only DNS and malware research. Despite the application of disclosure controls, we can still track infection status at very fine levels of granularity, attribute those infections to

[10]This model is deployed by the CRAWDAD data repository [7].

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

Figure 5: Modified utility profile after application of disclosure controls.

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

Figure 6: Modified risk profile after application of disclosure controls.

relatively accurate geographic areas, and extract significant information about typical DNS traffic as seen by a large DNS server with diverse client base.

With respect to the risk profiles, we can revisit the pertinent evaluation questions that we asked earlier and describe how the answers have changed due to the selected controls. The reason for the data release and nature of use of the data ("Why") remains unchanged from our earlier analysis since no significant utility changes have been instituted. The participants in the data sharing effort ("Who") also remain unchanged, though our choice of a virtually mediated environment with policy-based controls will decrease the likelihood that a malevolently motivated recipient will attempt to subvert the chosen controls. For researchers, this motivation was already quite low, and the use of these access control methods ensures that those researchers remain partly accountable for the safety of the data. The type of data being released ("What") no longer uniquely identifies clients by IP address and provides a reasonable basis for preventing Internet user activities from being learned. The changes to client IP address, along with the mediation controls and delayed release of the data, limit the ethical risks associated with the data sharing scenario and also meet the requirements imposed by private agreements with the U.S. government. The modified risk profile is shown in Figure 6.

In summary, we have articulated a clear justification for our chosen disclosure controls and demonstrated the reasonableness of our consideration of risk in the context of our decisions and actions to share the data in this scenario. By doing so, we were able to develop a nuanced approach that provides much more utility while simultaneously meeting the requirements of the various laws, regulations, and other risk considerations. The one-size-fits-all disclosure control approach used today, on the other hand, would have removed all DNS data and replaced IP addresses with pseudonyms that have no connection to the real-world locations of the infected clients.

6 Conclusion

Once the provider has engaged the three phases of the disclosure framework: (1) identify the data risk profile and utility objectives; (2) apply the disclosure control(s); and (3) assess the impact on risk and utility in the context of the threat environment, the provider may evaluate whether the overall residual risk and utility is acceptable. If the data risk is not acceptable, the provider can notionally apply additional and/or alternate operational, technical, or policy controls to determine how the risk may be modified. In general, the quantity and quality of the disclosure controls have a direct relation to the data risk, whereas they are inversely proportional to utility objectives– high data risk demands more restrictive controls, and vast utility calls for lower disclosure encumbrances. Balance is achievable when the application of disclosure controls lowers the data risk to acceptable levels and does not modify the utility objectives such that the purpose of the disclosure is rendered unattainable. This framework enables a provider to craft a disclosure strategy along a spectrum of risk thresholds and utility needs according to a repeatable, transparent and evidence-based process. As such, this design embeds a risk management approach to data disclosure that allows providers to demonstrate a documented and reasoned process that can be audited, measured and compared over time, both within and across provider organizations.

References

[1] M. Bezzi. An Entropy-based Method for Measuring Anonymity. In *Proceedings of the* 3^{rd} *International Conference on Security and Privacy in Communications Networks and the Workshops*, pages 28–32, 2007.

[2] CAIDA Data. `http://www.caida.org/data/`.

[3] Bee-Chung Chen, Daniel Kifer, Kristen LeFevre, and Ashwin Machanavajjhala. Privacy-Preserving Data Publishing. *Foundations and Trends in Databases*, 2(1–2):1–167, January 2009.

[4] S. Coull, F. Monrose, M. Reiter, and M. Bailey. The Challenges of Effectively Anonymizing Network Data. In *Proceedings of the DHS Cybersecurity Applications and Technology Conference for Homeland Security (CATCH)*, pages 230–236, March 2009.

[5] S. Coull, C. Wright, F. Monrose, M. Collins, and M. K. Reiter. Playing Devil's Advocate: Inferring Sensitive Information from Anonymized Network Traces. In *Proceedings of the* 14^{th} *Annual Network and Distributed System Security Symposium*, pages 35–47, February 2007.

[6] S. E. Coull, C. V. Wright, A. D. Keromytis, F. Monrose, and Michael K. Reiter. Taming the Devil: Techniques for Evaluating Anonymized Network Data. In *Proceedings of the* 15^{th} *Network and Distributed Systems Security Symposium*, pages 125–135, 2008.

[7] CRAWDAD: A Community Resource for Archiving Wireless Data at Dartmouth. `http://crawdad.cs.dartmouth.edu`.

[8] DShield: Cooperative Network Security Community. `http://www.dshield.org/`.

[9] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating Noise to Sensitivity in Private Data Analysis. In *Proceedings of the* 3^{rd} *Theory of Cryptography Conference*, pages 265–284. Springer, 2006.

[10] EFF Panopticlick. `https://panopticlick.eff.org/`.

[11] The EFF SSL Observatory. `https://www.eff.org/observatory`.

[12] FBI DNSChanger Malware. Accessed at: `http://www.fbi.gov/news/stories/2011/november/malware_110911/`.

[13] T. Kohno, A. Broido, and K. Claffy. Remote Physical Device Fingerprinting. In *Proceedings of the IEEE Symposium on Security and Privacy*, pages 93–108, May 2005.

[14] A. Kounine and M. Bezzi. Assessing Disclosure Risk in Anonymized Datasets. In *Proceedings of FloCon*, 2008.

[15] Latanya Sweeney et al. Comments from Data Privacy Researchers. Accessed at: `http://dataprivacylab.org/projects/irb/DataPrivacyResearchers.pdf`.

[16] LBNL/ICSI Enterprise Tracing Project. `http://www.icir.org/enterprise-tracing/`.

[17] R. Pang, M. Allman, V. Paxson, and J. Lee. The Devil and Packet Trace Anonymization. *ACM Computer Communication Review*, 36(1):29–38, January 2006.

[18] PREDICT: Protected Repository for the Defense of Infrastructure Against Cyber Threats. `http://www.predict.org`.

[19] REN-ISAC Security Even System. `http://www.ren-isac.net/ses/`.

[20] B. Ribeiro, W. Chen, G. Miklau, and D. Towsley. Analyzing Privacy in Enterprise Packet Trace Anonymization. In *Proceedings of the* 15^{th} *Network and Distributed Systems Security Symposium, to appear*, 2008.

[21] Mark A. Rothstein. Is Deidentification Sufficient to Protect Health Privacy in Research? *The American Journal of Bioethics*, 10(9):3–11, 2010.

[22] Joseph L Schafer. Multiple Imputation: A Primer. *Statistical Methods in Medical Research*, 8(1):3–15, 1999.

[23] The Spamhaus Project. `http://www.spamhaus.org/`.

[24] L. Sweeney. k-Anonymity: A Model for Protecting Privacy. *International Journal on Uncertainty, Fuzziness, and Knowledge-based System*, 10(5):557–570, 2002.

[25] Keren Tan, Jihwang Yeo, Michael E. Locasto, and David Kotz. Catch, Clean, and Release: A Survey of Obstacles and Opportunities for Network Trace Sanitization. In Francesco Bonchi and Elena Ferrari, editors, *Privacy-Aware Knowledge Discovery: Novel Applications and New Techniques*, chapter 5, pages 111–141. January 2011.

[26] U.S. Census Data Protection. Accessed at: `http://www.census.gov/privacy/data_protection/`.

[27] U.S. Census Research Data Centers. Accessed at: `http://www.census.gov/ces/rdcresearch/`.

[28] J. Xu, J. Fan, M. H. Ammar, and S. B. Moon. Prefix-Preserving IP Address Anonymization: Measurement-Based Security Evaluation and a new Cryptography-Based Scheme. In *Proceedings of the* 10^{th} *IEEE International Conference on Network Protocols*, pages 280–289, 2002.