

DENIABILITY TOOLS ARE ESSENTIAL TO STRONG PRIVACY

by Glen Slade
glen@stegostik.co.uk

Contents

1. Executive summary.....	2
2. Privacy framework.....	3
3. Privacy tools.....	5
4. Deniability.....	9

© 2010 StegoStik Limited
www.stegostik.co.uk



1. Executive summary

The importance yet vulnerability of personal privacy have been highlighted by many contemporary factors. The socialisation of the Internet is transforming the way that individuals make data available. Governments are achieving greater capabilities in monitoring populations. And companies are still struggling to protect data against human fallibility and malicious attack.

This paper demonstrates that the deniability of data and communication is an essential tool in the quest for strong privacy, particularly in the context of malicious adversaries and incompetent third parties.

Using Daniel Solove's privacy taxonomy as a reference point, the paper considers the measures available to a data subject to protect her privacy. Focusing on Solove's 'information collection' group, deniability tools are shown to make a unique contribution to protecting the privacy of both static and active data.

Furthermore, by facilitating the creation of multiple aliases and the 'watermarking' of data, deniability tools can confer some increase in protection from the potential harms of Solove's 'information processing' and 'information dissemination' groups.

The tools available today for deniable file systems, particularly StegoStik and Truecrypt, lack operating system integration but are practical enough for routine use. In contrast tools for deniable communication, such as Tor or Freenet, impose too great an overhead for widespread adoption; however, integration with a deniable file system would create new potential for speed and privacy.

StegoStik's uniquely powerful patented deniable file system can form the foundation for a new generation of user-friendly systems that have strong privacy built in. This requires integration with an operating system then adaptation of browser technology. While these are not small tasks, they are clearly tractable and stand to create a unique and highly valuable global consumer platform.

Chapter 2 builds the framework for this analysis, combining Daniel Solove's privacy taxonomy with a structured representation of a data subject's environment. Chapter 3 documents the privacy tools available to a data subject for each element of the analysis framework. Chapter 4 concludes by summarising the contribution of deniability to privacy and considering the future for such tools.

2. Privacy framework

Understanding privacy

In his book 'Understanding Privacy', Daniel J Solove investigates the concept and value of privacy in detail. Based on his analysis of existing privacy theories and the challenges of working with them, particularly in a legal and policy context, Solove concludes that the nature of privacy is multi-faceted.

He then proposes a sixteen-part taxonomy of privacy that can serve as a checklist of factors to be considered in different situations (as opposed to the elements of a monolithic privacy concept). Solove's taxonomy has four groups and sixteen subgroups; their headings are explained in further detail when they are considered in the following chapter.

Solove's taxonomy

Information collection	Information dissemination
<input type="checkbox"/> Surveillance	<input type="checkbox"/> Breach of confidentiality
<input type="checkbox"/> Interrogation	<input type="checkbox"/> Disclosure
Information processing	<input type="checkbox"/> Exposure
<input type="checkbox"/> Aggregation	<input type="checkbox"/> Increased accessibility
<input type="checkbox"/> Identification	<input type="checkbox"/> Blackmail
<input type="checkbox"/> Insecurity	<input type="checkbox"/> Appropriation
<input type="checkbox"/> Secondary use	<input type="checkbox"/> Distortion
<input type="checkbox"/> Exclusion	Invasions
	<input type="checkbox"/> Intrusion
	<input type="checkbox"/> Decisional interference

Solove deliberately avoids advocating any particular policy or law to address the privacy issues he identifies. The reader (or legislator) is left to weigh the factors for and against protection and invasion of privacy.

An extended framework

Law and policy may be used to establish what is or is not an "appropriate" degree of privacy, notwithstanding Solove's observation that this will vary between cultures and over time. Yet for data subjects and data holders the question remains of what to do to protect privacy and avoid liability.

To draw out the issues in achieving personal privacy, this paper notes four dichotomies relating to the data subject's predicament:

- Privacy and security.
- Physical and virtual.
- Stasis and activity.
- Friendly and adversarial.

None of these dichotomies is clear cut, but they help identify important considerations.

Privacy and security

In seeking to understand the privacy of a data subject, Solove's taxonomy actually analyses how the actions of data holders and other parties may harm the data subject. In the context of tools to protect privacy, it is important to distinguish between those employed by a data subject to preserve her privacy and those employed by a data holder to ensure security of private data. This paper focuses on the former.

Deniability Tools are Essential to Strong Privacy

Physical and virtual

The information that may be collected, processed and disseminated by a data holder can pertain either to a data subject's physical existence, such as her skin colour, or her virtual existence, meaning any electronic data or entity controlled by her such as a data file or chat room alias. This paper will consider the physical but concentrate on the virtual world.

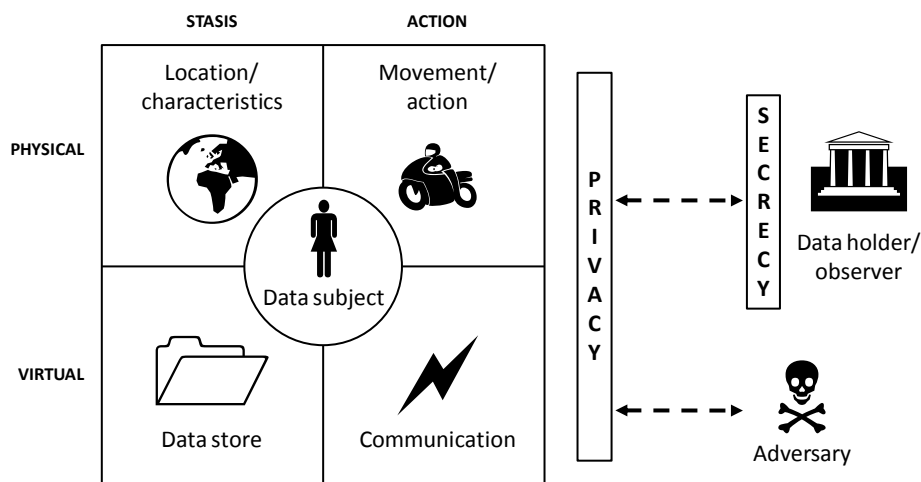
Stasis and activity

From the perspective of the data subject, it is important to distinguish between static data, such as the location of a birth mark (physical) or files on a hard drive (virtual), and active data including visiting a friend (physical) or sending an email (virtual).

Friendly and adversarial

In the real world, a data subject will experience both friendly and adversarial interactions with third parties. She may choose to rely on legal recourse against adversaries or deploy further privacy measures to seek protection against illegal activity.

These dichotomies are summarised in the following structured representation of a data subject's environment:



The data holder has been marked explicitly as also being an observer to indicate that their role is not passive. The adversary is shown separately to highlight that there can be no expectation of secrecy in that case.

Privacy tools analysis

The following chapter reviews the elements of Solove's taxonomy against the structured representation of the data subject's environment. For each aspect it will consider what privacy tools are available and the potential role and contribution of deniability.

3. Privacy tools

Overview

Before considering the 16 detailed elements of Solove's taxonomy, it is helpful to consider how the four high level groups relate to a data subject and her pursuit of privacy.

Information collection is the group that involves obtaining data from the subject. It is therefore the most critical group in relation to protecting privacy, since once data has left the data subject it is beyond her control.

Nonetheless, in the context of information processing and information dissemination, Solove's second and third groups, it is possible for the data subject to have taken measures to mitigate some of the potential privacy harms.

The fourth group, invasions, relates to privacy harms that do not involve information and will not be considered further in this paper.

Information collection

Solove defines the two subgroups of information collection as follows:

- ❑ *Interrogation* consists of the various forms of questioning or probing for information.
- ❑ *Surveillance* is the watching, listening to or recording of an individual's activities.

In terms of a data subject's environment:

- ❑ *Privacy and security*. Information collection is specifically about privacy since the data are not yet obtained by a data holder/observer/adversary.
- ❑ *Physical and virtual*. These are considered within each subgroup below.
- ❑ *Stasis and activity*. These correspond broadly to the interrogation and surveillance subgroups respectively.
- ❑ *Friendly and adversarial*. These are considered within each subgroup below.

The central aim of privacy tools is to enable a data subject to release information about herself, including any data held, at her discretion and on her terms.

Interrogation

In the physical world, a data subject may wish to protect against third parties establishing her existence, bodily characteristics or personal details including location. Depending on the situation this may require a combination of:

- ❑ *Isolation*. Every contact with another person or system is a clue to existence and typically reveals certain physical attributes.
- ❑ *Disguise*. Concealing or modifying physical attributes before interacting with other persons or systems can achieve anonymity or impersonation.
- ❑ *Withholding data*. By declining to answer questions during an interaction, the data subject can protect such information as is not directly observable.
- ❑ *Giving false data*. If withholding answers is not an option or would create suspicion, the data subject could supply false answers.

Faced with a malicious adversary, these protections become harder to achieve. Isolation becomes 'hiding'; a disguise must be robust against close scrutiny although it is likely to fail against physical assault. Withholding data may become impractical under duress, which places greater emphasis on the ability to give false data. The success of the latter strategy depends critically on the plausibility of the 'cover' story; adversaries could deploy drugs or fMRI to try to access the truth.

Deniability Tools are Essential to Strong Privacy

In the virtual world, interrogation relates specifically to stored data. (Data in transit represents a communication or action, which is considered in 'Surveillance' below.)

A data subject's options for protecting against a third party establishing the existence or contents of stored electronic data have parallels with the physical case:

- ❑ *Isolation.* Holding sensitive data on systems that are not directly connected to a network prevents hacking.
- ❑ *Disguise.* Encryption with a strong algorithm and strong key renders data unreadable. An adversary may use duress or covert observation to try to obtain the key.
- ❑ *Withholding data.* A data subject can decline to supply data. Against an adversary, the data would need to be deniably hidden for this to be successful.
- ❑ *Giving false data.* In some situations a data subject may need to construct decoy files to substitute withheld versions.

Steganography is the art of hiding information and can be seen to provide an important complement to encryption to facilitate withholding data against an adversary (who might otherwise force access to encrypted or isolated files). As in the physical world, deniability must be plausible to be effective. This is discussed further in the next chapter.

Surveillance

Surveillance is not only a direct privacy threat but also often a trigger for interrogation. Furthermore, the data collected through surveillance become the subject of information processing and dissemination as discussed below.

Modern technology has made mass surveillance in both the physical and virtual worlds a practical reality. This reduces the efficacy of 'safety in numbers' as a form of anonymity.

In the physical world, an individual's movements and actions can be tracked extensively through CCTV networks, mobile phone signals and other electronic interactions, such as using a credit card. The options for protecting privacy are:

- ❑ *Isolation.* Avoiding contact with surveillance systems can achieve invisibility but constrains movement and action. Adversaries may use covert surveillance.
- ❑ *Disguise.* Anonymity can be achieved by removing the evidence of identity, for example by covering the face, using pay-as-you-go mobiles and paying with cash.
- ❑ *Withholding data.* A data subject can block data either by legal recourse or by direct action such as turning off a mobile. Adversaries may not comply with the law.
- ❑ *Giving false data.* It is possible to construct decoys and alibis in the physical world, for example by placing pseudo-identifiers such as a mobile to corroborate a cover story.

In the virtual world, surveillance covers all online activity, whether sending or receiving emails, 'chat', blogs, web pages, streams, files or other data. For simplicity this paper will assume that the Internet is the communication network being used.

Steps that a data subject can take to protect the privacy of her communications are:

- ❑ *Isolation.* Communication cannot, by definition, be done in isolation. The minimum contact is direct transmission on a secure line; an alternative is a dead drop.
- ❑ *Disguise.* While a data subject can easily create multiple aliases for sending and receiving data, skilled adversaries will be able to trace physical sources.
- ❑ *Withholding data.* A data subject can only withhold a limited amount of data specific to a communication, such as cookies, without corrupting it.
- ❑ *Giving false data.* Tools such as anonymous proxies can make it more difficult, but not impossible, for an adversary to trace a communication.

The brief statements above only hint at the complexity and difficulty of achieving strong privacy in online communications. Systems such as Tor and Freenet have significant time

Deniability Tools are Essential to Strong Privacy

and bandwidth penalties and still cannot guarantee privacy (due to not knowing which nodes may be compromised).

The following chapter will expand on the role of deniability in communications and hence the importance of having a deniable data store that does not contradict online activity.

Information processing

Solove defines five subgroups of information processing as follows:

- ❑ *Aggregation* involves the combination of various pieces of data about a person.
- ❑ *Identification* is linking information to particular individuals.
- ❑ *Insecurity* involves carelessness in protecting stored information from leaks and improper access.
- ❑ *Secondary use* is the use of collected information for a purpose different from the use for which it was collected without the data owner's consent.
- ❑ *Exclusion* concerns the failure to allow the data subject to know about the data that others have about her and participate in its handling and use.

In terms of a data subject's environment:

- ❑ *Privacy and security*. These are discussed below.
- ❑ *Physical and virtual*. Now the data is collected, this dichotomy is unimportant.
- ❑ *Stasis and activity*. Now the data is collected, this dichotomy is unimportant.
- ❑ *Friendly and adversarial*. These are discussed below.

In the scenario of a friendly data holder, the data subject becomes dependent on the security of the data holder's systems and actions. Where the data holder is an adversary (or an incompetent 'friend'), the ability to harm the data subject's privacy could have been mitigated by earlier manipulation of the data supplied.

By the time a data holder or adversary is processing information about a data subject, it is too late to use 'isolation' or 'withholding data' privacy tools. However, earlier use of 'disguise' or 'giving false data' can mitigate privacy harms. This is explored briefly for Solove's five subgroups below.

Aggregation

Aggregation of data is much easier to avoid in the virtual world, where carefully managed aliases can in principle be kept separate. In the physical world, unless operating with a bogus identity, it becomes possible for disparate data to be correlated.

Identification

The crux of the privacy threat is mapping data to a real individual. Practical considerations mean that some parts of a data subject's life do require identification, for example access to bank accounts. The issue is that other parts of a data subject's life can get linked, through aggregation, to her identity. To combat this, the data subject needs to deploy isolation or disguise whenever not wishing data to be linked to her identity.

Insecurity

Where a 'friendly' data holder allows an adversary unauthorised access to read or modify the information held about a data subject, the minimum harm is unwanted disclosure but more seriously can amount to so-called identity theft.

Identity thieves do not, of course, steal identities, but rather achieve a sufficiently good impersonation of a data subject for their own gain. In this situation, a potential defence of a data subject is to provide the most difficult-to-impersonate information to prove her

Deniability Tools are Essential to Strong Privacy

identity. This is her DNA; ironically many people are opposed to supplying this non-deniable data to anybody.

Secondary use

A data subject's main privacy tool to deter or seek remedy for a breach of trust by a data holder is to introduce a limited amount of false data to create a form of 'watermark'. For example, by deliberately misspelling her surname when ordering goods, a data subject will still receive the items but also then be aware if that name/address pair is used for purposes contrary to her stated preferences.

Exclusion

This subgroup is not relevant to this paper.

Information dissemination

Solove defines seven subgroups of information dissemination as follows:

- ❑ *Breach of confidentiality* is breaking a promise to keep a person's information confidential.
- ❑ *Disclosure* involves the revelation of truthful information about a person that affects the way others judge her reputation.
- ❑ *Exposure* involves revealing another's nudity, grief, or bodily functions.
- ❑ *Increased accessibility* is amplifying the accessibility of information.
- ❑ *Blackmail* is the threat to disclose personal information.
- ❑ *Appropriation* involves the use of the data subject's identity to serve another's aims and interests.
- ❑ *Distortion* consists of disseminating false or misleading information about individuals.

In terms of a data subject's environment:

- ❑ *Privacy and security*. These are discussed below.
- ❑ *Physical and virtual*. Now the data is collected, this dichotomy is unimportant.
- ❑ *Stasis and activity*. Now the data is collected, this dichotomy is unimportant.
- ❑ *Friendly and adversarial*. These are discussed below.

As with information processing, consideration of Solove's subgroups highlights opportunities for a data subject to manipulate what data she does supply to mitigate the harm by dissemination:

- ❑ Disguise: Maintaining multiple aliases (assisted by deniability)
- ❑ Giving false data: 'Watermarking' disclosed data

Considering the situation for each subgroup highlights an additional benefit of deploying deniable privacy tools in the case of information dissemination. The potential for traceable aliases and 'watermarked' data can create a disincentive for adversaries to disclose data illegally.

4. Deniability

Role of deniability

The previous chapter showed that deniability can make a unique and substantial contribution to protecting a data subject's privacy across three of the four main groups in Solove's taxonomy: information collection, information processing and information dissemination.

While conventional privacy tools may suffice in idealised situations, deniability is essential in seeking protection against adversaries and incompetent data holders.

Deniability is particularly applicable to the 'disguise' and 'giving false data' privacy options, since it enables a data subject to withstand strong challenges to her cover.

This chapter will focus on deniability in the context of information collection by addressing deniable data storage and deniable communication. Once these have been achieved, the benefits of deniable aliases and deniable false data will flow through to protect from privacy harms in information processing and information dissemination.

Deniable data storage

Deniable data storage is not only critical to protection against interrogation, but is also the foundation of protection against online surveillance.

Steganography, or data hiding, is the principal technique to provide deniability in the context of electronic data. Early systems hid text strings in media files. The quest for a file system with full utilisation of storage capacity and user-friendly management of collisions (with the invisible files) has led to two systems that are used on any scale: StegoStik and Truecrypt. These are compared in a one page analysis available from StegoStik's website at <http://www.stegostik.co.uk/privacy/SSvTC.pdf>.

Effective use of either system requires careful consideration of the traces left by the operating system. Currently there are two relatively inconvenient options for managing this vulnerability. The first is to run a dedicated virtual operating system that can wipe its memory clean on exit; the second is to use a disk cleaning program after using decrypted private data.

Deniable communication

While deniable storage can be a little cumbersome, it is positively user-friendly compared with deniable communication.

Elaborate systems are needed to enable a data subject to establish a connection to a general-purpose communication network (such as the Internet) in such a way that her identity is hidden from adversaries that can monitor her incoming and outgoing traffic. This results in significant reduction in bandwidth and reliability.

Two examples of deniable communication systems are Tor and Freenet. Despite their slow operation, they are still not able to fully guarantee security because it is difficult or impossible to establish which network nodes can be trusted.

They suffer a further issue in that they do not directly integrate with a deniable data storage system, so a data subject is forced to store a fully visible copy of data during the protracted transmit or receive process.

The future of deniability

Deniability is a powerful tool that could enhance the privacy of every data subject. As with encryption, adoption can be increased through education and making the software solutions simpler to use, ideally integrated with the normal operation of the computer.

Wider adoption of deniable storage and communication also reinforces the plausibility that a data subject has deployed the tools not because she has something to hide, but rather because she is using best practice measures to protect her privacy.

StegoStik's vision is to build a PC that is steganographic by default. Each set of data files is revealed when the correct passphrase is entered and is undiscoverable otherwise. Moreover, the operating system will be designed not to leave unwanted traces when working with user data. This will provide the foundation for full integration with a deniable web browsing capability, allowing data requests to be hidden immediately after they are specified, and for responses to be received without being revealed until required.

Deniability is more feasible today than ever before as the continued growth in affordable computing power, data storage and bandwidth make redundancy a reasonable use of resources. However, there is a long journey ahead to provide all individuals with robust data privacy. StegoStik's uniquely powerful patented steganographic file system can be the foundation for this goal.