Best Data Practices for Online Service Providers from the Electronic Frontier Foundation

Introduction

Online service providers (OSPs) are vital links between their users and the Internet, offering bandwidth, email, web and other Internet services. Because of their centrality, however, OSPs face legal pressures from all sides: from users, industry, and government. As an intermediary, the OSP finds itself in a position to collect and store detailed information about its users and their online activities that may be of great interest to third parties. The USA PATRIOT Act also provides the government with expanded powers to request this information. As a result, OSP owners must deal with requests from law enforcement and lawyers to hand over private user information and logs. Yet, compliance with these demands takes away from an OSP's goal of providing users with reliable, secure network services. In this paper, EFF offers some suggestions, both legal and technical, for best practices that balance the needs of OSPs and their users' privacy and civil liberties.

Are you an OSP?

If you think you might be an OSP, you probably are. As defined by the Digital Millennium Copyright Act (DMCA)[1], an OSP is any "entity offering the transmission, routing, or providing connections for digital online communications" or any "provider of online services or network access, or the operator of facilities therefor." The Electronic Communications Privacy Act (ECPA) defines two subcategories of OSPs: "electronic communication services"[2] and "remote computing services."[3] Access to users' information under ECPA is determined in large part by which of these subcategories fits your OSP. As a general rule, email and connectivity services would be electronic communication services, while website hosting would be considered a remote computing service. This means that virtually *any* website or access intermediary, not just established subscriber-based businesses, can be considered an OSP under the law. Indeed, even individuals may be "accidental OSPs" if they set up WiFi access points to share Internet connectivity with friends and neighbors.

How can OSPs develop sane network policies to protect themselves from legal liability and respond to subpoenas and court orders?

A key strategy is to minimize the amount of information OSPs collect and store in the first place. Unless they are in a specially regulated industry (finance or health care, for example), no law requires OSPs to collect and store information about their users. This means that OSP owners and operators are free to develop and implement reasonable data retention policies. Our suggestions for these policies, elaborated below, are for

[1] See http://www4.law.cornell.edu/uscode/17/512.html

[2] "any service which provides to users thereof the ability to send or receive wire or electronic communications…" http://www4.law.cornell.edu/uscode/18/2510.html

[3] "the provision to the public of computer storage or processing services by means of an electronic communications system." http://www4.law.cornell.edu/uscode/18/2711.html

informational uses only. If you have any specific questions or concerns about your OSP, please consult an attorney. (EFF contacts are listed below.)

Legal Issues with Requests for User Data or Transactional Information

When law enforcement officers conduct civil or criminal investigations, they must obtain subpoenas, warrants or court orders to retrieve personal information from OSPs. The government may obtain basic subscriber information[4] with only a subpoena, but generally needs a warrant or a court order for more detailed records. These court orders might request the identity of the user, email message content, visited URLs, search queries, or any other kind of recorded information.

While the ECPA requires OSPs to disclose information in response to a legal process, it also prohibits certain disclosures without a proper request. For example, the ECPA prohibits an electronic communications service provider from producing the contents of electronic communications (i.e. the body of an email message or arguments in a URLs query string), even if served with a subpoena, except in limited circumstances. Thus, the OSP must evaluate the legal process carefully before retrieving the information and furnishing it to law enforcement. Often, this takes a great deal of time and resources, and the OSP should consult an attorney.

An OSP can keep its costs and risks down by setting clear policies about data retention. There are no laws that require OSPs to retain personally identifiable information (PII) or activity logs about users, unless this information is subject to other government regulation (such as financial transactions) or the OSP has received a backup preservation request from the government.[5] EFF believes that PII about users should be kept only so long as it is operationally necessary, and in no event for more than a few weeks. (We explore this issue in more detail in the technical section below.)

OSPs cannot be forced to provide data that does not exist. EFF suggests that OSPs draft an internal policy that states that they collect only limited information and do not retain any logs of user information on their networks for more than a few weeks. If a court order requests data that is more than a few weeks old, the OSP owner can simply point to the policy and explain that he cannot furnish the requested data. This saves the OSP time and money, while also providing the OSP with a X-week long cushion to examine their own logs.

Civil or criminal subpoenas may also be issued for identifying information called "subscriber information." This includes name, address, phone number and any other personal information that the OSP has collected from the user. Subpoenas for subscriber information are usually aimed at uncovering the identities of people who are posting anonymous comments. A typical scenario would be someone posting negative comments about a company. The company lawyer sends a subpoena for subscriber information about the poster, perhaps to determine whether it is an employee who can be fired or sued. Sometimes, these demands are simply used as a form of harassment, without any sound legal basis or intent to follow through with the legal process. In many cases, once

[4] Such as the user's name, address, records of session times and duration, IP or other network address.

[5] See http://www4.law.cornell.edu/uscode/18/2704.html.

the user's identity has been forcefully revealed, the requesting company takes extra-legal action against the user by firing or taking other forms of retribution against him.

Another common civil subpoena is a DMCA “Subpoena To Identify Infringer,” which requires an OSP that hosts allegedly infringing material to disclose “information sufficient to identify the alleged infringer … to the extent such information is available to the service provider.” Unlike an ordinary subpoena, the DMCA subpoena does not require a lawsuit to be filed first, but it must be accompanied or preceded by a notification of alleged infringement that has specific requirements. However DMCA subpoenas only apply to OSPs that actually host a work; not ISPs that merely provide connectivity, such as in the case of peer-to-peer filesharing. DMCA subpoenas also only apply to claims of copyright infringement.

In other circumstances, individuals may request information about a particular user, complaining that the user has engaged in harassment or other bad acts. In such cases, the OSP may be sympathetic to the alleged victim and be tempted to provide the information directly. However, an OSP has no way to verify the truth of the story and providing this information without legal process could subject the OSP to liability from the user. The safest course is to require a subpoena or other legal process before providing user information to anyone.

Remember, Internet users have a right to anonymous free speech under the First Amendment. An OSP receiving one of these subpoenas should notify the user as quickly as possible before responding to it.[6] This will give the user an opportunity to object to disclosure of his or her identity (technically, by filing a “motion to quash the subpoena”). Both Virginia and Arkansas currently require OSPs to give notice to users prior to turning over PII; California is considering a similar bill. Similar laws may soon be enacted in other states. Giving notice may also protect the OSP against lawsuits from users. It is important to set a data retention policy in place now that will protect your users’ privacy and your own legal liability.

Technical Issues

Up until now, we have discussed EFF’s recommendations for best practices to help OSPs minimize the cost of legal overhead. There is also a technical side to this issue. By being consumer-conscious about logging PII, network administrators can proactively save company resources and protect the privacy of their users at the same time. Upon receipt of a court order, OSPs are compelled by law to comb through their logs to extract the requested data using their own resources.[7] Thus, the cost of handling court orders scales proportionally with the retention of user traffic logs.

A general best practice to mitigate this problem is to log only enough information to maintain and upkeep the OSP’s intended services—no more, no less. Logs should be

[6] On occasion, court orders to provide user information to the government may be accompanied with a request not to notify the user. In such circumstances, OSPs should consult with an attorney.

[7] In some cases, OSPs can seek reimbursement for the costs of compliance. See e.g. http://www4.law.cornell.edu/uscode/18/2706.html. However, reimbursement may not capture all the costs associated with legal compliance.

stored for a minimal amount of time. The “correct” strategy for a particular OSP will depend on the services they provide to their users. We outline some possible strategies below.

OSPs must first pinpoint, on every server, all logs where PII is being recorded. It's important to remember that IP addresses and MAC addresses are crucial sources of identity-revealing information, and they are often requested in court orders. The most common locations for PII include:

- DHCP logs (IP address-to-MAC address assignments, session times)
- RADIUS logs (user name, IP address assignment, callback telephone number, session time, etc.)
- Web and FTP server logs (client IP address, files accessed, request time, query string, etc.)
- Email server logs (sender/recipient addresses, message date and time, relay hostnames, etc.)
- Firewall and IDS logs (IP addresses, packet payloads, date and time of connections, protocol used, etc.)
- User contact information databases (mailing address, phone number, billing information, etc.)

For each piece of PII being recorded, it is imperative that network administrators justify why they are keeping the information and consider a realistic time limit for retaining the information. These decisions should be recorded in an internal data retention policy. We outline three possible methods for PII-elimination below: these are obfuscation, aggregation and deletion.

Obfuscation

The easiest, but least protective, strategy is to periodically scrub the logs to obfuscate all explicit or deducible PII. Since virtually all OSPs maintain multiple logs and user information databases, providers must ensure that user identity cannot be gleaned when matching two or more processed logs. Setting a reasonable time duration before PII obfuscation allows OSPs to administer and troubleshoot their networks in real-time. The amount of time PII-exposed logs are stored will depend on the service requirements, but of course PII should never be kept any longer than necessary.

Key solutions to wipe PII from logs include:

- Obscuring the last octet of all IPv4 addresses by either using a randomly seeded one-way hash, or replacing it with an arbitrary integer (between 1 and 254).
- Obscuring the third, fourth and sixth octets of all MAC addresses in the same way as above. This will obfuscate both the exact manufacturer ID (first three octets) and the specific device ID (last three octets) being used.
- Obscuring the last four digits of phone numbers, or replacing it with ‘0000’, but keeping the area code and exchange.
- Obscuring or deleting all usernames in e-mail addresses.

- Obscuring or deleting all query strings in URLs (http://www.google.com/search?~~q=electronic+frontier+foundation~~).
- Obscuring or deleting all filenames from URLs (http://www.eff.org~~/IP/DMCA/unintended_consequences.pdf~~).

Some tactics that should not be used include:

- Encrypting PII with either symmetric or asymmetric keys: Any subpoena or court-order can still force OSPs to turn over the encryption keys along with the encrypted data.
- Hashing PII with a non-random, well-known one-way hash: Using trial-by-error, one could match hashed candidate IP addresses with the encrypted IP address to reveal the original data.

When implemented in a timely fashion, obfuscation gives OSPs the flexibility to glean general usage patterns without retaining PII; implemented poorly, OSPs will continue to be subject to the legal consequences of information requests.

Aggregation

A better strategy is to use aggregation techniques to compile general usage statistics followed immediately by log deletion. This allows OSPs to fully discard all logs, including PII-obfuscated logs, after a specified duration of time, but still keep tabs on network access patterns. OSPs can save a substantial amount of resources using this technique, since aggregation requires minimal hard disk space. It also ensures that no specific PII will be retained on OSP servers in the long term.

Consider an OSP which hosts an Internet search engine and wants to track popular search queries. Obfuscation of the query string would not work because it would mask the data the OSP wants to track. Obfuscation of only the IP address (while exposing the query string) could still lead to potential IP address matches and PII leaks. Using aggregation techniques, the OSP can simply extract the query strings from the log file, tally the number of times each query was made, and then delete the file completely. One OSP reported to us they automatically aggregate their web server logs every night, then immediately delete the previous day's logs. This method fully decouples users' identities from their search queries while allowing the OSP to keep track of popular search topics.

Deletion

Obfuscation and aggregation are only effective when used in tandem with log deletion. A strict policy which dictates when the OSP should fully purge logs from hard drives is a mandatory step in minimizing the potential challenges of legal compliance. Decisions on log retention time intervals will vary drastically. Free, open WiFi providers may delete connection logs immediately after log-off, while pay-per-use WiFi providers must keep logs for weeks until billing and collection have been completed. OSPs should note that different types of log files may have different data retention intervals.

Even after logs have been deleted from disk, the PII may still reside on the disk until that memory segment is reused and written over. Even then, advanced forensic searches of server hard drives could still reveal past data stored on them. These processes may cause OSPs significant disruptions. If possible, you should use strong deletion utilities to fully scrub the hard drives containing deleted logs. This will ensure the removal of all sensitive PII.

The best way to protect against the risk of log artifacts on disk is to never create any user logs in the first place. This is the ideal and safest solution even though it is often impractical. By reconfiguring the logging preferences in server applications, one can easily change the log level to record nothing about network events. But for most OSPs, these logs are necessary for network troubleshooting and security precautions. This is also virtually impossible for large, for-profit providers that need to maintain billing and subscriber contact information. Thus, the best tactic for an OSP is to come up with a safe and sane network policy in which logs are retained for the shortest possible time.

Summary of Recommendations

a. Develop procedures for dealing with legal information requests and providing notice to users.
b. Collect the minimum amount of information necessary to provide OSP services.
c. Store information for the minimum time necessary for operations.
d. Effectively obfuscate, aggregate and delete unneeded user information.
e. Maintain written policies addressing data collection and retention.

Conclusion

OSPs need to understand their legal risks and obligations when codifying their logging practices. They must adopt a reasonable internal data retention policy and follow this policy consistently. Being strict about deleting all PII on servers will protect OSPs from many hidden costs. By taking proactive technical steps, and knowing their legal rights and obligations, OSPs can simultaneously maximize the privacy of users and protect themselves from the damaging effects of the DMCA, the ECPA and other data disclosure laws.